# 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 $\alpha$. 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 $\Delta$. 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 $\Delta$ (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 $\pi_{dd}^{FFS}$ 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)<br>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 |

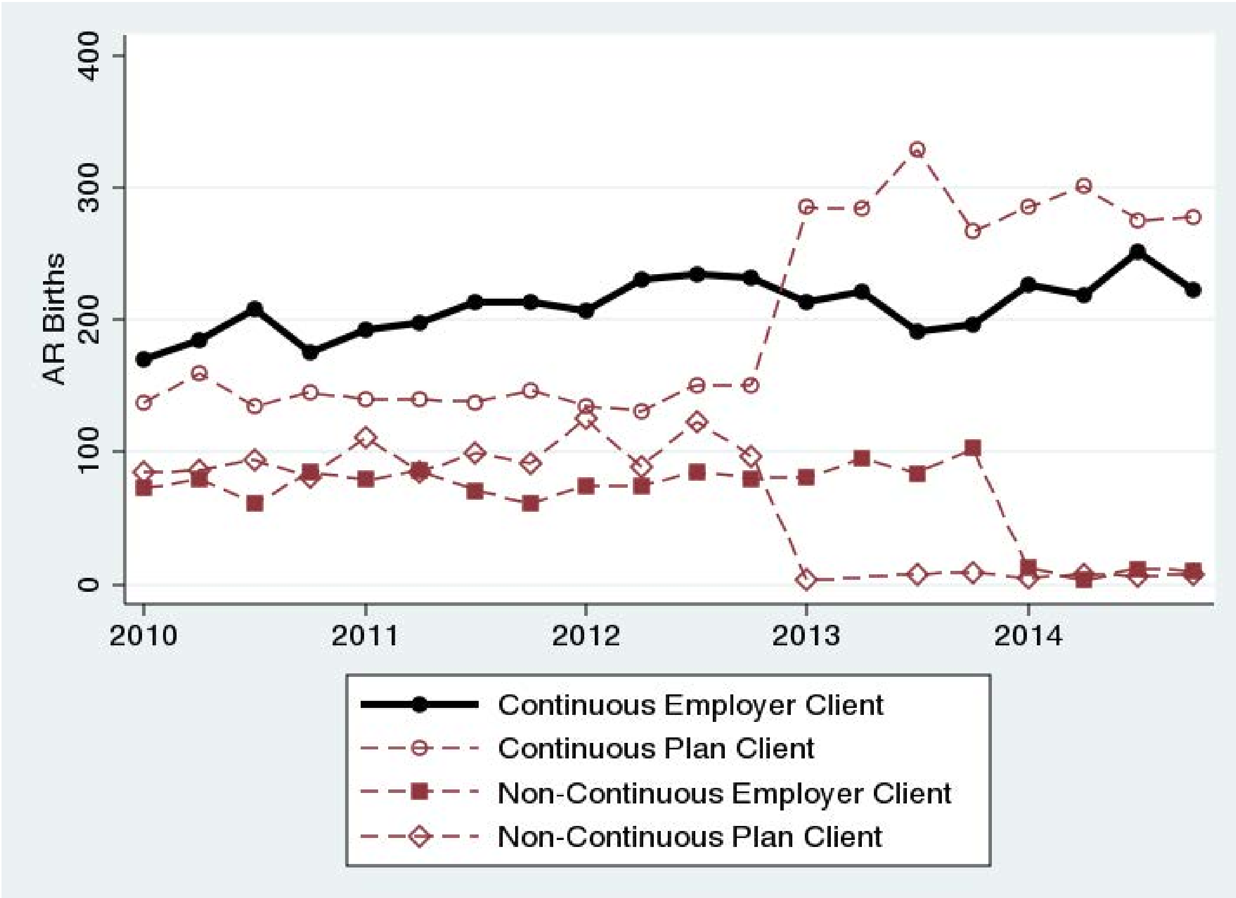
Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates 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)<br>Cesarean Section | (2)<br>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 |

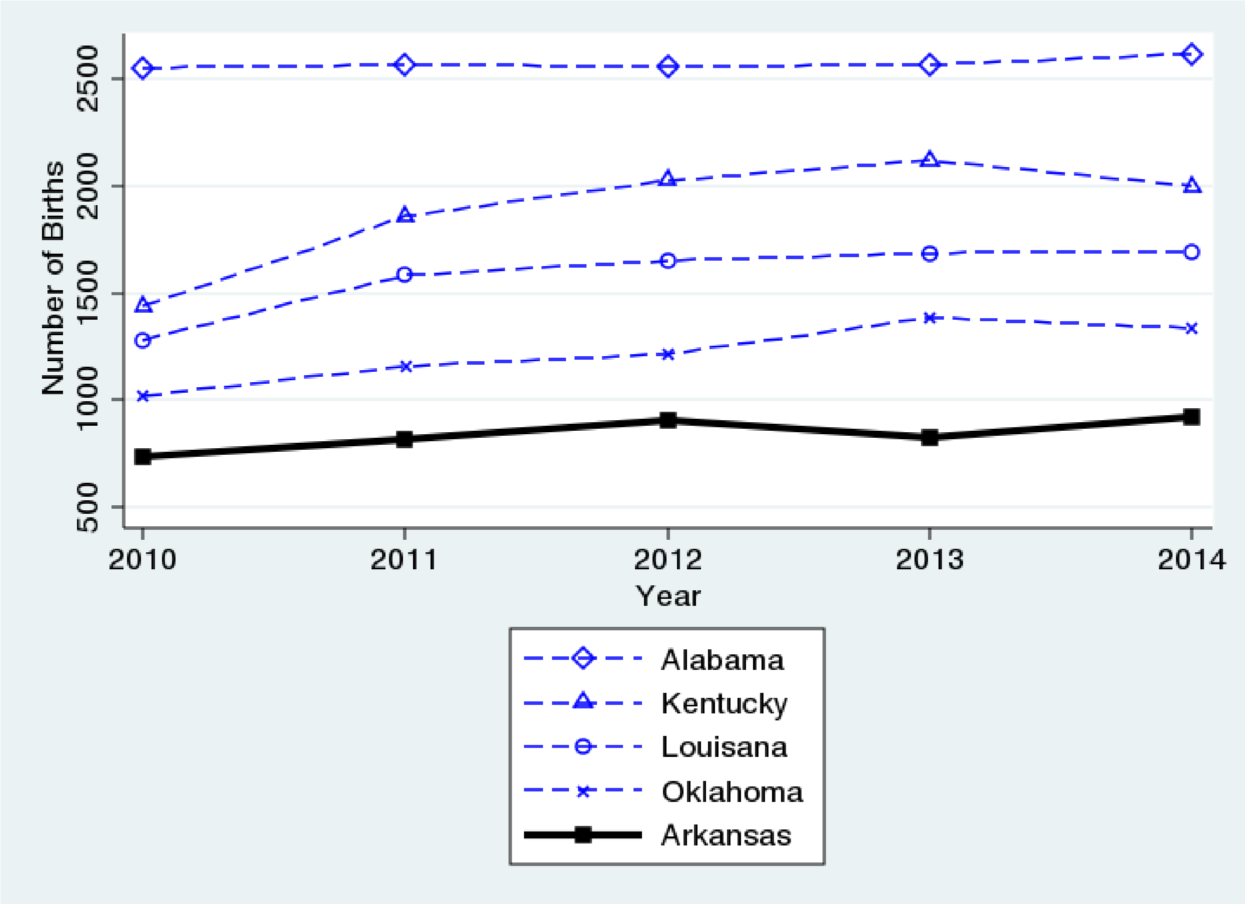
Notes: Sample estimates from Truven claims, using data from perinatal episodes with births between 2010 and 2014. Table cells include DD coefficients with standard errors in parentheses. Standard errors are clustered on MSA. Control states include Alabama, Kentucky, Louisiana and Oklahoma. Covariates 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.