Economics Bulletin

Volume 46, Issue 3

Household composition and infant baby food expenditure around the 2022 federal Medicaid postpartum coverage extension: A triple-difference from the Consumer Expenditure Survey

Gokhan Kumpas
California State University, Los Angeles

Abstract

Beginning in 2022, a federal option let states extend Medicaid postpartum coverage from sixty days to a full year, lowering out-of-pocket medical exposure for new mothers. Where does the freed budget surface? Using the Consumer Expenditure Survey Diary (2013–2024) and a triple-difference comparing households with and without an infant across Section 9812 adopters and comparison states that adopt only later, I find that the infant–non-infant baby food gap widens after adoption by $3.51 in households with two or more children and $3.76 in larger households, both significant at the 5 percent level and stable across falsification, weighting, and sample-composition checks. The contribution is a household-weighted descriptive correlation and its location: the association is estimated precisely in the cells that purchase baby food most often, a pattern consistent with a budget-relaxation reading but equally consistent with the greater statistical power of those cells, and cell-level staggered estimators, once a formula-shortage base-quarter artifact is corrected, bound any state-aggregate association near zero.

This study uses publicly available Consumer Expenditure Survey Diary microdata from the U.S. Bureau of Labor Statistics and required no institutional review board approval. I received no external funding and declare no conflicts of interest. Replication package is available on request. All errors are my own.

Citation: Gokhan Kumpas, (2026) "Household composition and infant baby food expenditure around the 2022 federal Medicaid postpartum coverage extension: A triple-difference from the Consumer Expenditure Survey", *Economics Bulletin*, Volume 46, Issue 3, pages 1025-1038
Contact: Gokhan Kumpas - gkumpas@calstatela.edu.
Submitted: June 02, 2026. **Published:** September 30, 2026.

1. Introduction

Section 9812 of the American Rescue Plan Act of 2021 let states extend Medicaid postpartum coverage from sixty days to twelve months at the standard federal-state cost-share; by the end of 2022-Q3, 34 jurisdictions had adopted under the single federal authorization, and nearly all states followed by 2024. A growing literature documents the policy's health and labor margins: extended postpartum coverage raised treatment for perinatal mood and anxiety disorders (Gordon et al., 2024), opioid use disorder treatment (Gupta et al., 2025), and maternal labor supply (Ejughemre et al., 2026). The margin this paper takes up is more basic for the economics of the household: how the extension reaches the infant household's *budget*. The motivating channel is a budget constraint. A coverage extension lowers out-of-pocket exposure to postpartum and infant medical costs, and Medicaid is known to shield household finances (Gross and Notowidigdo, 2011); the freed resources are reallocated across competing needs, and a budget-relaxing transfer raises food expenditure in particular (Hoynes and Schanzenbach, 2009; Levy et al., 2019). Baby food is the natural place to look: it is the household's primary infant-specific food outlay and is separately recorded in the Diary, and a budget-relaxation reading suggests looking hardest where several children share the food budget (Deaton and Paxson, 1998). That composition cut motivates the analysis, but it cannot by itself distinguish a budget mechanism from statistical power (Section 4 makes the power arithmetic explicit), so I treat it as a descriptive lens rather than a test.

This letter documents the pattern in the Consumer Expenditure Survey (CEX) Diary 2013–2024. In a triple-difference comparison of households with and without an infant across wave-1 adopting states and comparison states untreated during the wave-1 window, the infant–non-infant baby food gap widens after adoption by \$3.51 in two-or-more-child households (SE 1.11) and \$3.76 in larger households (SE 1.30), both significant at the 5 percent level or better and robust to dropping the CEX sampling weights (\$2.59 unweighted, also significant). The contrast is one of *precision*, not magnitude: the point estimates are similar across composition cuts (\$2.94 in single-child versus \$3.51 in two-or-more-child households; \$4.64 in smaller versus \$3.76 in larger households), and a formal test cannot reject equality across cells; what differs is where the estimate is precise enough to exclude zero, which it is in the multi-child and larger-household cells. Two features bound the claim. The exercise is descriptive (the object is the conditional-mean difference across groups, not a causal treatment effect, and I use associational language throughout). And the Diary records no income, while the precision pattern is equally consistent with the larger, more variable baby food purchasing of multi-child households, so the contribution is a household-weighted descriptive correlation and its location, not evidence of a budget mechanism.

2. Data and design

The data are the Diary file of the BLS Consumer Expenditure Survey Public Use Microdata for 2013-Q1 through 2024-Q4 (102,933 household-diary periods in the comparison sample), each recording all expenditures during a two-week window. The outcome is baby food expenditure (Universal Classification Code 180620) per two-week period: jarred and processed baby food and infant cereal. Infant formula is recorded under a separate code and never

enters the outcome; the 2022 formula-shortage quarters are nonetheless handled with an exclusion check below, because the shortage could shift spending toward baby food as a substitute even though formula itself is not in the outcome. Among infant households reporting any baby food the mean is $31.93, the conditional-positive baseline against which the dollar estimates are scaled. A household has an infant if any member is aged two or younger; this proxy is coarser than the policy's mother-in-postpartum-window margin, so it attenuates any true signal toward zero and the estimates read as lower bounds. State adoption dates come from the CMS state-by-state tracker. The comparison sample contrasts the 2022-Q3 (wave-1) adopters with three smaller states whose coverage begins only at 2024-Q1 (Alaska, Delaware, Idaho; Idaho appears in the estimation sample only through 2014, so the comparison group is effectively Alaska and Delaware in later years), untreated throughout the identifying window; wave-2 states and the two large 2024-Q1 adopters (Nebraska, Texas) are set aside so the household-weighted contrast is not dominated by one large state. The estimation sample spans 34 states (31 wave-1 adopters with Diary coverage and three comparison states), the 34 clusters used for inference. Because the comparison states' own coverage begins at 2024-Q1, quarters from 2024 onward contribute no cross-state contrast (every sampled state is covered): they identify only a common before-after shift in the infant/non-infant gap, so the pooled coefficient mixes the clean 2022-Q3–2023-Q4 contrast with this weaker variation. Table III reports a clean-window estimate ending 2023-Q4, which is larger. The staggered national rollout of the Section 9812 option is shown in Figure 1.

The object is the triple-difference in conditional means (Olden and Møen, 2022). For household h in state s and quarter t,

$$y_{hst} = \beta \left(\text{Adopter}_{st} \times \text{HasInfant}_{ht} \right) + \mu_{st} + \mu_{\text{infant}} + \varepsilon_{hst}, \tag{1}$$

where Adopter_{st} equals one once state s has adopted, HasInfant_{ht} equals one for infant households, μ_{st} is a state-by-quarter fixed effect absorbing every state-level shock common to all households in a quarter, and μ_{infant} is a has-infant fixed effect.[1] The coefficient β summarizes how the infant-versus-non-infant baby food gap differs between adopting and comparison states; to keep the comparison clean, equation (1) additionally nets out has-infant-specific differences associated with six other contemporaneous state family policies (diaper and menstrual-product sales-tax exemptions, refundable EITC expansions, state child tax credits, paid family leave, and abortion-trigger laws) and the state unemployment rate. Standard errors are clustered by state; with only 34 clusters, asymptotic cluster-robust inference is unreliable, so the significance stars below come from a wild cluster bootstrap with Webb weights (Cameron et al., 2008; MacKinnon and Webb, 2017) (9,999 replications). For reference, Table I summarizes the estimation sample: infant households are 6.8 percent of household-diary periods and far more likely to report any baby food in the two-week

[1] Equation (1) retains the state-by-quarter and has-infant fixed effects but not the quarter-by-has-infant and state-by-has-infant interactions of a fully saturated triple-difference (Olden and Møen, 2022). Because 31 of the 34 sample states adopt and the untreated contrast is effectively two states, a quarter-by-has-infant effect absorbs the national infant baby food time path that the treatment response itself dominates, so the saturated specification is close to under-identified: the point estimate is stable to larger ($4.43 in the full sample, $8.86 in multi-child households) but its wild-bootstrap interval no longer excludes zero. I therefore report the reduced form, whose identifying variation the placebo-date, formula-shortage, and leave-one-comparison-state-out checks bound directly.

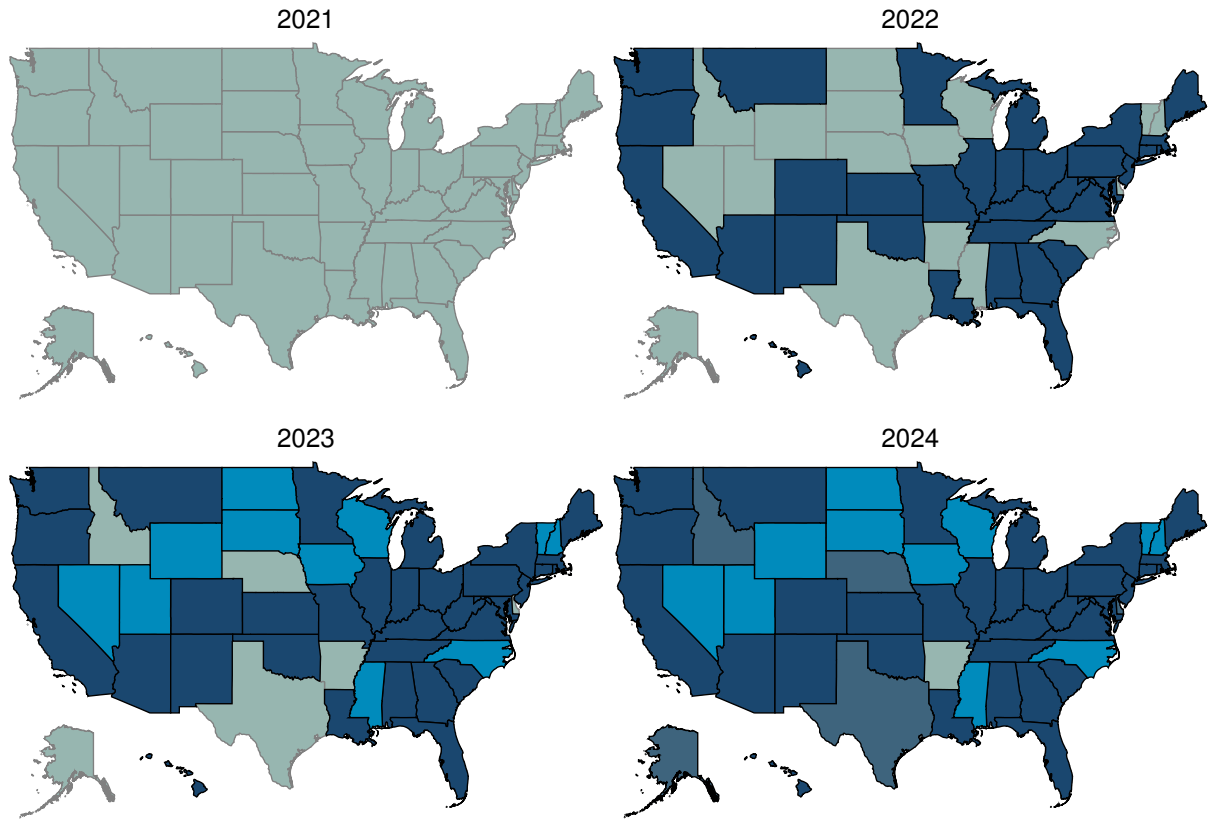


Figure 1: Staggered state adoption of the Section 9812 twelve-month postpartum Medicaid coverage option, 2021–2024

Notes: Panels shade states by the year their twelve-month postpartum coverage took effect (legend at base of figure). The triple-difference in this note compares wave-1 adopters (navy; effective 2022-Q3) with the three smaller states adopting only at 2024-Q1 (Alaska, Delaware, Idaho), untreated throughout the identifying window; wave-2 states and the two large 2024-Q1 adopters (Nebraska, Texas) are set aside so the household-weighted contrast is not dominated by one large state. Adoption dates are from the CMS state-by-state tracker.

window (16.0 versus 1.8 percent, weighted) and larger (mean family size 4.3 versus 2.3).

Table I: Descriptive statistics, by infant status

	Non-infant households	Infant households
Diary-period observations	96,446	6,487
Share of sample (%, weighted)	93.2	6.8
Reports any baby food (%, weighted)	1.8	16.0
Mean baby food expenditure ($)	0.24	5.12
conditional on any purchase ($)	12.99	31.93
Mean family size	2.29	4.27
Mean children under 18	0.43	2.09

Consumer Expenditure Survey Diary, 2013–2024, for the estimation sample of wave-1 adopter and comparison states (102,933 household-diary periods). Means weighted by the CEX final weight; observation counts unweighted. "Reports any baby food" is the share of periods with positive baby food expenditure in the two-week diary window; the conditional-on-purchase row is computed over the 2,820 such periods.

3. Results

Infant households account for essentially all baby food purchasing in the panel (Figure 2): weighted mean expenditure for households with an infant fluctuates between roughly $3 and $9 per two-week period, against a near-zero, flat series for households without one.

Table II and Figure 3 report equation (1) on subsamples defined by household composition. In the pooled sample the coefficient is $3.97 per two-week diary period (SE 1.43), about 12 percent of the $31.93 conditional-positive baseline; I report this figure but do not lean on it, because it is a household-weighted summary that state-equal-weighted estimators do not reproduce (below). Splitting by the number of children under 18, the estimate is $2.94 in single-child households (SE 3.39, confidence interval spanning zero) and $3.51 in households with two or more children (SE 1.11, significant at 1 percent). Splitting by family size gives $4.64 in smaller households (size three or fewer; SE 3.08, spanning zero) and $3.76 in larger households (size four or more; SE 1.30, significant at 5 percent). The pattern is one of precision, not magnitude. The point estimates are similar across cuts, and indeed the largest, $4.64, is in the imprecise smaller-household cell; interacting the adoption-by-infant term with the composition indicator in a single regression (so the cell coefficients reproduce the subsample estimates), the multi-child-minus-single-child difference is $0.57 and the larger-minus-smaller difference is −$0.89, and in each case the bootstrap cannot reject equality. What differs across cells is whether the association is estimated precisely enough to exclude zero, which it is in the multi-child and larger-household cells; Section 4 explains why this location, on its own, cannot be read as evidence of a budget mechanism.

Robustness. The household-level contrast is stable across a compact battery of checks (Table III, Panel A). Dropping the CEX sampling weights leaves $2.41 (full) and $2.59 (two

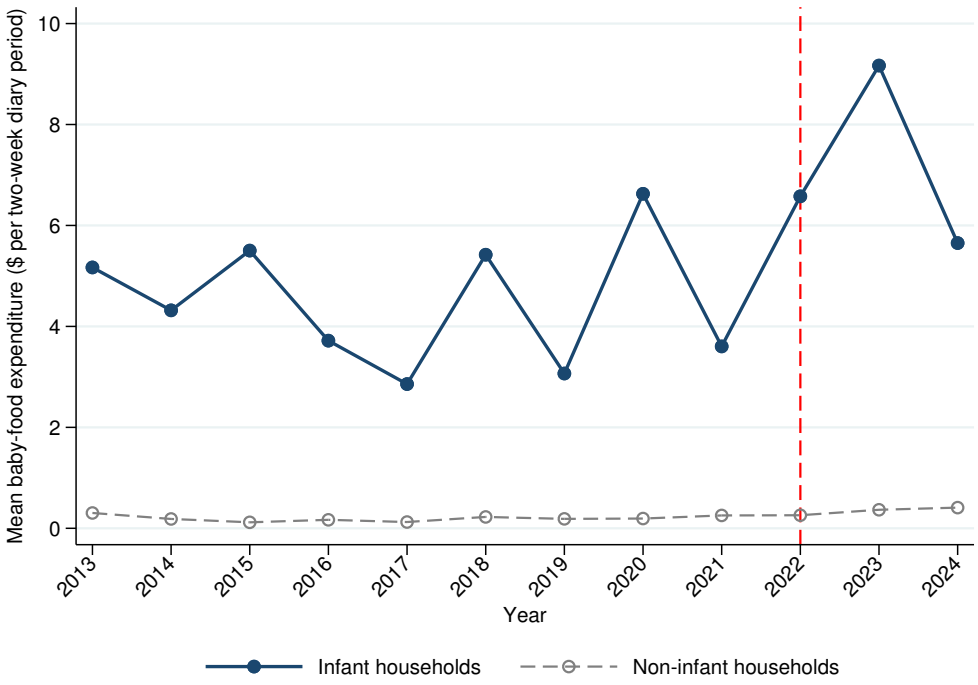


Figure 2: Mean baby food expenditure by infant status, 2013–2024

Notes: Weighted mean baby food expenditure per two-week diary period, by household infant status, for the estimation sample. Households with an infant (solid) account for essentially all baby food purchasing; the series for households without an infant (dashed) is flat and near zero. The vertical line marks the 2022-Q3 wave-1 adoption.

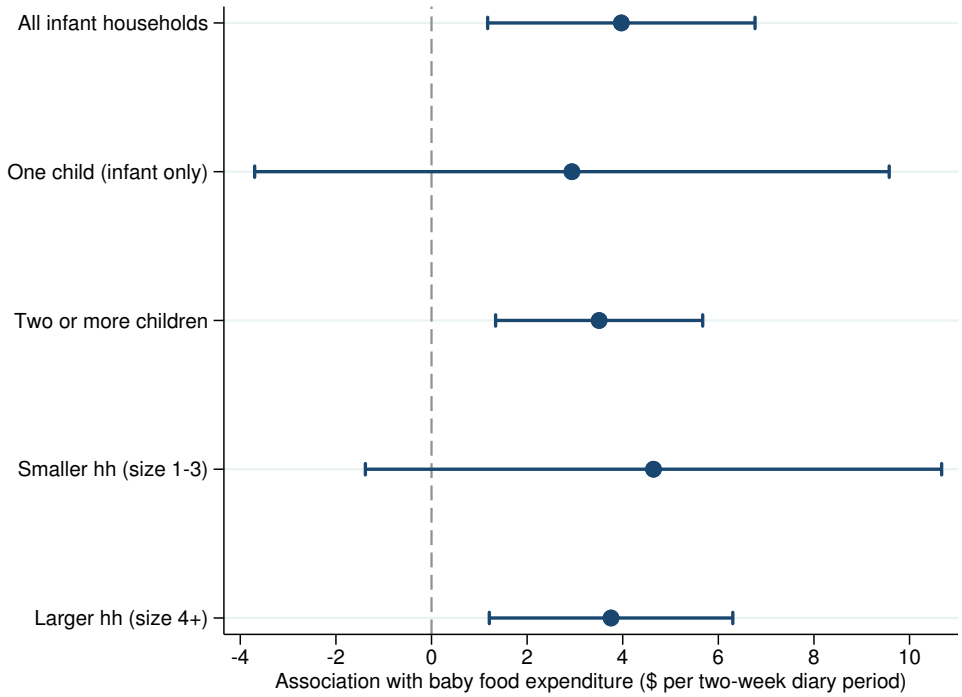


Figure 3: Triple-difference association between Section 9812 adoption and infant-household baby food expenditure, by household composition

Notes: Points are subsample estimates of β from equation (1); horizontal bars are 95 percent confidence intervals from analytic state-clustered standard errors (Table II reports wild-cluster-bootstrap significance for the same estimates). The confidence interval excludes zero in households with two or more children and in larger households, and spans zero in single-child and smaller households. The point estimates are similar across cuts; the contrast is in precision (a test cannot reject equality across cells), not in magnitude.

Table II: Baby food expenditure association by household composition

Subsample	Estimate (SE)	N
All infant households	3.97*** (1.43)	102,933
By number of children under 18		
One child (the infant only)	2.94 (3.39)	12,498
Two or more children	3.51*** (1.11)	15,568
By family size		
Smaller households (size 1–3)	4.64 (3.08)	82,405
Larger households (size 4+)	3.76** (1.30)	20,528

All rows estimate equation (1) on the indicated subsample, with state-by-quarter and has-infant fixed effects and the six-policy plus unemployment-rate controls described in the text. Standard errors clustered by state in parentheses. Significance from a wild cluster bootstrap (Webb weights, 9,999 replications): *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The children-count rows restrict to households with exactly one child under 18 and with two or more, respectively, infant and non-infant alike, so each cell retains its own non-infant comparison group (in an infant household with one child, that child is the infant); households with no children under 18 enter neither cell, so the rows need not sum to the full-sample row. The family-size split partitions all 102,933 observations at family size four.

or more children), the same sign and order of magnitude and still significant, so the contrast does not ride on the weights. Nor is it carried by the largest states: dropping California, New York, Illinois, and Florida *raises* the estimate to \$5.13 (SE 1.76), and separating the seven most populous adopters (California, Florida, New York, Pennsylvania, Ohio, Illinois, Virginia) from the remaining 24 gives \$3.42 (SE 2.01) among the seven and \$4.81 (SE 1.78) among the rest, so the association is present in both groups and, if anything, larger among the remaining 24, which is why excluding those four states raises rather than lowers the pooled figure. Nor does it hinge on the thin comparison group: dropping Alaska, Delaware, or Idaho one at a time leaves the estimate at \$3.98 to \$4.05 in the full sample and \$3.51 to \$3.53 in multi-child households, each significant at 5 percent or better, so no single comparison state pins the result. Restricting to the clean contrast window (2013-Q1 through 2023-Q4; from 2024-Q1 the comparison states are themselves covered) strengthens the estimate to \$4.80 (SE 1.86) and \$4.18 (SE 1.46); the 2024 quarters carry a smaller coefficient (Table IV, Panel B), so the full-window figures read as conservative relative to the clean contrast. Falsification checks pass: a placebo 2018-Q3 adoption date returns \$0.85 (full) and \$0.55 (two or more children), neither distinguishable from zero; a placebo outcome (adult food-away-from-home) does not move significantly in either sample; dropping the acute 2022 infant-formula-shortage quarters leaves the estimate unchanged or slightly larger (\$4.68 and \$3.76); and dropping the COVID years 2020–2021 leaves \$3.42 and \$2.48, the multi-child figure attenuating modestly but keeping its sign.

Because baby food is recorded in only 16 percent of infant-household diary periods, the

dollar estimate mixes a participation margin with a conditional-spending margin. Table IV, Panel A, examines the two margins separately: the any-purchase linear-probability version of equation (1) gives 1.5 percentage points (SE 1.7) in the full sample and 1.1 (SE 3.1) in the two-or-more-children cell, and the conditional-on-purchase version gives \$4.41 (SE 10.40) and \$2.64 (SE 14.93) on the 2,820 purchaser periods (1,013 in the two-or-more-children cell). Both margins point positive, but neither is separately distinguishable from zero at this sample size, and the two margin estimates combined at the sample participation rate and conditional mean account for only about \$1.17 of the \$3.97 total, a further sign that the margin-specific regressions are too noisy to apportion the association. The unconditional dollar coefficient is the estimable summary, and the margin split is reported for transparency rather than leaned on. The timing of the association matches the design's contrast (Table IV, Panel B): replacing the single post-adoption term with calendar-year interactions gives small, insignificant coefficients in the two 2022 adoption quarters, \$6.07 (SE 1.80, full) and \$6.33 (SE 1.71, two or more children) in 2023 (the one complete year in which adopters are covered and the comparison states are not), and attenuated values in 2024, when the comparison states' own coverage begins and the remaining variation is a common before-after shift rather than a cross-state contrast. Equality across years cannot be rejected in the full sample ($p = 0.23$) but is rejected in the two-or-more-children cell ($p = 0.004$): the association there is concentrated in 2023. That concentration is what the contrast structure predicts (partial-year exposure in 2022, no untreated comparison in 2024), but it also means the pooled \$3.51 averages over significant year-to-year variation.

A separate question is how the association behaves when each state-quarter cell, rather than each household, is the unit. Three heterogeneity-robust staggered estimators on the state-quarter cell panel (Callaway–Sant'Anna, Callaway and Sant'Anna, 2021; Borusyak–Jaravel–Spiess, Borusyak et al., 2024; and de Chaisemartin–D'Haultfœuille, de Chaisemartin and D'Haultfœuille, 2020; the cell panel uses the full staggered rollout across all adoption cohorts) return small, insignificant estimates of mixed sign (Table III, Panel B: −\$3.35, \$1.49, −\$3.91), and re-weighting the same estimators by each cell's infant-household mass does not recover the household-weighted figure (Callaway–Sant'Anna −\$6.45, SE 3.79; Borusyak–Jaravel–Spiess \$0.21, SE 1.48), so the divergence is not purely a weighting phenomenon. Table III, Panel C, locates its sources. First, the base quarter: Callaway–Sant'Anna and de Chaisemartin–D'Haultfœuille difference post-adoption quarters against the quarter just before adoption, which for the dominant wave-1 cohort is 2022-Q2, the trough of the acute infant-formula shortage, when the sparse cell-level baby food gaps were anomalous; the imputation estimator and TWFE instead lean on the entire pre-period. Excluding 2022-Q1–Q2 (so the base becomes 2021-Q4; the exercise changes the pre-sample along with the base, and each estimator moves within its own confidence band) brings the three equal-cell estimators into agreement around zero (\$0.69, \$0.70, −\$0.54; SEs 2.1–2.4): the sign disagreement was an artifact of the base quarter, and the resulting intervals contain the full-window household-weighted estimates, though the de Chaisemartin–D'Haultfœuille bound is marginal and the clean-window estimates sit at or just above the Borusyak–Jaravel–Spiess and de Chaisemartin–D'Haultfœuille upper bounds. The infant-household-weighted Callaway–Sant'Anna estimate also moves two-thirds of the way to zero (−\$2.16, SE 2.24, from −\$6.45) but keeps a negative point value with an interval that excludes the household-weighted figures. Second, the weighting unit within the TWFE family: on the same cell panel

over the clean window, TWFE weighted by each cell's infant-household mass gives \$3.32 (SE 1.58), close to the household-level estimate, while equal cell weights give \$0.90 (SE 1.39). The mechanics are the outcome's sparseness: the median state-quarter cell contains four sampled infant households (26 percent of cells contain none), so a small state's quarterly infant–non-infant gap swings wildly (the comparison states' gap moves from +\$32.98 in 2021-Q4 to −\$0.46 in 2022-Q2, on cells of six and two infant households), and equal weighting lets those high-variance cells dominate (Goodman-Bacon, 2021). Two conclusions follow. Cell-level inference is weak in every variant: at the cell level the treatment varies only across states, with three comparison clusters, and wild cluster bootstrap p-values for the two cell TWFE rows are 0.51 and 0.59, so the cell-level rows serve the weighting and base-period comparisons rather than inference. And the cell-level object is not the household-level object: cell aggregation discards the within-state-quarter infant/non-infant household contrast, the variation the household-level design's estimate and inference actually use, and replaces it with a few dozen noisy cell means, without the household controls or the CEX household weights, which is why even the corrected and re-weighted cell estimators sit below the household-level figures. I therefore read the cell level as bounding any state-aggregate association near zero while neither confirming nor contradicting the household-level association, and nothing in it indicates treatment-timing heterogeneity (the imputation estimator, which like TWFE uses the full pre-period, is small and non-negative throughout). The Callaway–Sant'Anna event study (Figure 4) shows a flat pre-trend (average pre-adoption estimate −\$0.82, leads jointly and individually indistinguishable from zero), with an imprecise post-adoption series whose relative-magnitude bounds (Rambachan and Roth, 2023) do not support a sign-definite state-equal-weighted association in either direction.

4. Discussion

A budget-relaxation reading motivates where this letter looks. The first link, that Medicaid coverage shields household finances, is well established (Gross and Notowidigdo, 2011; Levy et al., 2019); the second, that a budget-relaxing transfer is reallocated toward food, mirrors the in-kind-transfer literature (Hoynes and Schanzenbach, 2009). But the location of the precision is not evidence for that mechanism, and this letter does not read it as such. The power arithmetic is decisive: under a common true association equal to the multi-child estimate (\$3.51), and treating the cluster-robust standard errors as known, the single-child cell's standard error of 3.39 implies statistical power of roughly 18 percent at the 5 percent level, against roughly 88 percent in the multi-child cell (SE 1.11); the corresponding 80-percent-power minimum detectable effects are about \$9.50 and \$3.10. A precision pattern concentrated in the multi-child and larger-household cells is therefore exactly what power alone would produce under a homogeneous association, since those cells buy baby food more often and more variably (Deaton and Paxson, 1998), so the composition pattern cannot discriminate between a budget-relaxation channel and mechanical power. The Diary also records no income, so the income gradient the eligibility rules imply cannot be examined. The contribution is accordingly a household-weighted descriptive correlation and its location, for a causal literature with linked income data to interrogate.

The second discipline is the relation between the household-weighted and state-equal-

Table III: Robustness of the baby food association

	Full sample	Two or more children
Panel A. Household-weighted triple-difference		
Baseline (CEX sampling weights)	3.97*** (1.43)	3.51*** (1.11)
Unweighted	2.41** (1.11)	2.59** (1.25)
Drop California, New York, Illinois, Florida	5.13** (1.76)	3.39* (1.73)
Seven most populous adopters only	3.42* (2.01)	2.98* (1.02)
Remaining 24 adopters only	4.81** (1.78)	3.45 (2.00)
Clean contrast window (through 2023-Q4)	4.80*** (1.86)	4.18** (1.46)
Exclude COVID years 2020–2021	3.42** (1.36)	2.48* (1.35)
Exclude formula-shortage quarters (2022-Q1–Q2)	4.68*** (1.39)	3.76*** (1.20)
Placebo adoption date (2018-Q3)	0.85 (1.02)	0.55 (1.44)
Placebo outcome (adult food away from home)	4.19 (4.89)	−11.84 (8.34)
Panel B. State-equal-weighted estimators (full sample)		
Callaway–Sant'Anna (not-yet-treated)	−3.35 (4.78)	
Borusyak–Jaravel–Spiess	1.49 (1.83)	
de Chaisemartin–D'Haultfœuille	−3.91 (6.30)	
Callaway–Sant'Anna, infant-household-weighted	−6.45 (3.79)	
Borusyak–Jaravel–Spiess, infant-household-weighted	0.21 (1.48)	
Panel C. Reconciliation on the state-quarter cell panel		
Callaway–Sant'Anna, excl. 2022-Q1–Q2	0.69 (2.31)	
Borusyak–Jaravel–Spiess, excl. 2022-Q1–Q2	0.70 (2.06)	
de Chaisemartin–D'Haultfœuille, excl. 2022-Q1–Q2	−0.54 (2.37)	
Callaway–Sant'Anna, infant-hh-weighted, excl. 2022-Q1–Q2	−2.16 (2.24)	
Cell TWFE, infant-household-weighted (through 2023-Q4)	3.32 (1.58)	
Cell TWFE, equal cell weights (through 2023-Q4)	0.90 (1.39)	

Each Panel A cell is a separate estimate of β from equation (1) on the indicated sample, with state-clustered standard errors in parentheses and significance from a wild cluster bootstrap (Webb weights, 9,999 replications): *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The adopter-split rows separate the seven most populous wave-1 adopters (by summed CEX weight) from the remaining 24, keeping the comparison states in both; the clean-window row ends the sample at 2023-Q4, after which the comparison states are themselves covered. The placebo-date row reassigns adoption to 2018-Q3 on the pre-2022 sample; the placebo-outcome row replaces baby food with adult food away from home. Panels B and C report estimators on the state-quarter cell panel (infant minus non-infant mean baby food by state-quarter); the staggered-estimator rows use the full rollout across all adoption cohorts, while the two cell TWFE rows use the wave-1-versus-comparison design through 2023-Q4; analytic standard errors, no significance stars. The de Chaisemartin–D'Haultfœuille figures average the first four post-adoption quarters; the Callaway–Sant'Anna and Borusyak–Jaravel–Spiess figures average all available post-adoption periods. "Infant-household-weighted" weights each cell by its sampled infant-household mass. The four base-period reconciliation rows exclude the acute formula-shortage quarters 2022-Q1–Q2 (the two cell TWFE rows use the full clean window through 2023-Q4); for the wave-1 cohort the quarter just before adoption, 2022-Q2, is the base period in the short-base estimators, so the exclusion moves the base to 2021-Q4 (and shortens the pre-sample). Cell-level rows serve the weighting and base-period comparisons, not inference: at the cell level the treatment varies only across states (three comparison clusters), and wild cluster bootstrap p-values for the two cell TWFE rows are 0.51 and 0.59; the household-level design in Panel A instead identifies the coefficient within state-quarter from the infant/non-infant contrast, which is where its inferential power comes from.

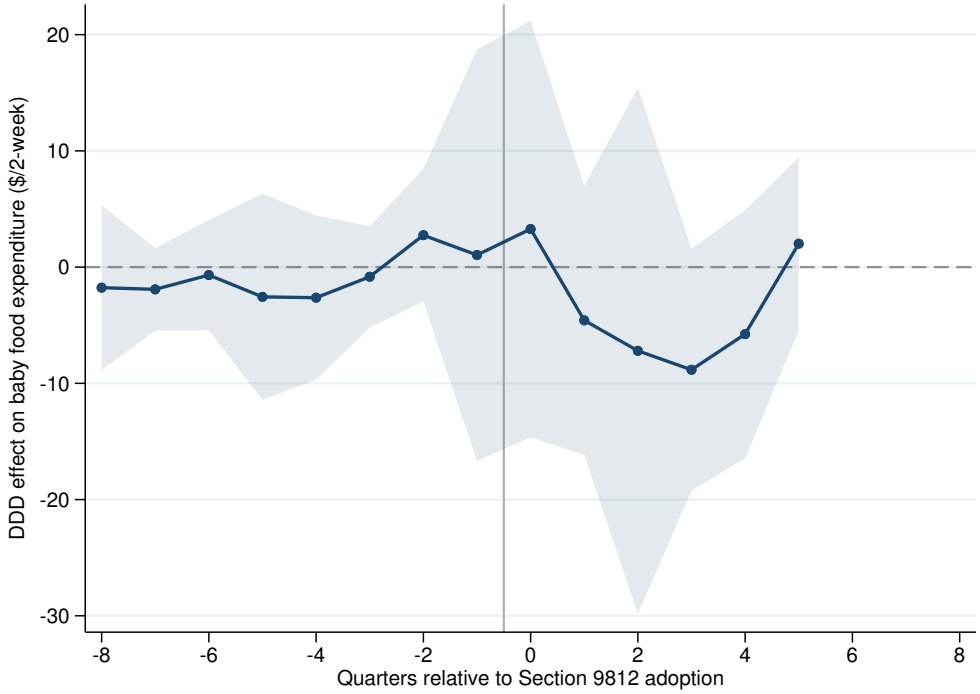


Figure 4: Event-study estimates of the Section 9812 triple-difference on infant-household baby food expenditure, by quarter relative to adoption

Notes: Callaway–Sant'Anna event-study coefficients on the state-quarter cell panel (state-equal weighting), with 95 percent confidence intervals; the vertical line marks adoption. The pre-adoption leads are flat and jointly indistinguishable from zero (average −$0.82), consistent with parallel trends. The post-adoption series is imprecise and not sign-definite; its negative tilt reflects the estimator's base quarter, 2022-Q2, the trough of the acute formula shortage (Table III, Panel C: excluding 2022-Q1–Q2 moves the average post-adoption estimate to $0.69, SE 2.31).

Table IV: Margins and timing of the household-weighted association

	Full sample	Two or more children
Panel A. Margins		
Any purchase (percentage points)	1.45 (1.68)	1.06 (3.08)
Expenditure conditional on purchase ($)	4.41 (10.40)	2.64 (14.93)
Panel B. Calendar-year profile ($)		
2022 (adoption quarters Q3–Q4)	2.93 (2.84)	0.38 (1.42)
2023 (adopters covered, comparison not)	6.07*** (1.80)	6.33*** (1.71)
2024 (comparison states also covered)	2.54 (1.81)	2.81* (1.51)
Equality across years (p)	0.23	0.004

Panel A re-estimates equation (1) replacing the dollar outcome with an any-purchase indicator (linear probability, coefficients in percentage points) and, separately, restricting to the diary periods with a positive purchase (2,820 in the full sample, 1,013 in the two-or-more-children cell; conditional dollars; the purchaser sample is selected and thin, so this margin is imprecise by construction). Panel B replaces the single post-adoption term with calendar-year interactions. Standard errors clustered by state in parentheses; significance from a wild cluster bootstrap (Webb weights, 9,999 replications): *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The equality row reports the Wald p-value for equal coefficients across the three years.

weighted estimates, where the diagnostics below settle two specific points with evidence, while a residual difference remains. The apparent disagreement among the staggered estimators (mixed signs, none significant) traces to the base quarter: the short-base estimators difference post-adoption quarters against the wave-1 cohort's just-before-adoption quarter, 2022-Q2, the trough of the acute infant-formula shortage; excluding 2022-Q1–Q2, the three equal-cell estimators agree on a band around zero ($0.69, $0.70, −$0.54; SEs 2.1–2.4) whose intervals contain the full-window household-weighted estimates. Within the TWFE family, the household-versus-cell gap behaves exactly as sparseness predicts: with a median of four sampled infant households per state-quarter cell (and a quarter of cells empty), cell-equal weighting hands the average to a handful of volatile small cells ($0.90), while weighting cells by their infant-household mass restores the household-level point estimate ($3.32). The residual is a level difference: even corrected and re-weighted, the cell-level estimates sit below the household-level figures (the infant-household-weighted Callaway–Sant'Anna remains at −$2.16, SE 2.24, after the base-quarter correction), which is what one expects when the object changes: cell aggregation discards the within-state-quarter infant/non-infant household contrast, the household controls, and the CEX household weights. Cell-level inference is weak in every case (the treatment varies only across states, with three comparison clusters), and that asymmetry is itself the deepest point of the reconciliation: the household-level design draws its power from the within-state-quarter contrast, which every cell-level estimator discards. I therefore read the cell level as bounding any state-aggregate association near zero, neither confirming nor contradicting the household-level association, with nothing in it indicating treatment-timing heterogeneity. Two scope limits complete the picture. Baby food is the Diary's infant-specific, separately recorded food outlay, but it is one claim among many on a relaxed budget: other components of spending may absorb part of any relaxation, and the adult food-away-from-home placebo probes only one alternative, so this letter

documents the association for one salient category rather than tracing incidence across the budget. And the household-level contrast, precise in multi-child and larger households, imprecise elsewhere, robust to the weighting choice, and strongest in the one calendar year with a fully clean contrast, remains a description of where the association sits, not an estimate of a policy effect. Documenting that location around a major coverage extension, and showing exactly why cell-level estimators cannot see it, is a useful descriptive step toward the causal analyses that linked administrative data with income will eventually support.

References

Borusyak, K., X. Jaravel, and J. Spiess (2024) "Revisiting event-study designs: Robust and efficient estimation" *Review of Economic Studies* **91**, 3253–3285.

Callaway, B. and P. H. C. Sant'Anna (2021) "Difference-in-differences with multiple time periods" *Journal of Econometrics* **225**, 200–230.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008) "Bootstrap-based improvements for inference with clustered errors" *Review of Economics and Statistics* **90**, 414–427.

de Chaisemartin, C. and X. D'Haultfœuille (2020) "Two-way fixed effects estimators with heterogeneous treatment effects" *American Economic Review* **110**, 2964–2996.

Deaton, A. and C. Paxson (1998) "Economies of scale, household size, and the demand for food" *Journal of Political Economy* **106**, 897–930.

Ejughemre, U. J., W. Lyu, and G. L. Wehby (2026) "Effects of Medicaid coverage on work: Evidence from extending postpartum Medicaid coverage" *Health Services Research* **61**, e70055.

Goodman-Bacon, A. (2021) "Difference-in-differences with variation in treatment timing" *Journal of Econometrics* **225**, 254–277.

Gordon, S. H., S. Lee, M. W. Steenland, N. Deen, and E. Feinberg (2024) "Extended postpartum Medicaid in Colorado associated with increased treatment for perinatal mood and anxiety disorders" *Health Affairs* **43**, 523–531.

Gross, T. and M. J. Notowidigdo (2011) "Health insurance and the consumer bankruptcy decision: Evidence from expansions of Medicaid" *Journal of Public Economics* **95**, 767–778.

Gupta, S., Y. Ge, J. C. Maclean, and M. D. Eisenberg (2025) "Medicaid coverage and postpartum opioid use disorder treatment" NBER working paper number 34541.

Hoynes, H. W. and D. W. Schanzenbach (2009) "Consumption responses to in-kind transfers: Evidence from the introduction of the Food Stamp Program" *American Economic Journal: Applied Economics* **1**, 109–139.

Levy, H., T. C. Buchmueller, and S. Nikpay (2019) "The impact of Medicaid expansion on household consumption" *Eastern Economic Journal* **45**, 34–57.

MacKinnon, J. G. and M. D. Webb (2017) "Wild bootstrap inference for wildly different cluster sizes" *Journal of Applied Econometrics* **32**, 233–254.

Olden, A. and J. Møen (2022) "The triple difference estimator" *Econometrics Journal* **25**, 531–553.

Rambachan, A. and J. Roth (2023) "A more credible approach to parallel trends" *Review of Economic Studies* **90**, 2555–2591.