NBER WORKING PAPER SERIES

EFFECTS OF EPISODE-BASED PAYMENT ON HEALTH CARE SPENDING AND UTILIZATION: EVIDENCE FROM PERINATAL CARE IN ARKANSAS

Caitlin Carroll
Michael Chernew
A. Mark Fendrick
Joe Thompson
Sherri Rose

Working Paper 23926
http://www.nber.org/papers/w23926

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017

We would like to thank Tom McGuire and Randall Ellis for helpful comments and suggestions. We gratefully acknowledge financial support from the Laura and John Arnold Foundation, the Agency for Healthcare Research and Quality T32 trainee program (Carroll) and the National Science Foundation Graduate Research Fellowship (Carroll). Thompson wishes to disclose his involvement in developing the Arkansas Health Care Payment Improvement Initiative, both as Arkansas Surgeon General and as President of the Arkansas Center for Health Improvement. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w23926.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Caitlin Carroll, Michael Chernew, A. Mark Fendrick, Joe Thompson, and Sherri Rose. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Effects of Episode-Based Payment on Health Care Spending and Utilization: Evidence from Perinatal Care in Arkansas
Caitlin Carroll, Michael Chernew, A. Mark Fendrick, Joe Thompson, and Sherri Rose
NBER Working Paper No. 23926
October 2017
JEL No. I1,I11

ABSTRACT

We study how physicians respond to financial incentives imposed by episode-based bundled payment (EBP), which encourages lower spending and improved quality for an entire episode of care. Specifically, we study the impact of the Arkansas Health Care Payment Improvement Initiative, a multi-payer program that requires providers in the state to enter into EBP arrangements for perinatal care. Because of its multi-payer nature and the requirement that providers participate, the program covers the vast majority of births in the state. Unlike fee-for-service reimbursement, EBP holds physicians responsible for all care within a discrete clinical episode, rewarding physicians not only for efficient use of their own services but also for efficient management of other health care inputs. In a difference-in-differences analysis of commercial claims, we find that perinatal spending decreased by 3.8% overall in Arkansas after the introduction of EBP, compared to surrounding states. We find that the decrease was driven by reduced spending on non-physician health care inputs, specifically the prices paid for inpatient facility care, and that our results are robust to a number of sensitivity and placebo tests. We additionally find that EBP was associated with a limited improvement in quality of care.

Caitlin Carroll
Harvard Medical School
Dept. of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
caitlincarroll@fas.harvard.edu

Michael Chernew
Harvard Medical School
Dept. of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
and NBER
chernew@hcp.med.harvard.edu

A. Mark Fendrick
University of Michigan
2800 Plymouth Road
Building 16 / Floor 4
Ann Arbor, MI 48109
amfen@umich.edu

Joe Thompson
Arkansas Center for Health Improvement
1401 W Capitol Ave
Victory Bldg., Suite 300
Little Rock, AR 72201
thompsonjosephw@uams.edu

Sherri Rose
Harvard Medical School
Dept. of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
rose@hcp.med.harvard.edu

1 Introduction

Understanding how physicians respond to financial incentives is a central issue in health economics. Compensation structures can improve efficiency in markets for physician services, addressing principal-agent problems that arise from informational asymmetries and non-contractible quality (McGuire, 2000). There is broad consensus that fee-for-service (FFS) reimbursement distorts care provision away from the social optimum and incentivizes overtreatment.[1] Recent debate over how to restructure physicians' financial incentives has included proposals to bundle related services into broader payments. Episode-based bundled payment (EBP) pays a case rate for an entire episode of care, imposing supply-side cost sharing for spending outside of a target range. Because EBP combines physician fees with payments to all other providers involved in the episode, EBP not only rewards physicians for efficient use of their own services, but also for efficient management of other health care inputs that have traditionally been separately reimbursed.

The influence of payment systems on physician behavior is widely recognized and has generated a robust theoretical and empirical literature.[2] Most empirical work is focused on capitation and salary contracts (Ho and Pakes, 2014; Jensen, 2014; Hennig-Schmidt et al., 2011; Dusheiko et al., 2006; Barro and Beaulieu, 2003; Gaynor and Gertler, 1995) and relatively little is known about how physicians respond to EBP. Existing studies of EBP are generally limited to observational evidence from small demonstration projects (Navathe et al., 2017; Dummitt et al., 2016; CMS, 2013; Casale et al., 2007; Cromwell et al., 1997), mostly in the Medicare market. This literature concludes that EBP is associated with efficiency gains, especially through improved management of non-physician health care inputs such as medical devices, post acute care and inpatient facility care. However, concerns about provider selection into these programs make causal inferences difficult. Moreover, the magnitude of physician responses to efficiency incentives is important, and the small scale of these demonstration projects limit the generalizability of the findings.

In this paper, we study physician responses to EBP in the context of the Arkansas Health Care Payment Improvement Initiative (APII), a state-wide program with mandatory provider participation. Implemented in 2013, the APII is among the most extensive EBP models in use and, to the best of our knowledge, the only large-scale EBP program that is obligatory for providers. The APII is built on a partnership of payers in the state, most notably Medicaid and Arkansas Blue Cross Blue Shield (AR BCBS), which accounts for

[1] More precisely, fees above marginal cost create incentives for oversupply; fees below marginal cost can create incentives for underprovision (McGuire, 2011).

[2] See McGuire (2000) for a review.

80% of the state's large group market (KFF, 2014b).[3] Under EBP, insurers set episode-level spending targets and hold a Principal Accountable Provider (PAP) responsible for a portion of any excess spending. EBP in Arkansas initially covered five episodes: Perinatal, Attention Deficit/Hyperactivity Disorder, Hip and Knee Replacement, Congestive Heart Failure, and Upper Respiratory Infection. Like many modern EBP programs, the APII employs a FFS model with reconciliation, rather than prospective EBP, which imposes a single, up-front payment to be shared by all providers involved in episode care.

The APII presents a novel opportunity to study the effects of EBP in the commercial market and to understand its impact at a system level. We study the effects of EBP on perinatal care in Arkansas. Studying EBP in the context of perinatal care has several advantages. First, because Medicare has such a small role in covering perinatal care, the partnership of Medicaid and commercial insurers covers the vast majority of births in the state and impedes physician selection into different payment systems. Second, spending on perinatal care in the United States is substantial and associated with large variations in episode costs, suggesting that savings under EBP are potentially large (Xu et al., 2015; Glantz, 2012; Main et al., 2012). Third, incentivizing provider coordination is important in the context of perinatal care, which typically occurs across a variety of clinical settings and involves management of multiple health care inputs. Lastly, perinatal care is well-suited to episode level bundling. From the physician's perspective, the total number of perinatal episodes is fixed, minimizing concerns about endogenous episode volume.[4] Moreover, episode spending accounts for a large share of total annual spending for this population, suggesting that efficiency gains to broader bundled-payments would be small.[5]

Our study makes several contributions to the literature. First, we develop a stylized model of physician behavior under FFS with episode-based reconciliation. We show that this EBP structure creates two broad incentives: (1) it discourages overprovision of services that PAPs deliver directly and are compensated for with FFS payments (e.g., performing a cesarean section) and (2) it incentivizes improved management of health care inputs that

[3]QualChoice Arkansas (QCA) also participates in EBP, but provider participation is voluntary (ACHI, 2016). QCA held 7% of the large group market in 2014.

[4]Under EBP, physicians are responsible for spending once an episode occurs, but are not held responsible for the frequency of episodes. Physicians therefore benefit financially from a high number of episodes as long as the case-rate is sufficiently high. These observations suggest that EBP is less appropriate in clinical areas where the volume of episodes is endogenous to physician effort, either through prevention or through "upcoding" (Dafny, 2005) related clinical conditions to appear as episodes.

[5]Capitated payments, for example, would cover more services but would require substantial integration across providers. Further, if capitated payments targeted primary care physicians, as is typical, patient-level bundling could introduce a new agency problem into care provision. In particular, informational asymmetries arise if primary care physicians know less about treatment options for perinatal care than the obstetricians they refer patients to.

PAPs influence indirectly but are not compensated for through the FFS system (e.g., referrals to teaching hospitals). We further argue that building EBP on top of the existing FFS reimbursement system mutes incentives to adjust provision of direct services, relative to indirect services. Second, we provide the first empirical evidence on physician responses to mandatory EBP and assess the potential for such payment reforms to influence spending on a large scale. We separately estimate the effect of EBP on PAPs' direct service provision and on PAPs' use of other health care inputs, testing the predictions of our conceptual model. Third, we contribute evidence about the effect of EBP in the commercial market, where payment reforms must contend with inefficiencies in prices (Cooper et al., 2015b) as well as utilization patterns.[6]

We estimate the effects of perinatal EBP using a difference-in-differences (DD) analysis of commercial claims, comparing Arkansas to its neighboring states after conducting extensive validity tests of the DD design. We find that total episode spending declined in Arkansas by 3.8% in the first year of full program implementation, relative to surrounding states, and that the decrease was driven by a change in indirect service provision. In particular, we estimate a 6.6% savings on inpatient facility spending and qualitatively smaller, statistically insignificant savings on professional inpatient spending and outpatient spending. We pursue a variety of sensitivity tests and find that these results are robust to alternative control groups, covariate selection, and sample restrictions.

We further explore the spending reductions under EBP by decomposing changes into quantity and price components. In the inpatient setting, we find little evidence of changes in utilization patterns, including cesarean section rates and lengths of hospital stays. Instead, we find that the reduction in inpatient facility spending was largely driven by the price of care. This conclusion is robust to placebo tests that repeat our analyses using services that were not subject to EBP incentives. In the outpatient setting, we study utilization of screening tests identified by Arkansas as markers of quality. We find only a modest improvement in the quality of perinatal care in Arkansas; out of six screening tests, we find increases in utilization of one under EBP.

Lastly, we identify two broad mechanisms that could underlie the decreases in episode spending under EBP, specifically the change in inpatient facility prices. Reductions in the price of care could reflect (1) referral patterns favoring low price facilities or (2) lower negotiated rates at a given facility. We develop indirect tests of these mechanisms and find preliminary evidence that a change in referral patterns is more likely.

[6] In early demonstration projects, for example, Medicare achieved spending reductions largely by imposing lower hospital payment rates, not through improvements in care efficiency (CMS, 2013; Cromwell et al., 1997). The ability of commercial payers to negotiate lower prices will depend heavily on local market structure and this avenue of spending reduction may only be feasible in pockets of the commercial market.

The paper proceeds as follows. Section 2 provides background on perinatal EBP in Arkansas. Section 3 lays out our conceptual model. Section 4 describes the data. Section 5 discusses our empirical strategy. Section 6 describes the spending, utilization and price results. Section 7 describes our robustness tests. Section 8 explores the mechanism behind the decline in inpatient facility prices and Section 9 concludes.

2 Perinatal EBP in Arkansas

2.1 Perinatal Episode Design

Under EBP in Arkansas, the perinatal episode is triggered by a live birth and includes all care associated with the delivery of the infant, prenatal care in the 40 weeks prior to the birth, and postpartum care in the 60 days after. One PAP is assigned responsibility for all episode spending. In the perinatal episode for AR BCBS, the PAP is the provider that oversees the delivery.[7] Episode spending is risk-adjusted based on documented patient comorbidities; PAPs can therefore mitigate financial risk by more thoroughly reporting patients' clinical conditions, analogous to "upcoding" under Medicare's Diagnosis Related Group (DRG) system (Dafny, 2005). Further details about the risk adjustment algorithm, for example which comorbidities trigger risk adjustment, have not been released by AR BCBS.

In addition to risk adjustment, certain comorbidities exclude episodes from EBP spending targets. These comorbidities are known publicly and include conditions such as type I diabetes and severe pre-eclampsia (AR BCBS, 2014).[8] Spending on valid (non-excluded) episodes determines gain- and risk-sharing amounts for each PAP. While episode exclusion protects PAPs against financial risk, it also raises concerns about gaming of exclusion criteria. In particular, by shifting expensive episodes to the excluded group, PAPs can lower spending on valid episodes, potentially earning gainsharing dividends (or avoiding risk-sharing penalties) with no real change in total episode spending.

PAPs receive quarterly reports that track their risk-adjusted spending, utilization patterns and quality of care. Each year, commercial insurers calculate a risk-adjusted spending average among valid episodes for every PAP, then assign gainsharing dividends or risk-sharing penalties to PAPs that have overseen at least five episodes. PAPs' spending averages are deemed Acceptable, Unacceptable, or Commendable according to pre-determined spending thresholds that are based on historical spending in Arkansas. PAPs with Unacceptable

[7] Under Medicaid perinatal EBP, the hospital where the delivery occurred can be the PAP, although this has not been common.

[8] Episodes can also be excluded for non-clinical reasons, for example if the delivering provider did not perform the majority of prenatal care. Because our ability to observe these disqualifications is limited, we do not pursue analyses of non-clinical exclusions.

spending are responsible for 50% of the excess spending beyond the Acceptable threshold and PAPs with Commendable spending can share in 50% of the savings. Acceptable spending thresholds appear relatively generous. Of the 134 PAPs overseeing AR BCBS episodes in 2012, 92 had Commendable spending averages and only 12 had Unacceptable spending averages (ACHI, 2015).[9] PAPs are disqualified from gainsharing if they do not meet certain quality metrics but are not otherwise penalized or rewarded based on quality. In the perinatal episode, quality metrics are outpatient process measures; PAPs satisfy quality requirements with 80% screening rates for group B strep, chlamydia and HIV.

2.2 EBP Program Diffusion

Perinatal EBP enjoys nearly universal participation in Arkansas. In 2013, Medicaid accounted for 67% of births in the state and implemented EBP for all beneficiaries, except for those also covered by Medicare (Allison, 2013).[10] In the private market in 2013, perinatal EBP applied to all AR BCBS fully-insured groups, 17% of AR BCBS self-insured members, AR BCBS employees, and Baptist Health employees; QualChoice Arkansas (QCA), the state's third largest commercial insurer, implemented EBP on a voluntary basis for providers. Thus, in 2013 the vast majority of births were covered by EBP, with the major exception being self-insured groups covered by AR BCBS. To the extent that providers cared for a mix of publicly and privately insured patients, however, the substantial Medicaid market share created an avenue for spillover effects (Glied and Zivin, 2002; Newhouse and Marquis, 1978). In 2014, AR BCBS expanded perinatal EBP to all self-insured groups. By 2014, therefore, almost all births were covered by EBP; the only major payer excluded from EBP in Arkansas was Medicare, which covers only a small fraction of births.[11]

2.3 Financial Incentives Before and After EBP

Under FFS, physicians receive a global maternity fee that covers all inpatient intrapartum professional care, as well as routine outpatient professional services provided during the prenatal and postpartum periods.[12] Physicians are separately reimbursed for certain outpatient

[9] In October 2012, the draft Acceptable spending category ranged from $8,650 to $10,100 per episode.

[10] Medicaid appears to have faced implementation challenges in 2013 and functionally applied EBP to a fraction of covered births (ACHI, 2015). Given that PAPs likely could not foresee these challenges, we assume that Medicaid EBP exerted pressure on spending in 2013.

[11] Medicare might cover a pregnancy if a beneficiary is eligible through disability or end-stage renal disease, as opposed to age.

[12] We use intrapartum care to mean services rendered during the childbirth hospitalization. This use of intrapartum departs from the clinical definition, which begins at the onset of labor and concludes with the delivery of the placenta.

procedures, such as laboratory work. Hospitals and other facilities are reimbursed separately for services rendered in those settings. Thus, while physicians decide on the intensity of facility spending, they do not stand to gain or lose financially from these management decisions.[13] Inpatient facility services are reimbursed via DRG-level bundled payments; DRG rates are higher for cesarean sections and for clinically complex patients. In the commercial market, inpatient facility payments also vary within a DRG type across hospitals.

Under EBP, physicians have a financial stake in all services provided in the episode window. The perinatal bundle maintains the underlying FFS structure, but sets an Acceptable spending range to cover the global maternity fee, any other professional fees and all facility fees. PAPs can reduce direct spending on their own services by decreasing professional inpatient spending, for example via cesarean section rates, or by decreasing spending on outpatient services that fall outside the global fee. Notably, EBP also creates a counteracting incentive to increase spending on outpatient services if it allows PAPs to meet quality requirements. PAPs can reduce indirect spending on other health care inputs by decreasing facility spending, for example by referring patients to low price hospitals or by adjusting hospitalization lengths; although length of stay (LOS) does not generally affect spending under a DRG system, facility reimbursements are higher if LOS exceeds a high trim threshold and lower if LOS is below a low trim threshold.[14] Cesarean section rate reductions will also reduce indirect spending via lower DRG payments. Lastly, PAPs can decrease both direct and indirect spending by documenting patient comorbidities such that expensive episodes are excluded from EBP incentives.[15]

Financial incentives under EBP are potentially strong for obstetricians. Inpatient facility spending for commercial perinatal care is substantial, suggesting a meaningful change in financial incentives under EBP (Truven, 2013). Additionally, the majority of OBGYNs are self employed and generate income from direct patient care, especially overseeing pregnancies and childbirths (ACOG, 2011).

3 Conceptual Model

In this paper, we are interested in how physicians adjust their treatment intensity in response to payment reform. We begin with a physician utility model in the style of Ellis and McGuire (1986), where PAPs maximize utility over profits and patient well-being. We extend the

[13]Examples include decision making around length of stay or referring patients to high or low price hospitals.

[14]Medicaid reimburses hospitals for perinatal care on a per-diem basis, causing incentives under EBP to diverge across payers.

[15]Relatedly, increased documentation of patient comorbidities will decrease episode spending via risk adjustment even if the episode is ultimately subject to EBP incentives.

model by relaxing the assumption that PAPs bear the full cost of treatment decisions. In particular, our model allows PAPs to influence the provision of other health care inputs, for example referrals to teaching hospitals, without facing the cost and revenue consequences of those decisions. Under FFS, PAPs maximize utility as follows:

$$\begin{aligned} U^{FFS}(q_d, q_i; \beta) &= \pi^{FFS}(q_d) + \alpha B(q_d, q_i; \beta) \\ &= R(q_d) - C(q_d) + \alpha B(q_d, q_i; \beta) \end{aligned}$$

where π is the PAP's profit, B is patient benefit and both functions are concave in treatment intensity. Patients vary only by their sickness level β and physicians have preferences over profit and patient benefit according to an agency parameter $\alpha \geq 0$. PAPs provide two types of services: direct services q_d that are delivered personally by the PAP, and indirect services q_i that are influenced by the PAP but delivered by a separate provider. Patient benefit depends on both types of services, but PAP profits depend only on revenues and costs associated with direct services. Revenues and costs derived from indirect services, $\hat{R}(i)$ and $\hat{C}(i)$, accrue to the provider that renders the treatment, for example a hospital providing bed days. For simplicity of exposition, we assume that direct and indirect services are independent in production ($B_{di} = 0$) and that PAP referrals to indirect services are costless. Utility maximization under FFS yields the following first order conditions:

$$U_d^{FFS} = R_d - C_d + \alpha B_d = 0$$

$$U_i^{FFS} = \alpha B_i = 0$$

where C_d is the marginal resource cost of providing direct services; marginal reimbursement for direct services, R_d, is driven by fees paid for physician services. These fees are generally not set by individual physicians. More often, they are dictated by payers who scale Medicare prices to create fee schedules in the commercial market (McGuire, 2011).[16] Because physician fees are set to cover fixed practice costs, fees are generally above short run marginal cost.

FFS reimbursement tends to incentivize overprovision of both direct and indirect services. Because PAPs do not profit from referrals to indirect services, they will choose q_i to maximize patient benefit ($B_i = 0$), as long as $\alpha > 0$. Socially optimal provision, in contrast, requires PAPs to weigh the benefit of treatment against the cost. In the case of direct services, PAPs will set q_d to maximize patient benefit in the special case of economically neutral fees. If fees are above marginal cost, PAPs may provide services past the point of maximizing patient

[16] Large physician groups sometimes negotiate fees with payers.

benefit. Taken together, these observations suggest that patient care under the FFS system is both too high and produced inefficiently ($\frac{B_i}{B_d} = 0 \neq \frac{\hat{C}_i}{C_d}$).

The Arkansas EBP model is structured as FFS with reconciliation and effectively makes two broad adjustments to the FFS model. First, it reduces the marginal reimbursement for direct services. Second, it imposes partial financial responsibility for indirect care, imposing negative fees for those services.

$$U^{AR}(q_d, q_i, \tilde{\beta}; \beta) = \pi^{AR}(q_d, q_i, \tilde{\beta}; \Delta, A_{min}, A_{max}) + \alpha B(q_d, q_i; \beta)$$

$$\pi^{AR}(q_d, q_i, \tilde{\beta}) = \begin{cases} \pi^{FFS}(q_d) + \Delta[\tilde{\beta}(R(q_d) + \hat{R}(q_i)) - A_{max}] & \text{if Unacceptable} \\ \pi^{FFS}(q_d) & \text{if Acceptable} \\ \pi^{FFS}(q_d) + \Delta[\tilde{\beta}(R(q_d) + \hat{R}(q_i)) - A_{min}] & \text{if Commendable} \end{cases}$$

where A_{max} and A_{min} are the maximum and minimum values of the Acceptable spending range. $\tilde{\beta}$ is the risk adjustment parameter for the episode. $\tilde{\beta}$ is driven by documentation of comorbidities and is therefore a choice variable for the PAP.[17] We take $\tilde{\beta}$ as given in our model, however, because the risk adjustment algorithm is unknown to us and to providers.

After payment reform, PAPs generate profits from direct services as before and additionally have a financial stake, Δ, in total (risk-adjusted) episode spending. Importantly, total episode spending reflects the revenue generated from direct services, $R(q_d)$, and indirect services, $\hat{R}(q_i)$. While revenue from indirect services continues to accrue to the provider that delivers the services, PAPs are held responsible for a portion of the spending via gain- or risk-sharing. When total spending falls outside of the Acceptable range, PAPs share in 50% of the savings or excess spending and $\Delta = -0.5$; otherwise, the profit function simplifies to its FFS form ($\Delta = 0$).

Although EBP in Arkansas exhibits discontinuous incentives around the Acceptable spending thresholds, we treat the PAPs optimization problem as continuous for two reasons. First, PAPs do not know where their average spending will be relative to these thresholds at the end of the performance period. Second, the risk adjustment algorithm is unknown to PAPs, so they do not know with certainty whether adjusted episode spending will fall outside the Acceptable range. We therefore assume that PAPs maximize utility over an expectation of Δ and face a smooth optimization problem characterized by the following first order conditions:

[17] We assume that documentation has revenue consequences through risk adjustment but is costless for PAPs.

$$U_d^{AR} = \pi_d^{FFS} + \Delta\tilde{\beta}R_d + \alpha B_d = 0$$

$$U_i^{AR} = \Delta\tilde{\beta}\hat{R}_i + \alpha B_i = 0$$

Our model has two core predictions. First, EBP creates incentives to reduce provision of both direct and indirect services. Second, incentives to reduce indirect care are greater than those to reduce direct care. We derive comparative statics, $\frac{\partial q_d}{\partial \Delta}$ and $\frac{\partial q_i}{\partial \Delta}$, formally in the Appendix. Intuitively, given the continuation of FFS reimbursement under Arkansas EBP, a decrease in q_d will weakly increase profits by reducing the risk-sharing penalty (or increasing the gainsharing dividend), but this gain will offset by forgone marginal reimbursement for the direct service, π_d^{FFS}. Thus, unless penalties are sufficiently large, reductions in direct service provision can decrease PAP profits under EBP, even in the Unacceptable and Commendable spending ranges. In contrast, reducing the quantity of indirect services unambiguously increases profits in the Unacceptable and Commendable spending ranges.[18,19]

While Arkansas EBP addresses care intensity, it does not necessarily improve productive efficiency. A formal exposition is included in the Appendix. Intuitively, because financial incentives to provide direct and indirect services diverge under FFS, payment reform generally cannot incentivize productive efficiency by applying one penalty parameter to both types of services. This result stands in contrast to the incentives for productive efficiency created by prospective EBP (Ellis and McGuire, 1986).[20]

4 Data

4.1 Data and Covariates

We use Truven Health MarketScan Commercial Claims and Encounters data. The database is a convenience sample of enrollees in commercial health plans and large self-insured firms. The Truven data confer several advantages. First, the data include enrollee identifiers, allowing us to follow patients across time and clinical settings. Second, the data contain information on actual payments to providers, rather than charges. Third, each observation

[18] Assuming quality targets are satisfied.

[19] If we relax our assumption about the independence of service types in production, our model is more complex. Optimal changes in q_d and q_i will depend on whether the services are substitutes or complements. In unreported comparative statics, we show that increases in PAP financial responsibility for care ($d\Delta < 0$) lead to lower treatment intensity of both direct and indirect services unless the service types are highly substitutable in production.

[20] Total patient benefit under EBP may be too high or too low, depending on α, but it will be produced efficiently.

in the Truven data corresponds to a line in an Explanation of Benefits form, allowing us to categorize spending into service areas and to distinguish between professional and facility spending. Lastly, the data include enrollee characteristics such as age, clinical characteristics and health plan identifiers.

We use Truven claims data from 2009-2014 to construct the perinatal episodes that determined PAP performance under EBP in 2013 and 2014, and group perinatal care into episodes in the pre-intervention period as if EBP had been in effect. We first identify all live births that occurred between 2010 and 2014, then track the mothers across time and flag all other claims in the relevant prenatal and postpartum periods. To form our analytic dataset, we collapse the claims into an episode-level database. We divide episodes into pre- and post-EBP years according to the date of the birth, following the assignment rule imposed by payers in Arkansas. Our main analytic sample excludes outlier episodes, namely those with high-outlier spending, although we test the robustness of our results to this exclusion. More detail about our sample construction is included in the Appendix.

For each episode, we construct measures of aggregate spending. Total episode spending is defined as the sum of all payments during the episode window. We further measure spending in the following categories: intrapartum inpatient facility (hereafter intrapartum facility), intrapartum inpatient professional (hereafter intrapartum professional), other inpatient and total outpatient.[21]

We additionally construct measures of utilization for each episode. In the inpatient setting, we measure cesarean section rates and intrapartum LOS. Since LOS outlier definitions vary by payer and are not observable in the data, we examine intrapartum LOS as an indirect test of movement around LOS trim points. In the outpatient setting, we identify the three quality metrics linked to gainsharing and additionally measure screening rates for asymptomatic bacteriuria, hepatitis B and gestational diabetes, which are tracked in the quarterly reports but not linked to gainsharing. Lastly, we classify episodes as valid or excluded based on clinical criteria. We categorize an episode as clinically excluded if a disqualifying diagnosis appears anywhere in a patient's claims during the episode window.

The data have several limitations. First, it is not possible to identify payers and we do not know if perinatal episodes in our sample were covered by AR BCBS or other EBP participants. Given the widespread implementation of perinatal EBP in the state, however, we expect that the program influenced care even among enrollees covered by non-participant insurers through spillover effects. That said, to the extent that the Arkansas perinatal episodes

[21] We could not reliably separate outpatient facility and outpatient professional payments in the Truven data. Regarding other inpatient spending, we do not know whether the PAP delivers the professional component of this care (direct service provision), or if the care is delivered by an on call attending physician (indirect service provision).

in our sample were associated with non-participant insurers, our estimates of the effect of EBP are likely attenuated. Second, we cannot reliably track providers in the Truven data. As a result, we cannot link episodes to PAPs nor to the hospital where the childbirth took place. Relatedly, we cannot characterize PAP responses to EBP according to the proximity of their pre-EBP spending to Acceptable spending thresholds. Given these limitations, we focus our analysis on the impact of EBP at the system level rather than the provider level.

5 Methods

We use a DD approach that compares perinatal episodes in Arkansas to episodes in neighboring states over three pre-implementation years, 2010 to 2012, and two post-implementation years, 2013 and 2014. While full implementation of perinatal EBP in the commercial market was not achieved until 2014, we begin our post period in 2013 to capture the effect of partial implementation and to allow for spillover effects of the Medicaid launch. Since we expect the effect of EBP in the commercial market to be substantively different in 2013 and 2014, we estimate the effect of EBP in Arkansas separately for each post-implementation year.

Our control group is drawn from states in the South Central Census Divisions.[22] From this initial pool of states, we first exclude Tennessee as a potential control based on contemporaneous changes in its reimbursement methods for perinatal care and recent volatility in its insurance markets (Gaither, 2015). We additionally exclude Texas based on its substantial share of urban hospitals relative to Arkansas.[23] Differences in treatment patterns across rural and urban areas have been well documented (ASPE, 2016). Moreover, many Texas hospitals are located in large urban centers that have no equivalent in Arkansas. For each remaining state in the South Central Divisions, we then compare pre-EBP trends in total episode spending to trends in Arkansas (Table A1). Alabama, Kentucky, Louisiana and Oklahoma have statistically similar trends and are included in the control group, though we test the robustness of our results to different sets of control states (Table 6).

We estimate the impact of EBP using the following equation:

$$\begin{aligned} Y_{e,s,t} = \beta_0 + \beta_1 AR_s * 2013_t + \beta_2 AR_s * 2014_t \\ + \delta X_{e,s,t} + \tau_t + \gamma_s + \epsilon_{e,s,t} \end{aligned} \tag{1}$$

where $Y_{e,s,t}$ measures spending, utilization or exclusion for episode e in state s in quarter t.

[22]West South Central: Arkansas, Louisiana, Oklahoma, Texas. East South Central: Alabama, Kentucky, Mississippi, Tennessee.

[23]73% of Texas hospitals and 25% of Arkansas hospitals are categorized as urban (THA, 2017; AHA, 2010).

AR_s is an indicator for Arkansas residence. The coefficients of interest, β_1 and β_2, measure the impact of payment reform in 2013 and 2014, respectively. γ_s and τ_t are state of residence and year-quarter fixed effects. $X_{e,s,t}$ is a vector of covariates including MSA fixed effects, maternal age bins (under 25, 25-29, 30-34, 35+), clinical characteristics (multiple gestation, previous cesarean, fetal malpresentation, fetal distress, preterm birth), plan type (HMO, PPO, POS, HDHP),[24] percent cost sharing quartile bins, fixed effects for the state where the birth took place,[25] and maternal policy holder status (policy holder, spouse, dependent). To determine cost sharing, we calculate each enrollee's percent contribution to intrapartum spending, which accounts for the majority of expenditure, and take an average across all enrollees within a plan; in the Truven data, plan indicates a unique benefit design issued by a particular insurer, rather than a unique insurer.[26]

Following Buntin and Zaslavsky (2004), we model episode spending using a one part generalized linear model (GLM) with a log link function.[27] We apply a modified Park test (Park, 1966) to determine the variance structure and find that the Gamma distribution is most appropriate. Binary variables, namely cesarean section rates, outpatient screening rates and episode exclusion, are estimated using logit models. Intrapartum LOS is estimated using a Poisson distribution, truncated at zero since LOS is positive by definition. Standard errors are clustered at the MSA level, allowing for separate clustering within non-metropolitan areas in each state. Although reimbursement policy varies across states, clustering errors at the state level could bias our standard errors toward zero given the small number of states in our analysis (Bertrand et al., 2004).

We pursue a variety of robustness checks. First, we repeat our analyses using four alternate control groups, dropping each control state in turn. Second, we test whether our results are sensitive to our choice of covariates. Third, we re-estimate our models including high-cost outlier episodes in the sample. Fourth, we assess the possibility that our results are affected by the contemporaneous implementation of the Affordable Care Act (ACA) in 2014. Lastly, we run placebo tests using claims from the study period that were not subject to EBP incentives.

[24] The abbreviations are as follows: HMO - health maintenance organization, PPO - preferred provider organization, POS - point of service, HDHP - high deductible health plan

[25] The majority of women give birth at a hospital located in their state of residence. In a sensitivity test, we define an alternative treatment group based on location of the childbirth; our results are largely unaffected (Table A2).

[26] When our sample includes only one beneficiary per plan, we necessarily use individual percent cost sharing as a control, rather than the plan average.

[27] Total episode spending is positive by definition, as is intrapartum spending. For these spending outcomes, therefore, the GLM must contend with skewness but not with a mass of zeros.

6 Results

6.1 Descriptive Statistics and Validity of Study Design

Table 1 summarizes characteristics of enrollees in Arkansas and the control states, before and after EBP implementation. Differences in enrollee characteristics across the treatment and control groups are small in the pre-EBP period. Prior to EBP, enrollees in Arkansas are slightly younger, less likely to be the holder of their insurance policies, more likely to live in a rural area and more likely to have a POS or HD health plan. There is little evidence of differential changes in enrollee characteristics across EBP implementation. The gap in maternal age grows in the post-implementation period, and there is a convergence in the prevalence of HD health plans. The magnitudes of these differences are small and we otherwise find that enrollee characteristics developed similarly over time.

The key assumption underlying the DD approach is that the treatment and control groups would have evolved similarly if not for the implementation of EBP. Although this assumption cannot be tested, we follow recent empirical literature (Antwi et al., 2015; Kolstad and Kowalski, 2012; Miller, 2012) and analyze pre-implementation trends to assess the validity of our design. Figures 1 and 2 plot unadjusted annual means of all spending and utilization variables, respectively, in Arkansas and the control states. While there is some variation over time in the pre-EBP period, visual inspection suggests that our outcome variables followed similar pre-reform trends across Arkansas and the control states.

We formally analyze pre-implementation trends in Table 2. Using only data from the pre-EBP period, we test for differential linear trends in all dependent variables. Our specification follows the main model, but the parameter of interest is an interaction term between an indicator for Arkansas residence and a linear time trend. The first panel of Table 2 tests for differential pre-EBP trends in all spending variables. By design, there is no trend difference in total episode spending between the treatment and control groups. For all other spending outcomes, we find that Arkansas and the control states exhibit statistically similar trends. Pre-implementation trends in inpatient utilization, likewise, are similar across the treatment and control groups. Of the six outpatient quality measures in our study, we find a statistically significant trend difference only for HIV screening. Based on this difference, we do not pursue an analysis of HIV screening rates. Overall, the results of our pre-trend test support our identification strategy.

6.2 Effect of EBP on Episode Spending

We begin by considering the effect of EBP on total spending, the primary target of reform. Regression estimates of Equation 1 are displayed in Table 3. In the first full year of EBP implementation, we find that total episode spending decreased by 3.8%, or $403, in Arkansas relative to the control states; we find a smaller, statistically insignificant decrease in total episode spending during 2013, consistent with attenuated results under partial program implementation (column 1). Figure 1 shows that these savings were driven by slower spending growth in Arkansas after EBP implementation, relative to consistent spending growth in the control states. In columns 2 and 3, we find similar patterns of spending decline among both valid and excluded episodes. Additionally, we find no evidence of differential changes in the rate of episode exclusion after the introduction of EBP (column 4), suggesting that gaming of EBP via episode exclusion was limited.[28]

To understand what is driving the decrease in total episode spending, we further estimate the effect of EBP on episode spending across service categories (Table 3). Consistent with the predictions of our conceptual model, we find that our results are largely driven by declines in indirect care, specifically intrapartum facility spending. By 2014, intrapartum facility spending decreased by 6.6%, or $329, accounting for over 80% of the overall savings (column 5). In columns 6 and 8, we estimate smaller, statistically insignificant decreases of 1.5% and 1.7% in intrapartum professional spending and outpatient spending, respectively. Our estimates of the effect of EBP on inpatient spending in the prenatal and postpartum periods suggest an increase, but the estimates are imprecise (column 7).

6.3 Effect of EBP on Utilization

In Table 4, we study the effect of EBP on utilization patterns. In the inpatient setting, we find little evidence of changes in treatment patterns. In column 1, we estimate a small, statistically insignificant decline in the cesarean section rate. We additionally estimate a negative but statistically insignificant change in intrapartum LOS, consistent with muted incentives under the DRG system (column 2). Given the reductions in intrapartum facility spending described previously, the stability of utilization patterns suggests that the spending reductions are driven by a price effect. We explore this pattern further in the following section.

In the outpatient setting, we find that EBP had only modest effects on utilization patterns (Table 4). While the rate of chlamydia screening increased significantly under EBP, we

[28] We calculate exclusion rates only for single births because multiple gestation is itself a disqualifying clinical condition.

find no evidence of changes in screening rates for the other quality measures (columns 3 through 7). These results are unsurprising given the pre-EBP screening rates for each of these tests. The baseline screening rate for chlamydia was significantly lower than the gainsharing threshold in the pre-EBP period, incentivizing PAPs to increase their testing rates. Screening rates for the other four tests were already high relative to the gainsharing threshold in the pre-EBP period, limiting potential financial gains from increases.

6.4 Intrapartum Facility Care: Price versus Quantity

In Table 5, we decompose the decrease in intrapartum facility spending into price and quantity components. To isolate the price-effect, we modify Equation 1 by adding a fixed effect for each childbirth DRG.[29] The DD estimates from this model are therefore identified off of variation within a DRG across hospitals. To estimate quantity changes, we standardize intrapartum facility claims to their median DRG payment, thereby holding prices constant, and estimate changes using Equation 1. Consistent with our analysis of inpatient utilization, we find that the decline in intrapartum facility spending is largely a price effect. In column 1, we show that the decline in intrapartum facility spending is mostly unaffected by the addition of DRG fixed effects, and estimate a price decrease of 6.1%. In column 2, we show that there was no meaningful change in price-standardized spending under EBP.

7 Robustness Checks

7.1 Total Episode Spending and Inpatient Facility Spending

In this section, we assess the robustness of our results by pursuing a variety of specification checks, focusing on total episode spending and intrapartum facility spending. We describe four sets of robustness checks in the main text (Table 6) and include additional checks in the Appendix (Table A2). First, we explore the sensitivity of our results to the choice of control group. We construct four alternative control groups by dropping each of the control states in sequence. In columns 1 through 4, we find that our results are not sensitive to the choice of controls. All estimated spending reductions are statistically significant, ranging from 3.1% to 4.2% for total episode spending and 5.6% to 7.7% for intrapartum facility spending.

[29] The relevant DRGs include: 765 Cesarean Section with complication or comorbidity (CC)/major complication or comorbidity (MCC); 766 Cesarean Section without CC/MCC; 767 Vaginal Delivery with Sterilization and/or Dilation and Curettage (D&C); 768 Vaginal Delivery with O.R. Procedure Except Sterilization and/or D&C; 774 Vaginal Delivery with Complicating Diagnoses; 775 Vaginal Delivery without Complicating Diagnoses.

Second, we test whether our results are sensitive to our choice of clinical covariates. One concern about including clinical covariates is that these measures are endogenous to physician effort and coding, both of which may be influenced by EBP. Conversely, our results could be affected by the clinical controls if they are not sufficient to account for differential changes in enrollee characteristics over time. We address these concerns with two specification tests. First, we drop all maternal characteristics from our covariate list. Second, we include an expanded list of maternal covariates, adding anemia, gestational diabetes, hemorrhage, hypertension and hydraminios. The results of these tests are displayed in columns 5 and 6. We find that adjusting the clinical controls does not meaningfully change our results.

Third, we investigate the impact of excluding episodes with high-outlier spending from our analytic sample. The purpose of the high-outlier restriction, defined following AR BCBS guidelines as episodes exceeding three standard deviations above mean spending in a state-half year cell, is to exclude episodes that fall outside a reasonable scope for EBP incentives due to severe health complications. We expect that including these episodes in the analytic sample will attenuate our results. To test the importance of the high-outlier exclusion, we re-estimate our main model, adding high-outliers back into the analytic sample. The results are displayed in column 7. When we include high-outliers, we estimate a statistically insignificant decrease in total episode spending of 2.6% and a statistically significant decrease of 6.5% in intrapartum facility spending.

Lastly, we assess the possibility that our results are affected by the contemporaneous implementation of the ACA in 2014. Since our analysis focuses on the large group market, the most pressing concern is that large employers in Arkansas were more likely to drop coverage under the ACA than employers in the control states. We present two pieces of evidence that suggest employer coverage declines are not a large factor in our analysis. First, data from the Kaiser Family Foundation show that employer shares in the commercial insurance market in Arkansas remained stable across ACA implementation, suggesting little scope for changes in employer insurance offers (KFF, 2014a). Second, our analytic sample relies on a stable subset of employers and the number of deliveries in our sample remains stable throughout the study period (Figure A2); we would expect to see a decrease in covered deliveries over time if employer coverage offers were changing.

7.2 Intrapartum Facility Prices

Next, we test the robustness of our estimated change in intrapartum facility prices. One concern with our analysis is that we are picking up the effects of a general price decrease in

Arkansas relative to the control states, rather than a reduction in perinatal facility prices more specifically. To address this concern, we estimate a series of placebo regressions. We construct a sample of surgical inpatient admissions that fell outside of EBP incentives and randomly select DRGs until we have at least 40,472 observations, matching our main analytic sample.[30] We then estimate the change in inpatient facility prices among these placebo DRGs, testing the hypothesis that EBP had no effect in 2014. We repeat this procedure a thousand times, generating a p-value for each iteration. The results of our placebo tests are plotted in Figure 3. The dashed vertical line denotes a p-value of 0.05 and the solid vertical line denotes the p-value from our analysis of perinatal inpatient facility prices (0.005). In approximately 96% of the placebo DRG samples, the change in intrapartum facility prices is not significant at traditional levels, consistent with a correct null hypothesis and uniformly distributed p-values. Likewise, about 99% of the placebo tests yield a p-value larger than the p-value associated with the intrapartum facility price change. Overall, the placebo tests support the validity of our empirical strategy.

8 Understanding the Decline in Intrapartum Facility Prices

In this section, we explore the mechanism behind the decline in intrapartum facility prices. A price decrease in this setting is consistent with two broad patterns of change: (1) a shift in referral patterns favoring low-price hospitals and (2) a decrease in negotiated rates for inpatient perinatal care in Arkansas relative to the control states. Given the limitations of the provider identifiers in the Truven data, we are unable to test either of these mechanisms directly. Discussions with AR BCBS likewise did not point definitively to either explanation. With these limitations in mind, we develop indirect tests of the two potential mechanisms. The findings from these tests are consistent with a shift in referral patterns but not lower negotiated hospital rates.

8.1 Referral Patterns

In the context of perinatal care and intrapartum facility prices, a change in referral patterns amounts to PAPs delivering more babies at low-price hospitals. A volume shift of this kind could occur if (1) PAPs changed their hospital affiliations to low-price facilities or (2) PAPs with existing affiliations at low-price facilities expanded their practice, stealing volume from

[30] We drop all maternity related admissions in addition to surgical admissions that were subject to their own EBP program.

neighboring hospitals.[31] In either case, a change in referral patterns requires that a market has multiple hospitals for PAPs and patients to choose from.[32]

If price decreases are driven by changing referral patterns, we would expect to find larger effects in multi-hospital markets. We test this prediction, defining markets based on MSA, the smallest geography available in the Truven data. Within Arkansas, MSAs surrounding Little Rock, Fayetteville, Fort Smith and Jonesboro qualify as multi-hospital markets (ADH, 2015; CMS, 2015).[33] Because MSAs are geographically larger than Hospital Service Areas (HSA), our definition of multi-hospital markets overstates the number of enrollees that truly have multiple hospitals available to them. To the extent that we include enrollees from single hospital HSAs in multi-hospital MSAs, our estimates of the price change in multi-hospital markets will be attenuated.

Table 7 presents price change estimates within single- and multi-hospital markets. To generate these results, we split the Arkansas episodes into separate treatment groups according to market type and estimate price changes relative to the full control group in two regressions. Consistent with a change in referral patterns, we find that intrapartum facility prices fell by 8.7% in multi-hospital Arkansas markets, relative to prices in surrounding states, and that there was little change in single-hospital MSAs.

Next, we consider the magnitude of referral changes necessary to generate the price decreases in our analysis. Health Care Pricing Project data on the Little Rock Hospital Referral Region (HRR) shows that the price of perinatal inpatient care is approximately $3,000 higher at the most expensive hospital in the market compared to the least expensive hospital (Cooper et al., 2015a).[34] This price difference holds within both vaginal and cesarean deliveries. If PAPs can save $3,000 by shifting patients from the highest-priced hospital in a market to the lowest-priced hospital, back-of-the-envelope calculations suggest that referring approximately 15% of patients in this manner would account for an 8.7% savings on inpatient facility spending.[35]

[31]It is relatively common for OBGYNs to have multiple hospital affiliations. In 2014 CMS Physician Compare data, approximately 40% of obstetricians in Arkansas had more than one hospital affiliation.

[32]We assume that cross-market movement in this setting is minimal. Given the clinical acuity of childbirth, we expect that patients travel across markets only if they have specific medical needs.

[33]Multi-hospital markets include the following MSAs: Little Rock-North Little Rock-Conway, Hot Springs, Pine Bluff, Fayetteville-Springdale-Rogers, Fort Smith and Jonesboro. We exclude the Memphis and Texarkana MSAs from the multi-hospital group. Although these MSAs contain multiple hospitals, they do not contain multiple hospitals in Arkansas. We expect movement of hospital affiliations across state lines to be limited, since such movement requires separate medical licenses for each state.

[34]Data are from 2008-2011. Pricing information for other Arkansas HRRs and for HSAs is not public.

[35]Data from the Health Care Utilization Project Network show that there were 13,198 commercially insured deliveries in Arkansas in 2012. Our analysis of the Truven data finds that inpatient facility spending on commercially insured births is $4,999 on average, implying that an 8.7% savings amounts to approximately $5.7 million. Thus, an 8.7% savings can be achieved by shifting approximately 1,913 deliveries from the

Our back-of-the-envelope calculations rely on several assumptions. First, we assume that hospital price variation is similar across multi-hospital HRRs. Second, we assume that the highest price hospital provides care to at least 15% of patients, and that these patients can be accommodated at the lowest price hospital. Regarding this second assumption, we note that there is significant variation in prices beyond differences between the highest and lowest price hospitals; savings via referral pattern changes need not be driven by shifts from the most to least expensive hospital, although these savings are by definition the largest.

8.2 Negotiated Hospital Rates

Next, we test whether our results are consistent with a decrease in negotiated rates for inpatient intrapartum care in Arkansas.[36] During our study period, AR BCBS negotiated a single base rate for each hospital in Arkansas and used Medicare Severity-DRG (MS-DRG) weights to dictate the spread of prices across clinical conditions.[37] Given this structure of AR BCBS-hospital negotiations, lower negotiated rates for perinatal care would have necessarily been accompanied by broader, hospital-level price changes. Our placebo tests find no evidence of an across-the-board price decrease in Arkansas; moreover, AR BCBS confirmed that hospitals did not accept lower rates from the insurer in 2014. Our findings thus suggest that lower negotiated hospital rates are unlikely.

9 Discussion

In this paper, we study physician responses to bundled payment incentives relative to FFS reimbursement. We study a mandatory EBP program in Arkansas that covered the vast majority of births in the state and evaluate physician responses in the commercial market. We find that EBP in Arkansas led to a 3.8% decline in total episode spending, driven by a change in physician management of hospital care. In particular, we estimate that spending on intrapartum facility services decreased by 6.6% in Arkansas, relative to surrounding states, and find that the reduction was largely a price effect. Outside of intrapartum facility spending, we find little evidence of changes in care under EBP. We estimate statistically insignificant reductions in inpatient professional spending and outpatient spending. In analyses of inpatient utilization patterns, we find no statistically significant changes in cesarean section rates or in LOS. Out of five outpatient quality measures, we find evidence of improve-

highest price to lowest price hospital.

[36]Although relatively large price changes in multi-hospital MSAs are consistent with referral shifts, they do not rule out changes in negotiated rates.

[37]Thank you to AR BCBS for clarification on this point.

ment in only one, chlamydia screening rates, where baseline performance was low. Extensive sensitivity tests support our main conclusions.

As payers continue to debate the expansion of bundled payment, our analysis suggests that EBP can be successful on a large scale. The magnitude of our results, however, indicates that system-wide bundled payment may have a modest impact compared to effects seen within smaller, voluntary programs. At 3.8% savings overall, our estimates are comparable to evaluations of the Bundled Payments for Care Improvement initiative that include all participants (Dummitt et al., 2016) and results from large population-based payment programs such as Accountable Care Organizations (McWilliams et al., 2015) and early implementation of the Alternative Quality Contract (Song et al., 2012, 2011). Our results are significantly smaller than savings achieved by small, selected groups of providers (Navathe et al., 2017; Casale et al., 2007; Cromwell et al., 1997). More generally, our results contribute to a growing empirical literature on physician agency and basis of payment, showing that such financial incentives affect physician behavior (Jensen, 2014; Schmitz, 2013; Hennig-Schmidt et al., 2011; Barro and Beaulieu, 2003).

Our results have several implications for efforts to expand bundled payment reforms. First, our work suggests that EBP arrangements can improve management of health care inputs. These results are generally consistent with literature documenting the effect of physician payment reform on hospital referral patterns (Alexander, 2016; Ho and Pakes, 2014; Dusheiko et al., 2006) and with studies of bundled payment demonstrations that observe changes in the use of medical devices (Navathe et al., 2017) and post acute care (Dummitt et al., 2016), rather than changes in direct care provision by the accountable provider. While we identify a change in spending driven by inpatient facility prices, we are unable to test directly for an underlying mechanism. We provide preliminary evidence that a change in referral patterns is a likely cause and encourage future work to consider the intersection of EBP reforms and provider market share.

Second, our work argues that modern EBP structures, namely FFS with reconciliation, impose different incentives than prospective EBP. In particular, the continuation of FFS reimbursement creates differential incentives for accountable providers to reduce direct care provision and to manage other health care inputs. From an implementation standpoint, this observation suggests that the designation of the accountable provider will influence the types of savings achieved.

Third, our results highlight potential challenges to coordinating compensation reform across payers. While multi-payer reforms can avoid the "free-rider" problem, designing meaningful incentives can be difficult if there are large differences in treatment patterns across insurance markets. For example, if commercially insured patients receive high quality care relative to Medicaid patients, then applying the same quality incentives across both populations places a higher burden of change on providers with a large Medicaid patient panel and does little to improve quality in the commercial market, although it minimizes administrative burden on providers.[38] Moreover, EBP in the commercial market incentivizes physicians to adjust both the quantity and price of care, causing incentives to diverge from an identically designed EBP program in a public insurance market.

Our analysis has several limitations and suggests potential avenues for future work. First, we focus on the effect of EBP in one clinical area and the results may not be applicable to EBP more generally. Perinatal care centers around an acute hospitalization, is typically under the direction of one provider and has a high and stable volume of episodes. The effect of EBP and challenges surrounding implementation will likely be different for chronic conditions, especially those that are predominantly outpatient based. Second, we are not able to identify providers or health plans in the Truven data; it will be useful for future work to address this gap and examine the effects of EBP at the provider level. Third, our analysis of quality is limited to outpatient process measures. Future work can expand upon these results by studying the effect of EBP on patient outcomes. Lastly, like all DD studies, our results rely on specific assumptions about the absence of unobserved shocks. We address this concern with a number of sensitivity analyses and our results are robust, but caution in interpreting any single study is warranted.

Growing discontent with the FFS system raises important questions about which payment reforms are most appropriate and how well those payment reforms will work. We present evidence that EBP can be effective in reducing spending on perinatal care. While our results are from a specific clinical context, our paper highlights benefits and challenges to bundled payment reform that are more general. Understanding if EBP incentives can be appropriately designed and implemented across diverse clinical settings and heterogeneous provider groups is a natural avenue for future research.

[38] Relatedly, EBP programs must balance imposing meaningful incentives for change with provider buy-in (Hussey et al., 2011). The quality incentives included in the final version of EBP in Arkansas were scaled back from initial discussions due to concerns in the provider community (APII, 2011). While a redesign of incentives could more meaningfully address quality of care in the commercial market, such a change would likely face similar challenges.

References

3M (2016). Implemented State Initiatives. 3M Health Information Systems.

ACHI (2015). Statewide Tracking Report. Arkansas Center for Health Improvement.

ACHI (2016). Arkansas Health Care Payment Improvement Initiative: 2nd Annual Statewide Tracking Report. Arkansas Center for Health Improvement.

ACOG (2011). The Obstetrician-Gynecologist Workforce in the United States. American Congress of Obstetricians & Gynecologists.

ADH (2015). Hospital Provider List and License Verification. Arkansas Department of Health.

AHA (2010). Arkansas Hospital Statistics, Facts and Figures. Arkansas Hospital Association.

Alexander, D. (2016). How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs. *Working Paper*.

Allison, A. (2013). Arkansas Medicaid Program Overview: SFY 2013. Arkansas Department of Human Services.

Antwi, Y. A., Moriya, A., and Simon, K. (2015). Access to Health Insurance and the Use of Inpatient Medical Care: Evidence from the Affordable Care Act Young Adult Mandate. *Journal of Health Economics*, 39:171–187.

APII (2011). Perinatal Workgroup Meeting #2. Arkansas Health Care Payment Improvement Initiative.

AR BCBS (2014). Arkansas Health Care Payment Improvement Initiative Perinatal Algorithm Summary. Arkansas Blue Cross Blue Shield.

ASPE (2016). Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs. Office of the Assistant Secretary for Planning & Evaluation.

Barro, J. and Beaulieu, N. (2003). Selection and Improvement: Physician Responses to Financial Incentives. *NBER Working Paper*.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Brantley, M. (2014). Crittenden Regional Hospital to close Sept. 7. Arkansas Times.

Buntin, M. and Zaslavsky, A. (2004). Too Much Ado about Two-Part Models and Transformation? Comparing Methods of Modeling Medicare Expenditures. *Journal of Health Economics*, 23(3):525–542.

Casale, A., Paulus, R., Selna, M., Doll, M., Bothe, A., McKinley, K., Berry, S., Davis, D., Gilfillan, R., Hamory, B., and Steele, G. (2007). ProvenCare: A Provider-Driven Pay-for-Performance Program for Acute Episodic Cardiac Surgical Care. *Annals of Surgery*, 246(4):613–623.

CMS (2013). Evaluation of the Medicare Acute Care Episode (ACE) Demonstration. Centers for Medicare & Medicaid Services.

CMS (2015). Bundled Payments for Care Improvement (BPCI) Initiative: General Information. Centers for Medicare & Medicaid Services.

Cooper, Z., Craig, S., Gaynor, M., and Van Reenan, J. (2015a). Hospital Prices for Vaginal Delivery Little Rock, AR 2008-2011. Health Care Pricing Project.

Cooper, Z., Craig, S., Gaynor, M., and Van Reenan, J. (2015b). The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured. *NBER Working Paper*, 21815.

Cromwell, R., Dayhoff, D., and Thoumanian, A. (1997). Cost Savings and Physician Responses to Global Bundled Payments for Medicare Heart Bypass Surgery. *Health Care Financ Rev*, 19(1):41–57.

Dafny, L. (2005). How Do Hospitals Respond to Price Changes? *American Economic Review*, 95(5):1525–1547.

Dummitt, L., Kahvecioglu, D., and Marrufo, G. (2016). Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. *Journal of the American Medical Association*, 316(12):1267–1278.

Dusheiko, M., Gravelle, H., Jacobs, R., and Smith, P. (2006). The Effect of Financial Incentives on Gatekeeping Doctors: Evidence from a Natural Experiment. *Journal of Health Economics*, 25(3):449–478.

Ellis, R. and McGuire, T. (1986). Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply. *Journal of Health Economics*, 5(2):129–151.

Gaither, K. (2015). Budget Reductions/Impacts for July 1, 2015. State of Tennessee Department of Health Care Finance and Administration.

Gaynor, M. and Gertler, P. (1995). Moral Hazard and Risk Spreading in Partnerships. *The RAND Journal of Economics*, 26(4):591–613.

Glantz, J. C. (2012). Obstetric Variation, Intervention, and Outcomes: Doing More but Accomplishing Less. *Birth*, 39(4):286–290.

Glied, S. and Zivin, J. (2002). How Do Doctors Behave When Some (But Not All) of Their Patients are in Managed Care? *Journal of Health Economics*, 21(2):337–353.

Hennig-Schmidt, H., Seltin, R., and Wiesen, D. (2011). How Payment Systems Affect Physicians Provision Behavior: An Experimental Investigation. *2011*, 30(4):637–46.

Ho, K. and Pakes, A. (2014). Physician Payment Reform and Hospital Referrals. *American Economic Review*, 104(5):200–205.

Hussey, P., Ridgely, M. S., and Rosenthal, M. (2011). The PROMETHEUS Bundled Payment Experiment: Slow Start Shows Problems In Implementing New Payment Models. *Health Affairs*, 30(11):2116–2124.

Jensen, V. M. (2014). Happy Doctor Makes Happy Baby? Incentivizing Physicians Improves Quality of Prenatal Care. *Review of Economics and Statistics*, 96(5):838–48.

KFF (2014a). Health Insurance Coverage of Adults 19-64. Kaiser Family Foundation.

KFF (2014b). Market Share and Enrollment of Largest Three Insurers- Large Group Market. Kaiser Family Foundation.

Kolstad, J. and Kowalski, A. (2012). The Impact of Health Care Reform on Hospital and Preventive Care: Evidence from Massachusetts. *Journal of Public Economics*, 96:909–929.

Main, E., Morton, C., Melsop, K., Hopkins, D., Giuliani, G., and Gould, J. (2012). Creating a Public Agenda for Maternity Safety and Quality in Cesarean Delivery. *Obstetrics and Gynecology*, 120(5):1194–1198.

McGuire, T. (2000). *Physician Agency*. Handbook of Health Economics.

McGuire, T. (2011). *Physician Agency and Payment for Primary Medical Care*. The Oxford Handbook of Health Economics.

McWilliams, J. M., Chernew, M., Landon, B., and Schwartz, A. (2015). Performance Differences in Year 1 of Pioneer Accountable Care Organizations. *New England Journal of Medicine*, 372(20):1927–1936.

Miller, S. (2012). The Effect of Insurance on Emergency Room Visits: An Analysis of the 2006 Massachusetts Health Reform. *Journal of Public Economics*, 96:893–908.

Navathe, A., Troxel, A., and Lioa, J. (2017). Cost of Joint Replacement Using Bundled Payment Models. *Journal of the American Medical Association*, 177(2):214–222.

Newhouse, J. and Marquis, M. S. (1978). The Norms Hypothesis and the Demand for Medical Care. *The Journal of Human Resources*, 13:159–182.

Park, R. (1966). Estimation with Heteroscedastic Error Terms. *Econometrica*, 34:888.

Schmitz, H. (2013). Practice Budgets and the Patient Mix of Physicians The Effect of a Remuneration System on Health Care Utilization. *Journal of Health Economics*, 32(6):1240–49.

Song, Z., Safran, D., Landon, B., He, Y., Ellis, R., Mechanic, R., Day, M., and Chernew, M. (2011). Health Care Spending and Quality in Year 1 of the Alternative Quality Contract. *New England Journal of Medicine*, 365(10):909–918.

Song, Z., Safran, D., Landon, B., Landrum, M. B., He, Y., Mechanic, R., Day, M., and Chernew, M. (2012). The Alternative Quality Contract, Based On A Global Budget, Lowered Medical Spending And Improved Quality. *Health Affairs*, 31(8):1885–1894.

THA (2017). Texas Hospitals. Texas Hospital Association.

Truven (2013). The Cost of Having a Baby in the United States. Truven Health Analytics.

Xu, X., Gariepy, A., Lundsberg, L., Sheth, S., Pettker, C., Krumholz, H., and Illuzzi, J. (2015). Wide Variation Found In Hospital Facility Costs For Maternity Stays Involving Low-Risk Childbirth. *Health Affairs*, 34(7):1212–1219.

Table 1: Summary Statistics of Arkansas and Control States, Before and After EBP Implementation

	Before EBP Implementation		After EBP Implementation	
	Arkansas	Controls	Arkansas	Controls
Maternal Demographics				
Average Age	28.38	28.89	27.78	28.81
Percent 35+	0.12	0.14	0.11	0.14
Percent Policy Holder	0.37	0.42	0.36	0.38
Percent Non-MSA	0.30	0.20	0.30	0.19
Clinical Characteristics (%)				
Fetal Malpresentation	0.02	0.01	0.01	0.02
Fetal Distress	0.07	0.07	0.06	0.07
Multiple Gestation	0.02	0.02	0.01	0.02
Preterm Birth	0.07	0.06	0.07	0.06
Previous Cesarean	0.18	0.20	0.18	0.20
Plan Characteristics (%)				
HMO	0.02	0.09	0.01	0.07
PPO	0.71	0.74	0.70	0.70
POS	0.14	0.08	0.11	0.06
High Deductible	0.13	0.09	0.18	0.17
Percent Cost Sharing	0.15	0.11	0.16	0.12
Episode Count	2,459	20,885	1,741	15,387

Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. EBP was first implemented in 2013. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Percent cost sharing equals enrollees' contribution to intrapartum spending, averaged across all enrollees within a plan; in the Truven data, plan indicates a unique benefit design issued by a particular insurer, rather than a unique insurer.

Table 2: Tests for Equality of Pre-EBP Trends in Spending and Utilization Outcome Variables

	Total Spending				Categorical Spending (All Episodes)			
	(1) All Episodes	(2) Excluded Episodes	(3) Valid Episodes	(4) Pr(Excluded)	(5) Intrapartum Facility	(6) Intrapartum Professional	(7) Other Inpatient	(8) Outpatient
Arkansas*Time Trend	0.000930	0.00371	0.000183	0.00141	0.00355	-0.00232	-0.00482	0.00401
	(0.00220)	(0.00403)	(0.00167)	(0.0120)	(0.00370)	(0.00125)	(0.0830)	(0.00323)
N	23344	7482	15862	22894	23344	23344	23344	23344
Dependent Variable Means								
Arkansas Mean	10493	11819	9900	0.29	4999	3084	170	2239
Controls Mean	12065	13612	11331	0.31	5977	3075	283	2731

	Intrapartum Utilization		Outpatient Screening Rates					
	(1) Cesarean Section	(2) LOS	(3) Group B Strep	(4) Chlamydia	(5) A. Bact	(6) Hepatitis B	(7) Gest. Diabetes	(8) HIV
Arkansas*Time Trend	0.0186	-0.00641	0.0218	0.00502	0.0144	0.0302	0.0188	0.0481**
	(0.0169)	(0.00390)	(0.0216)	(0.0213)	(0.0254)	(0.0196)	(0.0203)	(0.0170)
N	23344	23344	23344	23344	23327	23344	23344	23344
Dependent Variable Means								
Arkansas Mean	0.39	2.18	0.78	0.55	0.90	0.75	0.78	0.75
Controls Mean	0.40	2.60	0.80	0.56	0.88	0.75	0.79	0.74

Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2012. Table cells include regression coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects. Spending variables are modeled using a one part GLM with a log link and a gamma distribution. Episode exclusion, cesarean section rates and outpatient screening metrics are estimated using logit functions. LOS is modeled as a Poisson truncated at zero. Episode exclusion is based on clinical criteria among episodes with single births; multiple gestation is sufficient to exclude episodes from EBP incentives. Intrapartum refers to the entire childbirth hospitalization. To estimate trends in other inpatient spending, we forced model convergence after 50 iterations and the results should be interpreted with caution. Screening rates for group B strep, chlamydia and HIV are linked to gainsharing. * p<.05; ** p<.01; *** p<0.001

Table 3: Effect of EBP on Perinatal Episode Spending

	Total Spending				Categorical Spending (All Episodes)			
	(1) All Episodes	(2) Excluded Episodes	(3) Valid Episodes	(4) Pr(Excluded)	(5) Intrapartum Facility	(6) Intrapartum Professional	(7) Other Inpatient	(8) Outpatient
Arkansas*2013	-0.0209	-0.00329	-0.0307*	0.0665	-0.0122	0.0115	0.0209	-0.0506
	(0.0170)	(0.0340)	(0.0126)	(0.0769)	(0.0397)	(0.0157)	(0.245)	(0.0292)
Arkansas*2014	-0.0385*	-0.0446*	-0.0310*	-0.0702	-0.0660**	-0.0146	0.139	-0.0170
	(0.0154)	(0.0193)	(0.0146)	(0.112)	(0.0236)	(0.0156)	(0.277)	(0.0363)
N	40472	13106	27366	39720	40472	40472	40472	40472
Dependent Variable Means								
Arkansas Pre-EBP	10493	11819	9900	0.29	4999	3084	170	2239
Controls Pre-EBP	12065	13612	11331	0.31	5977	3075	283	2731
Arkansas Post-EBP	11365	12631	10794	0.30	5472	3186	185	2522
Controls Post-EBP	13434	15059	12633	0.32	6755	3208	263	3208

Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects. Spending variables are modeled using a one part GLM with a log link and a gamma distribution. Episode exclusion is estimated using a logit function among single births; multiple gestation is sufficient to exclude episodes from EBP incentives. Intrapartum refers to the entire childbirth hospitalization. * $p<.05$; ** $p<.01$; *** $p<0.001$

Table 4: Effect of EBP on Perinatal Utilization Patterns

	Intrapartum Utilization		Outpatient Screening Rates				
	(1) Cesarean Section	(2) LOS	(3) Group B Strep	(4) Chlamydia	(5) A. Bact	(6) Hepatitis B	(7) Gest. Diabetes
Arkansas*2013	0.0558	-0.00465	0.139	0.450**	0.0976	0.181	0.234
	(0.0523)	(0.0310)	(0.156)	(0.142)	(0.123)	(0.110)	(0.133)
Arkansas*2014	-0.0449	-0.0320	0.136	0.652*	0.00219	-0.134	-0.177
	(0.126)	(0.0219)	(0.115)	(0.255)	(0.197)	(0.235)	(0.136)
N	40472	40472	40472	40472	40472	40472	40472
Dependent Variable Means							
Arkansas Pre-EBP	0.39	2.18	0.78	0.55	0.90	0.75	0.78
Controls Pre-EBP	0.40	2.60	0.80	0.56	0.88	0.75	0.79
Arkansas Post-EBP	0.37	2.15	0.81	0.71	0.90	0.75	0.78
Controls Post-EBP	0.38	2.56	0.80	0.59	0.87	0.74	0.80

Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects. Cesarean section rates and outpatient screening metrics are estimated using logit functions. LOS is modeled as a Poisson truncated as zero. Intrapartum refers to the entire childbirth hospitalization. Screening rates for group B strep and chlamydia are linked to gainsharing. * $p<.05$; ** $p<.01$; *** $p<0.001$

Table 5: Price-Quantity Decomposition of Intrapartum Facility Spending

	Intrapartum Facility Spending	
	(1) Price Effect	(2) Quantity Effect
Arkansas*2013	-0.0144	0.00191
	(0.0408)	(0.00279)
Arkansas*2014	-0.0611**	-0.00404
	(0.0189)	(0.00594)
N	40472	40472

Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects. Spending variables are modeled using a one part GLM with a log link and a gamma distribution. Intrapartum refers to the entire childbirth hospitalization. The price effect is measured as the change in intrapartum facility spending when DRG fixed effects are included in the model. The quantity effect measures the change in price standardized spending, where each claim is re-priced to its median DRG payment. * p<.05; ** p<.01; *** p<0.001

Table 6: Robustness Checks for Total Episode Spending and Intrapartum Facility Spending

	Alternate Control Groups				Alternate Clinical Covariates		Alternate Outlier Restriction
	(1) Drop AL	(2) Drop KY	(3) Drop LA	(4) Drop OK	(5) Expanded Covars.	(6) No Covars.	(7) Include High Outliers
Total Spending							
Arkansas*2013	-0.0283	-0.0172	-0.0178	-0.0219	-0.0217	-0.0269	-0.000586
	(0.0170)	(0.0171)	(0.0171)	(0.0173)	(0.0185)	(0.0187)	(0.0299)
Arkansas*2014	-0.0415*	-0.0313*	-0.0420**	-0.0397*	-0.0377*	-0.0468**	-0.0262
	(0.0170)	(0.0152)	(0.0157)	(0.0167)	(0.0159)	(0.0150)	(0.0200)
Intrapartum Facility Spending							
Arkansas*2013	-0.0192	-0.00505	-0.0126	-0.0137	-0.0129	-0.0202	0.0182
	(0.0399)	(0.0397)	(0.0402)	(0.0399)	(0.0418)	(0.0423)	(0.0490)
Arkansas*2014	-0.0770**	-0.0558**	-0.0663**	-0.0656*	-0.0647**	-0.0730**	-0.0646**
	(0.0285)	(0.0208)	(0.0252)	(0.0255)	(0.0218)	(0.0223)	(0.0217)
Observations	27634	31042	32580	34360	40472	40472	41912

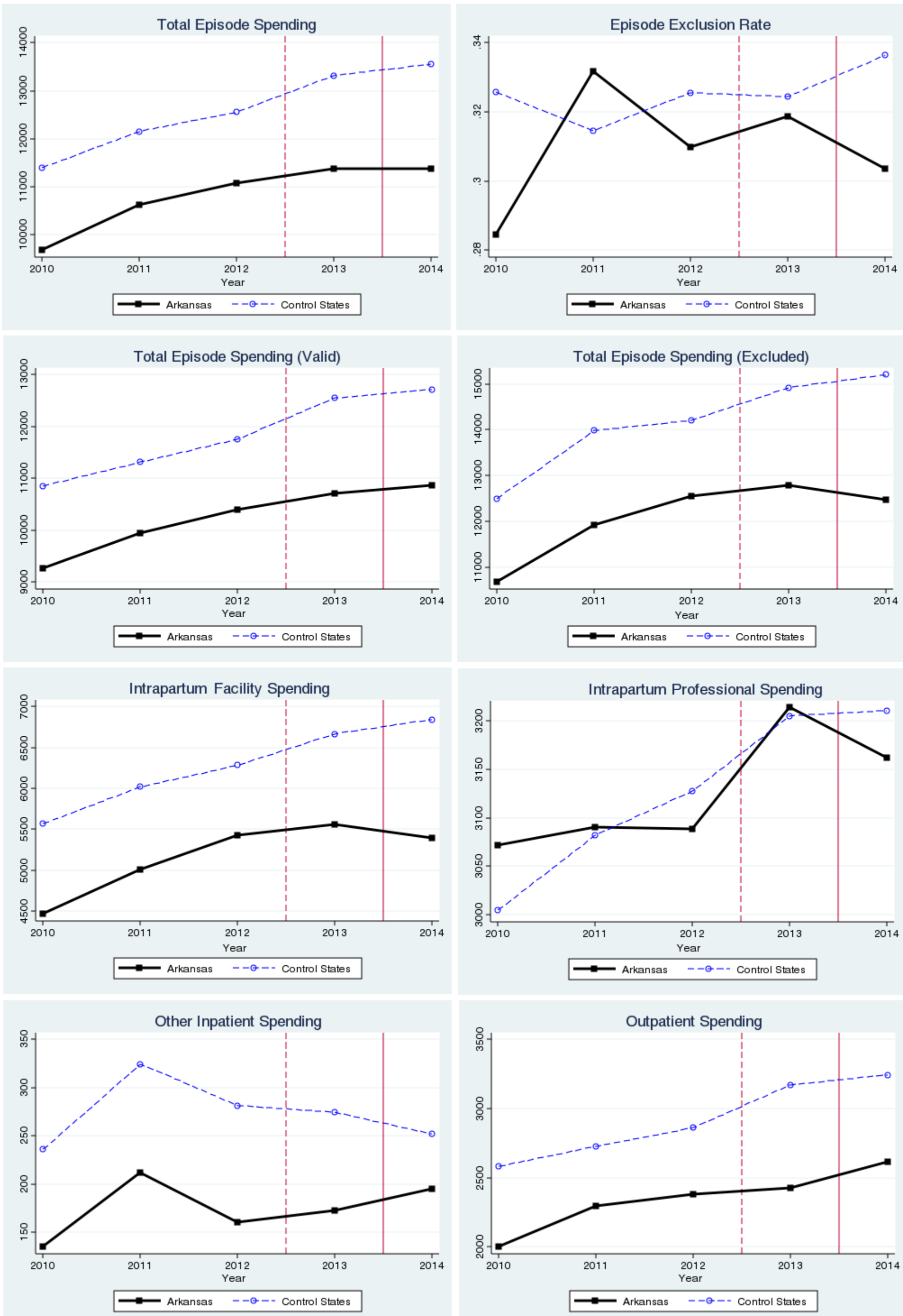
Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Spending variables are modeled using a one part GLM with a log link and a gamma distribution. Intrapartum refers to the entire childbirth hospitalization. In our preferred specification (Table 3), (1) control states include Alabama, Kentucky, Louisiana and Oklahoma and (2) covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects, where maternal clinical characteristics include multiple gestation, previous cesarean, fetal malpresentation, fetal distress, and preterm birth. In columns 1 to 4, we drop each control state in sequence. In column 5, expanded clinical controls include anemia, gestational diabetes, hemorrhage, hypertension and hydrominios. In column 6, we drop all maternal clinical characteristics from the model. In column 7, we include high outliers in our sample, defined as episodes with total spending more than three standard deviations above the mean within a state-half year cell. * $p<.05$; ** $p<.01$; *** $p<0.001$

Table 7: Intrapartum Facility Price Changes by Market Type

	Intrapartum Facility Prices	
	(1) Multi-Hospital	(2) Single Hospital
Arkansas*2013	-0.0589***	0.0581
	(0.0100)	(0.0391)
Arkansas*2014	-0.0868***	-0.0176
	(0.0203)	(0.0213)
N	38914	37830

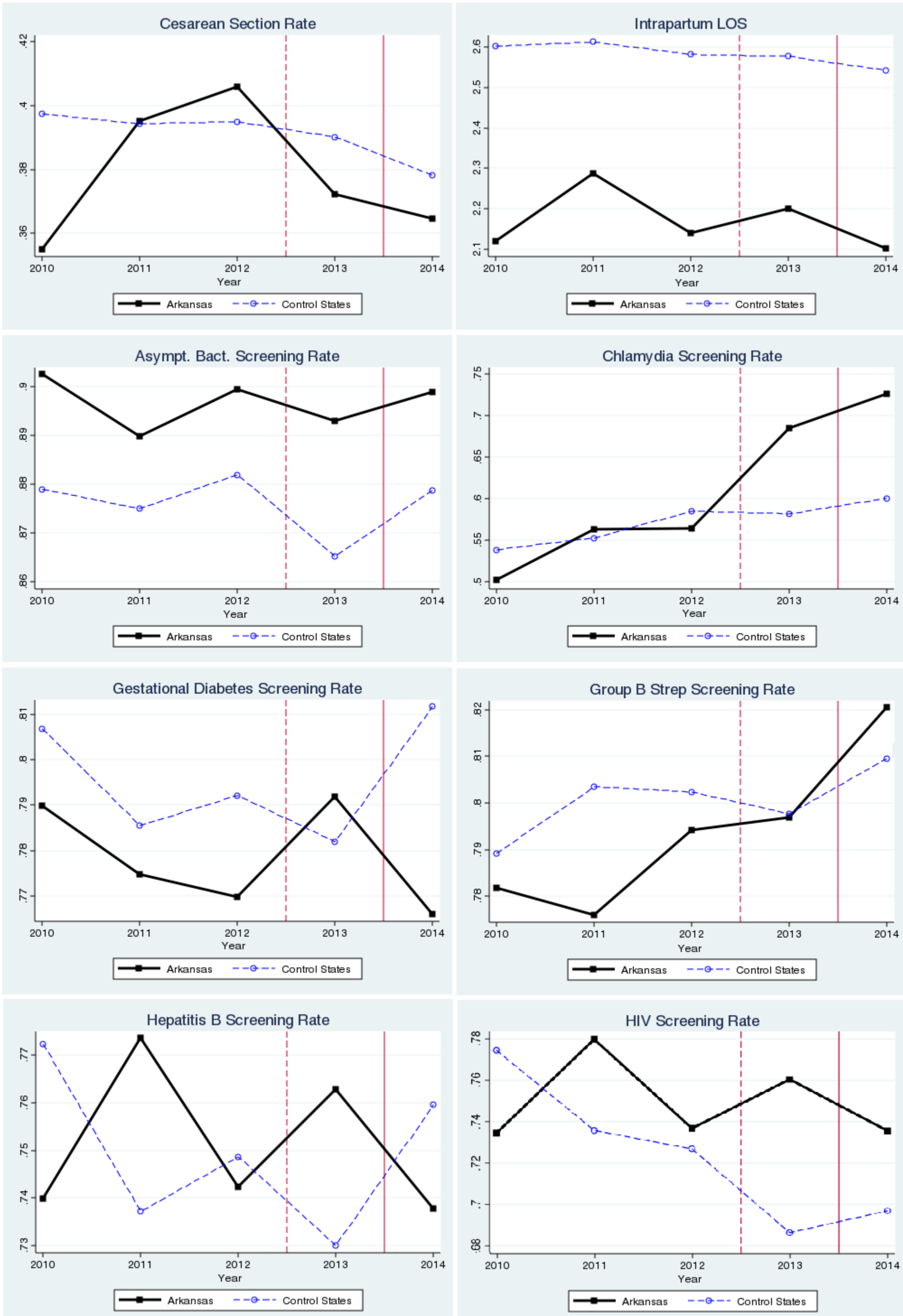
Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects. Inpatient prices are modeled using a one part GLM with a log link and a gamma distribution. Intrapartum refers to the entire childbirth hospitalization. In column 1, multi-hospital markets include the Little Rock, Northwest Arkansas and Jonesboro MSAs. The control group includes episodes from all relevant markets and we split the treatment group into multi-hospital and single-hospital markets across columns. * p<.05; ** p<.01; *** p<0.001

Figure 1: Trends in Spending Outcome Variables in Arkansas and the Control States



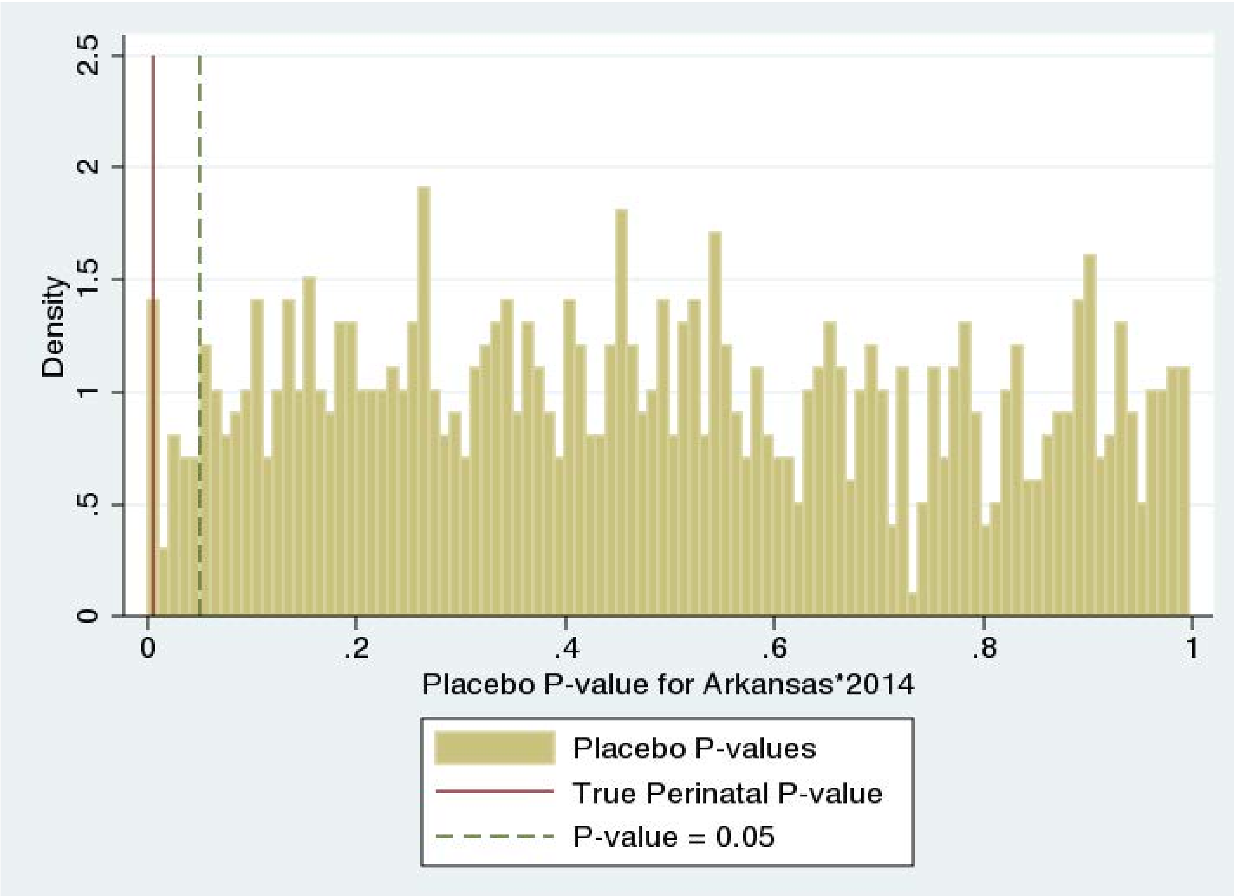
Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Data points are unadjusted, annual means. Control states include Alabama, Kentucky, Louisiana, and Oklahoma. The dashed vertical line denotes partial implementation of EBP in 2013 and the solid line denotes full implementation in 2014. Excluded episodes include those with a disqualifying clinical condition. Intrapartum refers to care delivered during the childbirth hospitalization.

Figure 2: Trends in Utilization Outcome Variables in Arkansas and the Control States



Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Data points are unadjusted, annual means. Control states include Alabama, Kentucky, Louisiana and Oklahoma. The dashed vertical line denotes partial implementation of EBP in 2013 and the solid line denotes full implementation in 2014. Intrapartum LOS is length of stay during the childbirth hospitalization. Screening rates for group B strep, chlamydia and HIV are linked to gainsharing.

Figure 3: Placebo Tests of Intrapartum Facility Price Effect



Notes: Sample estimates from Truven claims, using data from surgical hospitalizations that were not subject to EBP incentives between 2010 and 2014. We randomly select DRGs until the placebo sample has at least 40,472 observations, matching our main analytic sample. Using the random sample of DRGs, we estimate the "effect" of EBP in Arkansas and plot the p-values of the coefficient on Arkansas*2014 as a histogram. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, state and MSA fixed effects, and DRG fixed effects. Inpatient prices are modeled using a one part GLM with a log link and a gamma distribution. Intrapartum refers to the entire childbirth hospitalization. Standard errors are clustered on MSA. The solid vertical line denotes the p-value from our estimated decrease in intrapartum facility prices for perinatal care (Table 5). The dashed vertical line denotes a p-value of 0.05.

Appendix

A1 Sample Construction

As described in the main text, we construct perinatal episodes using Truven claims data from 2009-2014. We identify all live births between 2010 and 2014 and combine all claims in the prenatal and postpartum periods for each birth. Episodes triggered by births in the last two months of 2014 have partial coverage of the postpartum period. Likewise, episodes triggered by births in the first 9 months of 2010 have partial coverage of the prenatal period if the enrollee appears in the Truven data in 2010 but not 2009 (e.g., because their employer began contributing data in 2010).

Our episode construction process mirrors the AR BCBS algorithm except for four deviations. First, for a live birth to trigger an episode, AR BCBS requires both a relevant Current Procedural Terminology (CPT) code for physician services and a relevant DRG code for facility services. We relax this restriction and allow a relevant DRG code to trigger an episode, even when a relevant CPT code is not present. Second, in the prenatal and postpartum periods, the AR BCBS episode includes all inpatient and outpatient claims with a pregnancy classification according to the Episode Treatment Grouper algorithm. We do not have access to this grouper and instead include all claims in the relevant episode window. Third, the AR BCBS episode includes pharmacy claims and we chose to exclude this spending from the analysis. The Truven databases do not include information about potential rebates for drug payments and the prices are likely overstated. Fourth, AR BCBS counts screening tests toward quality metrics if they are conducted during the prenatal period. We calculate screening rates based on tests conducted in the outpatient setting at any time in the episode window.

Because Truven provides a convenience sample of claims that varies across years, our analysis requires several data restrictions to make our episode sample comparable over time. In particular, we subset the database to records from Truven's employer clients that continuously contributed their data between 2010 and 2014. As shown in Figure A1, the volume of Arkansas episodes from non-continuous Truven clients drops in 2013 and 2014, making it impossible to distinguish the effect of EBP from the effect of the changing sample. Among the continuous contributors, Truven advised us to include data from employer clients only. As demonstrated in Figure A1, health plan clients can add or subtract customers from their contribution over time and still be labeled continuous contributors by Truven.

From the continuous employer contributors, we obtain our main sample by removing episodes based on the following criteria: (1) maternal age less than 10 or greater than

55, (2) negative spending in any location-service category, (3) zero professional or facility spending during the intrapartum period, (4) more than one episode trigger during any day in the sample, (5) missing plan type information and (6) overlapping episode timelines, i.e., beneficiaries with pregnancies less than 11 months apart. Episode counts from our main sample are displayed in Figure A2.

A2 Control Group Selection

To select our control group, we compare pre-period spending trends in Arkansas to a pool of candidate control states and select the states where there was no statistical difference in the trends. Our specification follows the main model, but our parameter of interest is an interaction between a linear time trend and a vector of indicator variables for residence in each potential control state. Of our five candidate states, all but Mississippi passed the differential trends test (Table A1).

A3 Additional Robustness Tests

To build on the robustness checks described in the main text, we test the sensitivity of our spending estimates to additional modeling choices (Table A2). First, we test if our spending estimates are sensitive to covariate selection outside of maternal clinical characteristics. We find that our estimates are largely unaffected when we drop plan characteristics as covariates (column 1) and when we drop MSA fixed effects (column 2).

Second, we test if our results are sensitive to functional form. In columns 3 and 4, we run two OLS models, with untransformed spending and log transformed spending as the dependent variables, respectively. Our total spending estimates range from a statistically insignificant decline of 2.8%, using log transformed spending, to a statistically significant decline of 6.4% using untransformed spending. For inpatient facility spending, we find statistically significant decreases in both specifications.

Third, we test the sensitivity of our results to our definition of treatment exposure. In our main sample, we assign enrollees to treatment according to their state of residence. To the extent that Arkansas residents gave birth at out of state hospitals, where EBP incentives are less salient, our analysis underestimates the effect of EBP. Likewise, if control group residents gave birth at Arkansas hospitals and were exposed to the policy treatment, our estimates are biased toward the null. To address this concern, we re-run our analysis using a subset of enrollees where treatment and control assignment is unambiguous: those that gave birth at an in-state hospital. Our results are displayed in column 5. We find that our results are not meaningfully changed.

Next, we test whether our inpatient facility price analysis is confounded by contemporaneous changes in DRG weights over time. Given that AR BCBS employs the MS-DRG system (AR BCBS 2014), changes in DRG weights will affect our analysis if two conditions hold: (1) payers in the control states did not employ the MS-DRG weight system and (2) the weights for perinatal care changed differentially across systems over time. We find little evidence that these conditions hold. First, we do not find evidence that commercial payers in the control states used other DRG systems available in the market, namely the All Patient or All Patient Refined DRG systems (3M, 2016). Second, we do not find that changes in MS-DRG weights during our study period align with our estimated price decreases. Specifically, MS-DRG weights decreased in 2014 only for cesarean deliveries. We analyze the effect of EBP on the price of natural births and continue to find evidence of a price reduction (Table A3).

Lastly, we assess the possibility that our results are confounded by a contemporaneous growth in a low-price insurer in Arkansas or by the closure of high-price hospitals. We find little evidence of such market changes in Arkansas. Examining trends in the large group insurance market, we find that payers like AR BCBS, United and QCA maintained relatively steady shares in Arkansas during our study period (KFF, 2014b).[39] In the hospital market, we find evidence of only one hospital closure in Arkansas under EBP, and it did not close until September 2014 (Brantley, 2014).

A4 Conceptual Model

In this section, we consider the PAP's utility maximization problem under Arkansas payment reform in more detail. Following Ellis and McGuire (1986), PAPs maximize utility over profits and patient benefit according to an agency parameter α. Recall the first order conditions for utility maximization under Arkansas EBP:

$$U_d^{AR} = \pi_d^{FFS} + \Delta\tilde{\beta}R_d + \alpha B_d = 0$$

$$U_i^{AR} = \Delta\tilde{\beta}\hat{R}_i + \alpha B_i = 0$$

[39] With two exceptions, the large group market was relatively stable in the control states during our study period. In Kentucky, Wellpoint increased their market share from 60% in 2011 to 69% in 2014. In Louisiana, Aetna increased their market share in 2014, replacing Coventry as the market's third largest payer. Since our results are driven by changes in Arkansas, rather than the control states, we do not think that these market share changes are driving our results.

To predict how optimal q_d and q_i change under FFS with reconciliation, we fully differentiate the first order conditions with respect to Δ. Recalling that $\Delta = 0$ at baseline, we derive the following comparative statics:

$$\frac{\partial q_d}{\partial \Delta} = \frac{-\tilde{\beta} R_d}{\pi_{dd}^{FFS} + \alpha B_{dd}^{AR}}$$

$$\frac{\partial q_i}{\partial \Delta} = \frac{-\tilde{\beta} \hat{R}_i}{\alpha B_{ii}^{AR}}$$

where concavity implies that second own derivatives are negative, and marginal reimbursement for care is positive. Thus we find that $\frac{\partial q_d}{\partial \Delta}$ and $\frac{\partial q_i}{\partial \Delta}$ are unambiguously positive, confirming an intuitive result that a reduction in Δ (increase in the risk-sharing penalty) will reduce care provision. As discussed in the main text, the optimal change in care provision differs across direct and indirect services because of the continuation of FFS reimbursement. In particular, the incentive to change q_d is muted by the presence of π_{dd}^{FFS} in the denominator. Both $\frac{\partial q_d}{\partial \Delta}$ and $\frac{\partial q_i}{\partial \Delta}$ grow with marginal reimbursement levels (reflecting the fact that savings from service reductions are larger if those services are expensive) and are restrained by changes in patient benefit that accompany adjustments to care provision.

Relatedly, we find that FFS with reconciliation will not generally incentivize efficient care provision. From the first order conditions, we can characterize equilibrium care provision as follows:

$$\frac{B_i}{B_d} = \frac{-\Delta \tilde{\beta} \hat{R}_i}{-\pi_d^{FFS} - \Delta \tilde{\beta} R_d} \neq \frac{\hat{C}_i}{C_d}$$

Under FFS with reconciliation, treatment is provided such that the ratio of the marginal benefit functions equals the ratio of marginal reimbursements; productive efficiency, in contrast, requires equivalence with the ratio of marginal costs.

Table A1: Test for Equality of Pre-EBP Trends in Total Spending

	(1) Total Spending
Alabama*Time Trend	-0.001
	(0.002)
Kentucky*Time Trend	0.003
	(0.003)
Louisiana*Time Trend	-0.002
	(0.002)
Mississippi*Time Trend	-0.007***
	(0.002)
Oklahoma*Time Trend	-0.003
	(0.002)
N	32563

Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2012. Table cells include regression coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects. Total episode spending is modeled using a one part GLM with a log link and a gamma distribution. * $p<.05$; ** $p<.01$; *** $p<0.001$

Table A2: Additional Robustness Checks for Total Spending and Intrapartum Facility Spending

	Alternate Covariates		Alternate Functional Form		Alternate Treatment Exposure
	(1) Drop Plan Controls	(2) Drop MSA FE	(3) OLS	(4) Ln	(5) Residence and Childbirth Location
Total Spending					
Arkansas*2013	-0.0147	-0.0208	-454.4*	-0.0164	-0.00879
	(0.0160)	(0.0164)	(192.0)	(0.0157)	(0.0215)
Arkansas*2014	-0.0362*	-0.0343*	-678.3***	-0.0280	-0.0379*
	(0.0148)	(0.0159)	(177.3)	(0.0159)	(0.0187)
Intrapartum Facility Spending					
Arkansas*2013	-0.00781	-0.00812	-207.2	0.00168	0.00853
	(0.0388)	(0.0398)	(233.5)	(0.0313)	(0.0454)
Arkansas*2014	-0.0621*	-0.0624*	-534.6***	-0.0543**	-0.0682**
	(0.0245)	(0.0247)	(133.2)	(0.0196)	(0.0242)
N	40472	40472	40472	40472	35508

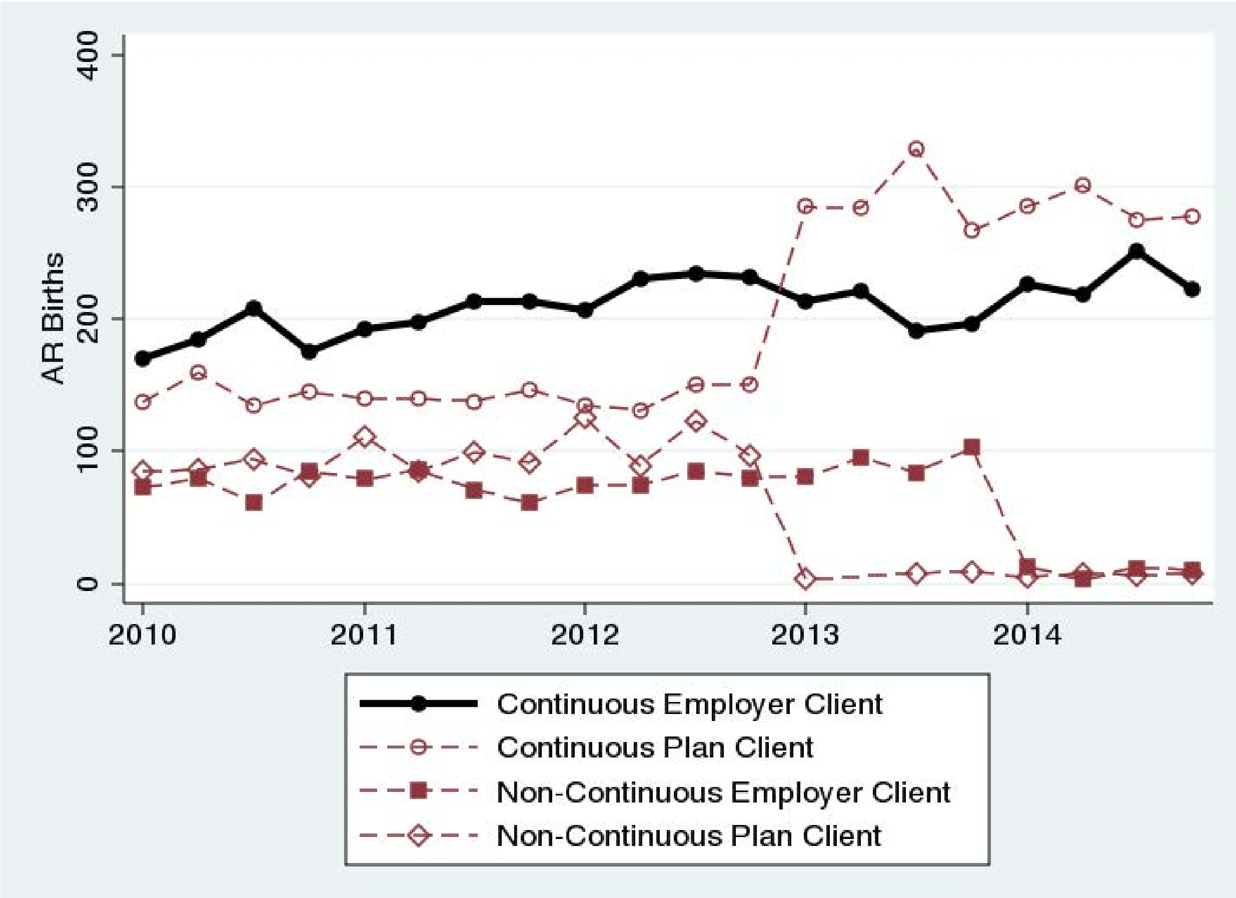
Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, and state and MSA fixed effects, unless otherwise noted. Spending variables are modeled using a one part GLM with a log link and a gamma distribution unless otherwise noted. Intrapartum refers to the entire childbirth hospitalization. In columns 1 and 2, we drop plan type controls and MSA fixed effects, respectively. In columns 3 and 4, we estimate the effect of EBP on untransformed and log transformed spending, respectively. In column 5, we restrict the sample to enrollees that gave birth in a hospital in their state of residence; EBP treatment is defined as residing in Arkansas and giving birth at an Arkansas hospital. * $p<.05$; ** $p<.01$; *** $p<0.001$

Table A3: Effect of EBP by Delivery Type

	Intrapartum Facility Prices	
	(1) Cesarean Section	(2) Vaginal Delivery
Arkansas*2013	-0.0425	0.00151
	(0.0351)	(0.0472)
Arkansas*2014	-0.0711***	-0.0572*
	(0.0186)	(0.0230)
N	15766	24706

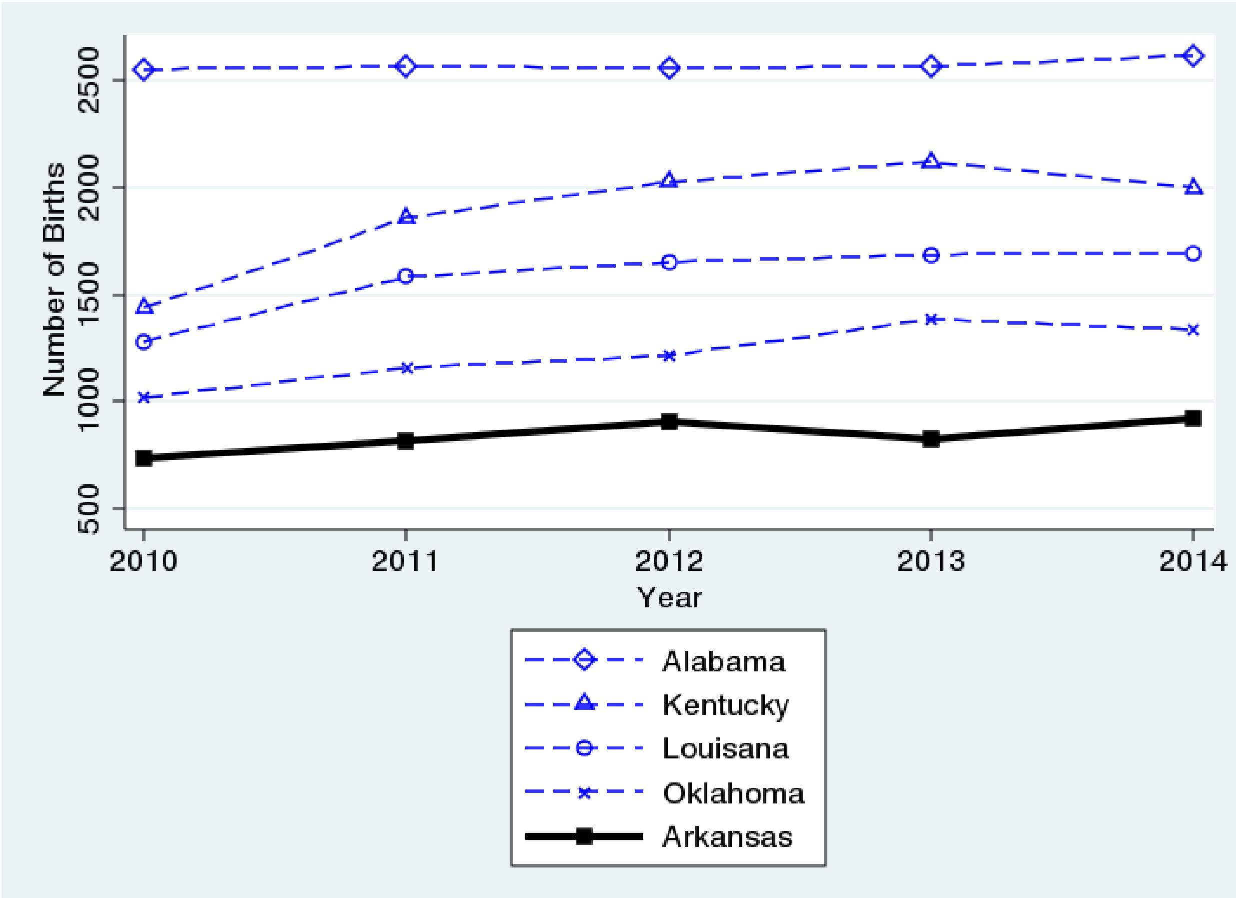
Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates include maternal characteristics, plan characteristics, state and MSA fixed effects and DRG fixed effects. Inpatient facility prices are modeled using a one part GLM with a log link and a gamma distribution. Intrapartum refers to the entire childbirth hospitalization. In columns 1 and 2, we restrict the sample to births that were delivered by cesarean section and vaginally, respectively. * p<.05; ** p<.01; *** p<0.001.

Figure A1: Arkansas Birth Counts by Truven Client Type and Contribution History



Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. The opaque circles indicate births covered by employer clients that continuously provided data to Truven throughout the study period. The hollow circles indicate continuous plan clients. The opaque squares indicate non-continuous employer clients and the hollow diamonds indicate non-continuous plan clients. Our main analytic sample is restricted to data from continuous employer clients.

Figure A2: Final Birth Counts by State: Continuous Employer Contributors Only



Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014.