Pour Decisions: Alcohol Privatization and Birth Outcomes

David Hall
Department of Economics, University of Oregon

August 24, 2026

Abstract

Alcohol-market liberalization can change prices, retail availability, convenience, and product choice simultaneously. Increased alcohol use can generate substantial social and health costs, particularly when exposure occurs during pregnancy. Does a broad market reform that expands access to alcohol lead to poorer birth outcomes? I study Washington's Initiative 1183, which privatized spirits retailing in June 2012 and expanded entry by grocery and big-box stores. I find that privatization substantially increased very local spirits access, bringing a larger share of the population within one mile of a spirits retailer. NielsenIQ Homescan data also show a 35% increase in recorded household spirits purchases after privatization. Substitution across beer, wine, and spirits, however, reduces the corresponding increase in total ethanol-equivalent purchasing to 9.2%, and the available household evidence suggests that much of the purchasing response occurred outside prime childbearing ages. Using restricted-use vital statistics and a repeated-cross-section synthetic difference-in-differences design, I find adverse point estimates for birth weight and fetal growth, but these estimates are not unusually large relative to fully refitted placebo-state estimates. I also find no large, coherent shift in the observed birth pool. The results show that the consequences of market liberalization depend not only on whether consumers respond, but on which dimensions of access change, substitution across products, and the incidence of the resulting behavioral response.

Keywords: Alcohol policy; infant health; birth weight; liquor privatization; synthetic difference-in-differences
JEL Codes: I12, I18, H75, J13

1 Introduction

Alcohol demand responds to prices, legal restrictions, and retail availability. A large literature documents these responses, while a separate literature shows that alcohol use can generate substantial health and social costs. While a large literature studies restrictions on alcohol access as a means of reducing specific health and social costs, less is known about what "access" means when an entire market is liberalized and how the resulting behavioral response translates into population-level harm. Liberalization can lower the generalized cost of acquiring alcohol through shorter travel, more sellers, broader product assortment, longer hours, and the ability to combine alcohol purchases with other shopping. Understanding its consequences therefore requires knowing not only whether access expands, but how it expands and who responds to that change.

This paper asks two related questions. First, which dimensions of access change when a restricted alcohol market is opened to private retail competition, and how strongly do consumers respond? Second, does that market response translate into measurable changes in birth health? The second question depends on more than aggregate alcohol demand. Consumers can substitute across beer, wine, and spirits, so a large increase in purchases of one beverage need not generate an equally large increase in total ethanol. The health consequences also depend on behavioral incidence: additional alcohol exposure can affect birth health only when the response occurs among people and periods relevant to pregnancy. A large aggregate response can therefore coexist with a much smaller effect on this particular health margin.

I study these questions using Washington's Initiative 1183, which privatized spirits sales in June 2012. Before the reform, Washington directly controlled the retail sale of spirits. I-1183 closed state stores and allowed qualifying grocery, warehouse, and other large retailers to enter the market. Privatization changed prices, retail entry, assortment, travel costs, shopping convenience, and market organization at the same time. This bundled reform makes it possible to study both how consumers respond when several dimensions of market access change together and whether those responses propagate into a population-level health outcome.

I find that I-1183 generated a large spirits-market response, but that response attenuates as the analysis moves from retail access to total ethanol purchasing and then to the population most relevant for birth health. I use pre-I-1183 grocery availability to predict post-privatization changes in several dimensions of local spirits access. The clearest response occurs at very short distances: one-mile population coverage increases substantially in counties with greater pre-policy grocery exposure, while five-mile coverage was already nearly saturated. NielsenIQ Homescan data likewise show a large purchasing response, with recorded household spirits purchases rising by 35 percent after privatization. Substitution across beverage categories reduces the increase in total ethanol-equivalent purchasing to 9.2 percent, as higher spirits and beer purchases are partly offset by lower wine purchases. The household incidence of that response is also uneven. Total ethanol purchasing rises most among households whose female head is age 45 or older, by 35.6 percent, compared with 16.3 percent among those ages 35–44; the estimate for households with a female head under age 35 is negative, although pre-policy fit is poor for that group. The aggregate market response was therefore not concentrated among the households most likely to be relevant for pregnancy, underscoring why both the margin of access and the incidence of the behavioral response matter for interpreting downstream health effects.

I-1183 can affect observed birth outcomes through two primary channels: changes in prenatal alcohol exposure and changes in the composition of the birth pool. I first estimate the overall effect of I-1183 on health among observed live births using restricted-use National Vital Statistics System records. Washington's counterfactual is constructed using a repeated-cross-section synthetic difference-in-differences design, with the primary donor and time weights selected using blinded pre-policy holdouts. Because there is only one treated state, I base primary coefficient-level inference on complete placebo refits in which each non-Washington jurisdiction is assigned treatment and the full design is reestimated (Arkhangelsky et al., 2021). I also evaluate related birth outcomes jointly rather than drawing conclusions from whichever individual coefficient is largest.

The statewide point estimates suggest some movement toward lower fetal growth. Average birth weight falls by 10.28 grams, low birth weight rises by 0.274 percentage points, small-for-gestational-age births rise by 0.357 percentage points, and large-for-gestational-age births fall by 0.396 percentage points. At the same time, every primary sampling interval includes zero, the prespecified fetal-growth and gestational-duration index is near zero, and neither the six-outcome fetal-growth family nor the complete birth-weight distribution is jointly unusual relative to the fully refitted placebo states. I separately examine whether selection into the observed birth pool helps explain these estimates. Pregnancy accounting and maternal composition reveal no large, coherent shift. The adverse point estimates therefore leave open the possibility of modest or concentrated effects, but the evidence does not establish a broad deterioration in birth health following privatization.

Rather than relying only on the statewide reform, my second design exploits cross-county differences in predetermined grocery infrastructure to predict where privatization produced the largest expansion in one-mile population coverage. I then test whether birth outcomes changed more in counties where I-1183 generated larger increases in very local spirits access. That first stage is strong and remains stable under finer geographic measurement and alternative definitions of short-distance access. Birth weight and small-for-gestational-age rates move adversely along the resulting policy-exposure gradient, but the broader outcome family is mixed and the health estimates are sensitive to several robustness checks. The within-Washington analysis therefore provides clear evidence that I-1183 changed very local retail access, but weaker evidence that those access changes caused worse birth health.

The paper contributes to two related literatures. Prior work establishes that alcohol purchasing and consumption respond to prices, legal availability, and spatial access, while studies of prenatal health use minimum-drinking-age laws, alcohol taxes, point-of-sale warnings, and temporary changes in availability to estimate consequences for pregnancy and birth outcomes (Fertig and Watson, 2009; Zhang, 2010; Cil, 2017; Nilsson, 2017; Luukkonen et al., 2023). I study a broader market liberalization that affected the full adult spirits market rather than targeting pregnancy or a narrowly defined policy margin. By observing changes in retail access, household purchasing, and birth health within the same reform, I can examine not only whether consumers respond to liberalization, but whether the incidence of that response is sufficient to generate measurable changes in birth health.

More broadly, the paper emphasizes the incidence of policy-induced behavioral responses. Aggregate demand alone is insufficient for predicting a specific downstream harm when consumers substitute across products and differ in the harms associated with additional consumption. In this setting, that distinction appears at two points: the 35 percent increase in spirits purchasing corresponds to a much smaller increase in total ethanol purchasing, and the available household evidence does not show that the remaining response was concentrated among the population most relevant for birth health. The results therefore show why the size of a market response cannot, by itself, determine the size of a particular health effect; what matters is how much the relevant exposure changes and who accounts for that change.

These reduced-form estimates can be used to recover an important component of an optimal corrective-tax calculation. The birth-health effects can be translated into valued social costs, while the purchasing response provides a quantity change over which to scale those costs. Recovering the optimal alcohol tax, however, also requires information on whose consumption changes, the marginal external harms associated with those consumers, substitution across products, existing price-cost wedges, and equilibrium tax pass-through. Section 5 develops a sufficient-statistics framework to show how the reduced-form evidence enters that broader policy problem.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting surrounding I-1183 and develops the access and incidence framework. Section 3 presents the data and empirical designs. Section 4 traces the reform from household purchasing and local retail access to statewide birth health, pregnancy accounting, and the within-Washington evidence. Section 5 places those estimates in the broader corrective-policy problem. The appendices report supporting empirical evidence and the formal sufficient-statistics derivation.

2 Liquor Privatization and Birth Outcome Pathways

2.1 Initiative 1183

Before privatization, Washington operated a state monopoly over the distribution and retail sale of spirits. Under Washington law, spirits are beverages containing alcohol obtained by distillation, excluding flavored malt beverages and including wines above 24 percent alcohol by volume. Private retailers, including grocery and convenience stores, could sell beer and wine, but spirits were sold through state-operated and contract stores. The Washington State Liquor Control Board (WSLCB) set uniform statewide spirits prices and applied a 51.9 percent markup over product cost, in addition to the state's spirits excise taxes. Such systems emerged after the repeal of Prohibition, when states were given broad authority to determine how alcohol would be distributed and sold (National Alcohol Beverage Control Association, 2026). Some states chose direct government control, in part to restrict access to higher-alcohol products. Seventeen states still retain some form of government control over the wholesale or retail sale of distilled spirits.[1]

The persistence of these systems reflects a longstanding tradeoff between limiting alcohol availability and allowing private competition in alcohol markets. That tradeoff was explicit in the debate over I-1183. The initiative characterized Washington's monopoly as "outdated, inefficient, and costly" and presented privatization as a way to reduce state operating costs, increase funding for state and local services, and move the government from selling liquor toward regulating private sellers (Washington Secretary of State, 2011). Supporters emphasized competition and consumer access, while opponents argued that a large expansion in liquor outlets would increase alcohol consumption and related harms (Washington Secretary of State, 2011). I-1183 therefore provides a useful setting for asking whether liberalizing a restricted alcohol market was accompanied by measurable downstream health consequences, including effects on health at birth.

[1] As of August 2026, the control states are Alabama, Idaho, Iowa, Maine, Michigan, Mississippi, Montana, New Hampshire, North Carolina, Ohio, Oregon, Pennsylvania, Utah, Vermont, Virginia, West Virginia, and Wyoming. Twelve of these states also retain some form of state control over off-premise retail sales, either through government-operated stores or designated private agents. Montgomery County, Maryland also operates as a local control jurisdiction.

Initiative 1183 replaced Washington's state-controlled spirits market with a private licensee system. Beginning in June 2012, the state closed its liquor stores and allowed qualifying private retailers, generally stores larger than 10,000 square feet, to sell spirits. This brought grocery stores, warehouse clubs, and other large retailers into a market from which they had previously been excluded. The state also auctioned the operating rights associated with its former store locations and disposed of its retail and distribution assets. The reform privatized wholesale distribution as well as retail sales and relaxed the conventional separation between producers, distributors, and retailers in the three-tier alcohol market (Williams et al., 2020). Finally, I-1183 eliminated the Board's authority to set retail spirits prices. State price setting was replaced by private pricing and new fees, including a 17 percent retailer fee on spirits revenue and a 10 percent distributor fee during the first two years.[2]

Privatization therefore changed several margins of Washington's spirits market at once. Retail availability changed visibly: the number of off-premise spirits outlets increased from roughly 330 to more than 1,400 and allowable selling hours expanded (Phillips et al., 2021). Consumers could also purchase spirits at grocery and big-box stores where they were already purchasing other goods. Post-privatization spirits prices remained high relative to neighboring states, while additional retailer entry expanded product assortment with little average-price response (Kerr et al., 2015; Williams et al., 2020; Illanes and Moshary, 2020). I-1183 thus changed prices, assortment, spatial access, shopping convenience, and market structure simultaneously. The statewide analysis consequently treats privatization as a bundled market reform rather than attempting to isolate any one of these margins.

Prior evidence establishes that these margins can affect alcohol markets. Reviews of retail privatization generally find increased sales of the privatized beverage, and spirits privatization in Alberta produced a persistent increase in spirits sales (Hahn et al., 2012; Trolldal, 2005). Alcohol purchasing and consumption also respond to legal and spatial access (Carpenter and Dobkin, 2011; Ahammer et al., 2022; Picone et al., 2010; Shakory et al., 2025).

Evidence from Washington clearly documents a market response, although its magnitude depends on what is measured. Using household scanner data, Barnett et al. (2020) find increased spirits purchases following I-1183, with larger responses among households that had previously been low or moderate purchasers. Administrative data reported by Seo (2019) show liquor quantity increasing by 5.97 percent in the year following privatization, or 4.89 percent per capita. Seo separately estimates that the convenience of one-stop shopping increases liquor demand by 16 percent holding prices, distance, and other market condi-

[2]I-1183 required the Liquor Control Board to close state liquor stores by June 1, 2012 and transition to a private licensee system. It also required the public auction of operating rights associated with state-owned store locations and the disposition of state retail and distribution assets.

tions fixed. Post-privatization spirits prices, however, remained high relative to neighboring states (Kerr et al., 2015; Williams et al., 2020). Survey evidence on realized drinking is less uniform: Kerr et al. (2018) find little change in total self-reported alcohol volume and a decline in reported spirits volume following privatization. These differences underscore that increased retail purchasing need not translate one-for-one into individual alcohol exposure. In Section 4.1, I provide new first-stage evidence using NielsenIQ Homescan data and separately estimate changes in spirits purchases and total ethanol-equivalent purchases following I-1183.

The reform also generated substantial variation in local retail exposure. Public control over liquor retailing can materially affect entry and the spatial distribution of stores (Seim and Waldfogel, 2013). In Washington, Illanes and Moshary (2020) find that additional liquor retailers expanded product assortment and increased purchasing with little average price response. I use the pre- and post-privatization retailer networks below to measure where I-1183 generated the largest changes in short-distance spirits access.

A smaller literature examines downstream consequences of I-1183 itself. Phillips et al. (2021) find no differential increase in hospitalizations for alcohol-related disorders following privatization, but find an increase in accidental-injury hospitalizations in metropolitan-urban Washington counties. Tabb et al. (2016) study alcohol outlets and violence in Seattle around privatization and find higher assault risk in areas with additional alcohol outlets. These findings suggest that downstream effects need not move uniformly with the aggregate market response. I extend this literature to reproductive and infant health, where the population capable of generating the relevant health effect is especially narrow.

2.2 Behavioral Incidence and Birth-Health Consequences

I-1183 changed more than the posted price of spirits. Before privatization, purchasing spirits generally required a separate trip to a state or contract liquor store. After privatization, many consumers could purchase spirits while already shopping for groceries or other goods. The relevant cost of acquiring alcohol therefore includes not only the monetary price of the product, but also travel time, search costs, store availability, product assortment, and the convenience cost of making an additional shopping trip. I-1183 changed several of these margins simultaneously.

These reductions in nonmonetary acquisition costs need not generate uniform behavioral responses. Consumers who place a higher value on convenience, live near newly eligible retailers, or were previously near the margin of purchasing spirits may respond more strongly than consumers whose alcohol demand was already relatively unconstrained by retail access.

Consistent with this mechanism, Seo (2019) estimates substantial consumer value from one-stop shopping after I-1183 and finds that the convenience effects of privatization can generate a meaningful increase in spirits demand.

A large response in spirits purchases need not generate a comparably large change in alcohol-related harm. The relevant exposure is ethanol rather than beverage volume itself. Spirits are particularly important in this setting because they contain substantially more ethanol per liter than beer or wine. Yet an increase in spirits purchases need not imply a proportional increase in total ethanol if consumers substitute across beverage categories. Moreover, even a large increase in total ethanol does not determine the effect on a particular health outcome. That mapping depends on the incidence of the behavioral response: who changes their alcohol use, by how much, and the health consequences of additional ethanol exposure for those consumers and circumstances.

A useful first-order representation is

$$\Delta H \approx \sum_i \Delta E_i \times MH_i, \tag{1}$$

where ΔE_i is consumer i's policy-induced change in ethanol consumption and MH_i is the marginal health harm associated with additional ethanol exposure in the circumstance relevant to the outcome under study. Aggregate ethanol demand summarizes $\sum_i \Delta E_i$, but the resulting health effect also depends on how that response is distributed across consumers with different marginal harms.

Birth-related health harm is not necessarily equivalent to an economic externality. Prospective parents may internalize some expected consequences of alcohol use for a future child through altruism or other preferences. At the same time, incomplete information, pregnancy before recognition, behavioral frictions, and harms borne directly by the child can create a wedge between the health harm generated by consumption and the harm internalized in the private decision. The reduced-form estimates below measure changes in health at birth; interpreting those changes as an external cost requires additional assumptions about how much of the resulting harm was already internalized by the decision maker.

For birth outcomes, this distinction is especially stark. Additional ethanol consumption can affect birth health only when the behavioral response occurs among people and at times when alcohol use can affect a pregnancy. A large increase in ethanol consumption concentrated outside that margin may therefore generate little change in birth health, while a much smaller aggregate response concentrated around pregnancy could generate a comparatively large effect.

This motivates a distinction between two empirical objects used throughout the paper.

The *market first stage* asks whether I-1183 changed spirits purchasing, total ethanol purchasing, and retail access. The *outcome-relevant first stage* asks whether the reform changed ethanol exposure among people whose alcohol use could affect a pregnancy. Household purchasing and retailer data identify the market response much more directly than this pregnancy-relevant exposure. The birth-health reduced form then measures whether the reform ultimately changed health at birth without requiring prenatal alcohol exposure itself to be observed.

2.3 Pathways to Observed Birth Outcomes

I-1183 can affect health among observed live births through two broad pathways. The first operates through prenatal exposure. Alcohol use may change before pregnancy recognition or during pregnancy, potentially affecting fetal growth and gestational duration. Prior economics research has studied these margins using minimum-drinking-age laws, alcohol taxes, point-of-sale warnings, and temporary changes in alcohol availability (Fertig and Watson, 2009; Zhang, 2010; Cil, 2017; Nilsson, 2017). These mechanisms motivate examining birth weight together with gestational duration and size for gestational age: a change in birth weight can reflect shorter gestation, altered fetal growth, or both.

The second pathway operates through which pregnancies enter the observed live-birth sample. Fertility responds to incentives and constraints (Becker and Barro, 1988), and alcohol access can also affect sexual behavior, pregnancy, and pregnancy resolution (Fertig and Watson, 2009; Cintina, 2015). Alcohol prices have likewise been linked to both abortions and birth outcomes (Luukkonen et al., 2023). Changes in conception, pregnancy continuation, or fetal survival can therefore alter average health among observed births even if health within a fixed set of pregnancies is unchanged.

Because natality records condition on live birth, the statewide reduced form combines prenatal-health and birth-pool pathways. I use reportable fetal deaths, reportable pregnancies, and maternal composition to examine whether the observed birth pool changed following I-1183. These measures capture only part of the selection process: the vital-statistics data do not recover abortions, early pregnancy losses, or conceptions that never enter the natality or fetal-death files.

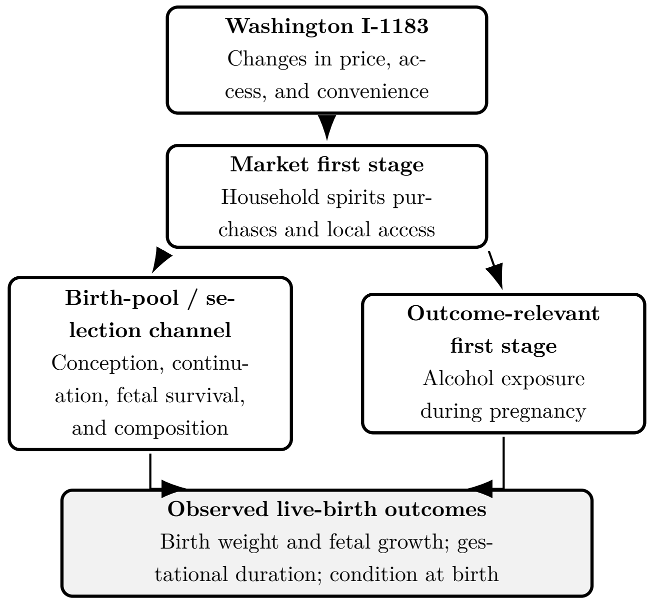

Figure 1: Market Response and Birth-Outcome Pathways

Notes: The market first-stage measures the aggregate purchasing and access response. Birth outcomes depend additionally on whether the response changes alcohol exposure during pregnancy or the set of pregnancies observed as live births. The data identify the market response more directly than prenatal exposure.

3 Data and Empirical Strategy

3.1 Vital Statistics and Birth Outcomes

I use restricted-use, individual-level natality and fetal-death records from the National Vital Statistics System (NVSS). The natality files cover the universe of registered live births in the United States and report infant and pregnancy outcomes, maternal characteristics, and place of residence.[3] The main sample includes births to mothers residing in the 50 states or the District of Columbia whose estimated fertilization month falls between January 2010 and December 2014. Washington births are defined by maternal residence rather than place of delivery. The final live-birth sample contains 19,805,760 records across 60 estimated-fertilization months; Appendix Table A.1 reports the complete sample construction and outcome-specific denominators.

The outcomes are chosen to capture clinically and economically meaningful dimensions of health at birth. Birth weight is a widely used summary measure of neonatal health: lower-weight infants face greater mortality and require more intensive medical care, and

[3]The National Center for Health Statistics provides descriptions of the NVSS natality data, fetal-death data, and annual public-use files and documentation. Restricted-use records additionally provide county geography.

differences in birth weight have been linked to later educational attainment, cognitive ability, and earnings (Almond et al., 2004; Black et al., 2005). Gestational duration captures a distinct margin of fetal development. Preterm birth carries substantial immediate medical costs and is associated with health and developmental consequences that can persist well beyond the birth hospitalization. These outcomes therefore measure health at delivery while also providing early indicators of potential longer-run consequences.

The primary outcomes distinguish fetal growth from gestational duration. I measure birth weight, low birth weight (less than 2,500 grams), very low birth weight (less than 1,500 grams), gestational age, early-term birth (37 or 38 completed weeks), preterm birth (fewer than 37 completed weeks), and very preterm birth (fewer than 32 completed weeks). Birth weight and gestational duration are related but not interchangeable: lower birth weight can result because pregnancy ends earlier or because fetal growth is slower at a given gestational age. To distinguish these margins, I additionally construct small-for-gestational-age (SGA) and large-for-gestational-age (LGA) indicators using the pre-I-1183 distribution among non-Washington singleton births within infant-sex-by-gestational-week cells. SGA denotes birth weight below the cell-specific 10th percentile and LGA birth weight above the 90th percentile. Because the 10th and 90th percentiles used to define SGA and LGA are themselves estimated from the birth data and could therefore shift with the policy, Appendix Table A.3 redefines SGA and LGA using the fixed, published sex-specific U.S. reference distribution from Aris et al. (2019).

The five-minute Apgar score provides a complementary measure of the infant's condition immediately after delivery. It sums five assessments—appearance, pulse, grimace response, activity, and respiration—each scored from zero to two. I use both the continuous score and indicators for scores below seven and below four. These cutoffs correspond to standard clinical categories: five-minute scores of 7–10 are considered reassuring, scores of 4–6 moderately abnormal, and scores of 0–3 low among term and late-preterm infants (American Academy of Pediatrics Committee on Fetus and Newborn and American College of Obstetricians and Gynecologists Committee on Obstetric Practice, 2015). A low Apgar score is therefore useful as a summary measure of neonatal condition, but it should not be interpreted on its own as evidence of birth asphyxia or as a predictor of an individual child's later neurologic outcome. I also construct indicators for congenital anomalies recorded on the birth certificate. I treat these measures as secondary outcomes because anomaly-item availability and reporting depend on the birth-certificate revision in use.

These birth-certificate measures do not capture all potential consequences of prenatal alcohol exposure. Prenatal alcohol exposure can generate a broad range of health and developmental consequences, most directly recognized clinically through fetal alcohol spectrum disorders (FASDs). FASD encompasses physical, cognitive, and behavioral impairments caused by prenatal alcohol exposure and is among the clearest established links between alcohol use during pregnancy and later child health (National Institute on Alcohol Abuse and Alcoholism, 2025). The NVSS records do not provide a direct measure of FASD or fetal alcohol syndrome (FAS). These conditions cannot generally be identified from a single measure at delivery, and many neurological, behavioral, and developmental manifestations become apparent only later in childhood. Some of the outcomes I observe, including low birth weight and congenital anomalies, can occur among children with FASD, but none is itself diagnostic of prenatal alcohol exposure (National Institute on Alcohol Abuse and Alcoholism, 2025).

Gestational age is measured using NCHS's edited combined gestation field (COMBGEST), reported in completed weeks. I treat values from 17 through 47 weeks as valid, corresponding to the allowable range in the NCHS files; records coded as unknown, missing, or outside this range are excluded. Through 2013, the combined measure is based primarily on gestational age calculated from the reported last menstrual period (LMP), with NCHS imputation or clinical/obstetric substitution when the LMP-based measure is missing or inconsistent. NCHS adopted the obstetric estimate as its preferred national gestational-age standard beginning in 2014 (Martin et al., 2015). To preserve a consistent measure across the analysis window, I use the edited combined field throughout and report the substitution and imputation flags by birth year in Appendix Table A.2. Records without valid gestational age are not assigned an estimated fertilization month; I impose no fallback.

Finally, the natality records condition on survival to live birth. The estimated effect on observed birth outcomes can therefore reflect both changes in health among pregnancies that continue to delivery and changes in which pregnancies enter the live-birth sample. I supplement the natality records with NVSS fetal-death data and maternal-composition outcomes to investigate this second pathway. The fetal-death analysis retains records meeting the NCHS tabulation criterion for stated or presumed gestation of at least 20 weeks.[4] I examine population-normalized live births and reportable fetal deaths, their sum as reportable pregnancies, and fetal deaths per 1,000 reportable pregnancies. These measures capture later fetal survival and pregnancies recorded in the vital-statistics system; they do not recover abortions, early pregnancy losses, or all conceptions. I likewise treat maternal characteristics as outcomes describing the observed birth pool rather than as controls in the primary health specification.

[4] The 2014 NVSS fetal-death user guide describes the tabulation rule and associated measurement changes.

3.2 Treatment Timing and the Statewide Estimand

Initiative 1183 took effect on June 1, 2012. Because alcohol exposure can matter throughout pregnancy, including before pregnancy recognition, I organize the primary analysis by estimated fertilization month rather than month of birth. Backdating births using gestational age to construct conception or fertilization cohorts is common in economics and epidemiology (Buckles et al., 2021; Noelke et al., 2019). The NVSS records report only month of birth and gestational age in completed weeks, however, so neither the exact delivery date nor the exact fertilization date is observed.

Let B_i denote the fifteenth day of infant i's reported birth month and let G_i denote gestational age in completed weeks. I estimate fertilization date as

$$\begin{aligned} \widehat{F}_i &= B_i - (7G_i + 3) + 14 \\ &= B_i - 7G_i + 11. \end{aligned} \tag{2}$$

Each adjustment corresponds to a feature of the available data. First, because the exact delivery day is unavailable, I assign delivery to the fifteenth day of the reported birth month. Second, G_i records completed weeks, leaving an unobserved remainder of zero through six days; I approximate this remainder by its midpoint of three days. Third, gestational age is conventionally dated from the beginning of the last menstrual period (LMP), approximately 14 days before fertilization. I therefore add 14 days when converting gestational age to an estimated fertilization date, consistent with the timing convention used in official conception statistics (Office for National Statistics, 2025). I then assign $\widehat{F}_i$ to its calendar month, which defines the estimated fertilization cohort used in the analysis.

For example, a birth recorded at 40 completed weeks is assigned a gestational duration of approximately 40 weeks and three days from LMP to delivery. Moving the date forward by 14 days from LMP places estimated fertilization approximately 269 days before the assigned delivery date. The procedure therefore approximates fertilization at about 38 weeks before delivery while making explicit the two sources of timing coarseness in the NVSS data: month-only delivery dates and gestational age recorded in completed weeks.

Let s denote maternal state of residence and t estimated fertilization month. The statewide treatment indicator is

$$D_{st} = \mathbf{1}\{s = \text{Washington}\} \times \mathbf{1}\{t \geq \text{June 2012}\}. \tag{3}$$

Washington births assigned to fertilization cohorts beginning in June 2012 are therefore treated, while contemporaneous cohorts in the donor states are untreated.

The binary treatment assignment necessarily simplifies the timing of pregnancy relative to the reform. Estimated fertilization is reconstructed from month of birth and gestational age in completed weeks rather than observed directly, so pregnancies close to June 2012 may change treatment status under modest changes in the timing rule. Moreover, pregnancies already underway when I-1183 took effect can overlap the post-policy environment for only part of gestation. I therefore use two timing sensitivities. First, I shift the estimated-fertilization treatment boundary by one cohort month in either direction to assess sensitivity to modest classification error. Second, I replace the single post-policy indicator with three fixed calendar windows corresponding to months 1–3, 4–6, and 7–9 after estimated fertilization and record whether each window occurs before or after June 2012. These specifications test whether the statewide conclusions depend on how pregnancies are aligned with the reform. Because the windows are constructed from calendar timing rather than realized biological exposure, I do not interpret their coefficients as trimester-specific causal effects. Appendix Table A.13 reports these timing sensitivities.

The natality data are repeated cross-sections: each state-month contains a new cohort of births rather than repeated observations of the same mothers or infants. The statewide estimand is the average effect of the post-I-1183 environment on outcomes among observed live births in Washington fertilization cohorts beginning after implementation, relative to the untreated counterfactual path constructed in Section 3.3. Because I-1183 may also change which pregnancies appear in the live-birth records, this estimand incorporates both changes in health within pregnancies and changes in the composition of observed births. It is therefore an intention-to-treat effect of the bundled reform on health among observed live births, not a biological dose response to prenatal ethanol exposure.

The preferred sample retains estimated-fertilization cohorts through December 2014. Two overlapping policy changes motivate shorter-window sensitivities. Licensed recreational cannabis sales began in Washington in July 2014 (Washington State Liquor and Cannabis Board, 2014), so I first exclude cohorts whose estimated fertilization occurs after cannabis retail commenced. This restriction removes conceptions beginning under the new cannabis regime, although pregnancies conceived earlier may still overlap with cannabis availability later in gestation. Washington's temporary $15.50-per-barrel beer-tax increment expired on June 30, 2013. I therefore also report estimates through June 2013 fertilization cohorts. Because that cutoff does not ensure that the full pregnancy occurred before the tax expired, I additionally report a stricter specification ending with October 2012 cohorts, whose fixed nine-month gestational windows end no later than June 2013.[5] Appendix Tables A.11 and A.14 report these sensitivities. These exercises assess robustness to overlapping policy

[5]Washington's Department of Revenue describes the temporary increase as applying through June 30, 2013; see *Comparative State and Local Taxes 2012*.

regimes rather than provide alternative headline estimates.

3.3 Constructing Washington's Counterfactual

Washington is the only treated jurisdiction, so the central identification problem is constructing the path its birth outcomes would have followed after June 2012 in the absence of I-1183. A conventional difference-in-differences design using all untreated states would require Washington's untreated outcome trend to be well approximated by the average trend across the donor states. That is unnecessarily restrictive in this setting: states differ persistently in birth outcomes and in their pre-policy trajectories, while the long monthly pre-period provides information that can be used to construct a more comparable counterfactual.

I therefore use synthetic difference-in-differences (SDID) (Arkhangelsky et al., 2021). SDID combines the reweighting logic of synthetic control with the fixed-effects structure of difference-in-differences. Donor weights place greater weight on untreated jurisdictions whose pre-policy outcome histories resemble Washington, while time weights place greater weight on pre-policy periods that make the weighted pre- and post-period comparisons more comparable. The subsequent weighted difference-in-differences regression absorbs persistent unit and time differences. This is useful here because identification depends on reproducing Washington's pre-policy evolution, rather than requiring Washington and its comparison group to have identical outcome levels.

The original SDID estimator is developed for panel data. NVSS instead consists of repeated cross-sections: the births observed in one state-month are different individuals from those observed in the next. I therefore implement the repeated-cross-section adaptation of Morin (2024). Because treatment occurs at the state level, state-by-estimated-fertilization-month outcomes can be used to construct the SDID donor and time weights even though the underlying individuals change across periods. In the preferred no-control specification, applying these weights to the state-month means is algebraically equivalent to the corresponding weighted individual repeated-cross-section regression. The donor pool contains the other 49 states and the District of Columbia.

Rather than choose a different synthetic comparison for every outcome, I use one common set of donor and time weights for the primary birth-health analysis. This keeps the estimated treatment effects tied to a common counterfactual and exploits shared pre-policy variation across related outcomes (Sun et al., 2025; Tian et al., 2026). I construct the primary weighting target from birth weight, preterm birth, and gestational length. These outcomes span the two principal dimensions of health studied here: birth weight captures fetal growth, while gestational length and preterm birth capture both the level and a clinically meaningful lower tail of pregnancy duration. I prefer these broad measures for constructing the counterfactual to more derived outcomes such as SGA, LGA, or low birth weight, whose definitions depend on additional thresholds or on combinations of the underlying birth-weight and gestational-age measures. The resulting index therefore summarizes common pre-policy variation in fetal growth and gestational duration without tailoring the synthetic comparison to any one reported treatment outcome.

For component k, let μ_k^D and σ_k^D denote its mean and standard deviation across donor-state-by-month observations in the pre-policy period. The index is

$$H_{st} = \frac{1}{3}\left[-\frac{BW_{st} - \mu_{BW}^D}{\sigma_{BW}^D} + \frac{Preterm_{st} - \mu_{Preterm}^D}{\sigma_{Preterm}^D} - \frac{Gestation_{st} - \mu_{Gestation}^D}{\sigma_{Gestation}^D}\right]. \tag{4}$$

Higher values indicate worse fetal growth or shorter gestation. As a robustness check, I also estimate a common-vector specification that enters these outcomes separately into the weighting problem rather than collapsing them into a scalar index, along with alternative outcome-index constructions reported in Appendix A.2.

I selected the weighting target using only pre-I-1183 data. I sequentially treated the final four, six, and nine pre-policy months as pseudo-post periods and evaluated how well each candidate design predicted outcomes that were known but withheld from the weighting procedure. The comparison included the three-outcome index, outcome-specific weights, alternative scalar and common-vector multiple-outcome targets, and an unweighted all-state comparison. Across 14 outcomes and three holdout windows, the selected index produced the lowest average scaled RMSPE and the smallest worst-case prediction error. Washington's actual post-I-1183 outcomes were never used to select the weighting target. Appendix Table A.4 reports the complete blinded comparison.

Applying SDID to the selected index yields donor weights ω_s and pre-period time weights λ_t. I then hold these weights fixed across the primary birth outcomes and estimate, for outcome j,

$$\bar{Y}_{stj} = \tau_j D_{st} + \alpha_{sj} + \delta_{tj} + u_{stj}, \qquad a_{st} = \Omega_s \Lambda_t, \tag{5}$$

where $\Omega_s = 1$ for Washington and $\Omega_s = \omega_s$ for donor state s. In the pre-policy period, Λ_t is given by the estimated SDID time weight λ_t; post-policy months receive equal weight. Thus, every outcome is evaluated using the same synthetic Washington and the same weighting of the pre-policy period. Differences across the estimated τ_j therefore reflect differences across birth outcomes rather than differences in the counterfactual used to estimate them.

The underlying NVSS data are individual repeated cross-sections, but the preferred specification contains no individual-level controls. Because treatment and the SDID weights vary only at the state-by-estimated-fertilization-month level, estimation on the corresponding state-month outcome means is algebraically equivalent to applying those weights to the individual birth records. I use the individual records to construct outcomes and characterize the composition of observed births; Equation (5) makes the aggregation explicit.

The preferred specification deliberately does not condition on individual- or state-level covariates. When the untreated outcome path can be credibly identified without conditioning on observed characteristics, covariate adjustment is not required for difference-in-differences identification; covariates become relevant when identification instead relies on parallel trends conditional on observed characteristics (Sant'Anna and Zhao, 2020; Roth et al., 2023). In the present setting, SDID uses pre-policy outcomes themselves to construct Washington's counterfactual path. Moreover, several natural individual-level covariates, including maternal demographic characteristics, may respond to privatization through changes in the composition of the observed birth pool. Conditioning on such variables could therefore absorb part of the policy response and change the estimand away from the total effect of the reform. I use the unconditional, no-control specification as the primary design and evaluate its credibility using pre-policy outcome fit rather than conditioning on potentially endogenous characteristics.

Identification then requires that the weighted donor mixture approximate the path Washington's outcomes would have followed after June 2012 in the absence of I-1183. Because SDID includes unit fixed effects, Washington and the weighted donors need not have identical outcome levels; what matters is whether the weighted comparison captures Washington's untreated evolution. I assess the plausibility of this assumption using blinded pre-policy holdouts, pre-treatment fit, donor concentration, alternative weighting targets, influential-donor refits, and placebo assignments. Appendix A.2 reports the complete counterfactual audit.

3.4 Inference with One Treated State

The inferential strategy has a fixed hierarchy. At the coefficient level, the primary question is whether Washington's estimated effect is large relative to the sampling variation implied by applying the same design to untreated states; Algorithm 4 placebo-variance intervals are the primary uncertainty measure for that question. At the family level, the question is whether the pattern of effects across related outcomes is collectively unusual relative to placebo states; joint full-refit statistics are the primary evidence for that question. Placebo effect ranks and RMSPE ranks are complementary diagnostics that show how unusual Washington's estimate and post-policy fit are within the comparison distribution. Finally, fixed-counterfactual Newey–West standard errors provide a time-series sensitivity conditional on the chosen synthetic comparison, but they are not used for primary inference.

The nearly 20 million birth records make state-month outcomes precisely measured, but they do not create additional independent policy assignments. I-1183 occurred in a single state, so inference cannot rely on the number of individual births or on conventional cluster-robust asymptotics with many treated units. The primary uncertainty calculation instead asks how much estimated treatment effects vary when the same design is applied to untreated jurisdictions.

My primary sampling interval follows the Algorithm 4 placebo-variance procedure of Arkhangelsky et al. (2021). I successively designate each of the 50 non-Washington jurisdictions as treated beginning in June 2012, exclude Washington from that jurisdiction's donor pool, and reestimate the complete design. Each refit therefore reconstructs the outcome standardization, donor weights, time weights, and treatment effect rather than applying Washington's fitted counterfactual mechanically to another state. Let $\hat{\tau}_j^P$ denote the resulting estimate for placebo jurisdiction j. The placebo-variance standard error is

$$\widehat{\sigma}_{\mathrm{A4}} = \left[\frac{1}{J} \sum_{j=1}^{J} \left(\hat{\tau}_j^P - \bar{\tau}^P \right)^2 \right]^{1/2}, \qquad J = 50, \tag{6}$$

and the primary 95 percent sampling interval is $\hat{\tau}_{WA} \pm 1.96\widehat{\sigma}_{\mathrm{A4}}$.

The procedure uses the variation in estimated effects across the placebo states to measure uncertainty around Washington's estimate. Intuitively, it asks how much apparent treatment effects vary when the same design is applied to states that did not experience I-1183. This requires the placebo states to provide a reasonable benchmark for the amount of unexplained variation in Washington's estimate.

I also report placebo effect ranks to show how unusual Washington's estimate is relative to estimates obtained for untreated states. For each outcome, the two-sided rank compares the absolute value of Washington's estimated effect with the absolute effects from the fully refitted placebo assignments. A low rank means that few placebo states produce an effect as large as Washington's. Because I-1183 was not randomly assigned across states, I interpret these ranks as empirical diagnostics rather than exact randomization-inference p-values.

I also test related outcomes jointly. This matters because several outcomes can move in the same adverse direction even when no single estimate is individually precise. For each pre-specified outcome family, I summarize the estimated effects in two ways: the largest absolute standardized effect, M, and the sum of squared standardized effects, Q. The first detects whether any one outcome is unusually large; the second detects whether the collection of effects is unusually large overall. I calculate the same statistics for each fully refitted placebo state and compare Washington with that distribution. These joint diagnostics therefore ask whether the overall pattern of Washington's estimates is unusual, rather than whether a particular coefficient happens to be large. Appendix A.3 reports the construction and results.

I also compare Washington's post-policy divergence with its pre-policy fit. The post/pre RMSPE statistic asks whether the Washington–synthetic gap becomes substantially larger after I-1183 than it was before the reform. I rank Washington's post/pre RMSPE against the corresponding values from the fully refitted placebo states. As a sensitivity, I also repeat this comparison after excluding placebo states whose pre-treatment fit is substantially worse than Washington's. This provides a diagnostic of counterfactual fit and post-policy divergence.

Finally, I report a time-series sensitivity based on the monthly gap between Washington and its estimated synthetic comparison. Monthly birth outcomes are persistent, and adjacent Washington–synthetic gaps may therefore be serially correlated. Treating those monthly observations as independent could understate uncertainty. Holding the synthetic counterfactual fixed, I calculate Newey–West standard errors that allow for autocorrelation in the gap over time. I use six monthly lags as the baseline HAC specification. Appendix Table A.9 reports 3-, 6-, and 12-month bandwidths for the complete primary outcome family and index. The bandwidth changes some conditional standard errors but does not change any conventional 5-percent classification; this comparison does not alter the paper's inferential hierarchy.

This exercise asks how precisely the average post-policy gap is estimated conditional on the synthetic comparison already having been chosen. It does not account for uncertainty from having only one treated state or from constructing the counterfactual itself, so I treat it as a robustness check rather than primary inference.

The inferential hierarchy is maintained throughout the paper: Algorithm 4 intervals provide the primary coefficient-level uncertainty; joint full-refit statistics provide the primary family-level assessment; placebo effect ranks and RMSPE ranks describe the comparison distribution and counterfactual fit; and fixed-counterfactual HAC estimates are reported only as a sensitivity.

3.5 Measuring the Market First Stage

I use NielsenIQ Homescan consumer-panel records from 2010 through 2014 to measure the market response to I-1183. Homescan follows a large longitudinal panel of U.S. households that record purchases intended for personal, in-home use across retail outlets. During this period, the panel contains roughly 40,000–60,000 active households per year and is designed to approximate the U.S. household population through demographic and geographic recruitment targets and household-specific projection factors (Kilts Center for Marketing, 2026). For each recorded shopping trip, I observe the products purchased, purchase date, quantity, and retailer; linked UPC files provide product characteristics including package and multi-pack size. The household files additionally report geography and demographic characteristics, including household composition, income, presence and age of children, and characteristics of the household heads.

Homescan is particularly useful here because these data allow me to observe both the product composition and household incidence of the market response. I classify purchases as spirits, beer, or wine, which allows me to distinguish increased spirits purchasing from substitution across beverage categories and to construct total ethanol-equivalent purchases. Household characteristics then provide suggestive evidence about which types of households account for the purchasing response. NielsenIQ's projection factors make the panel more representative of the underlying household population, although they do not imply that every state-month or demographic subgroup is perfectly representative. Validation work finds recording error in Homescan transactions but substantially better reporting of purchased quantities than prices (Einav et al., 2010).

I convert product size and multipack information into beverage liters and apply NielsenIQ's household projection factors when aggregating purchases to state-by-month cells. The resulting outcomes are monthly liters purchased per projected household, where the projection factors reweight participating households to represent the underlying household population. Because ethanol rather than beverage volume is the relevant exposure in the framework of Section 2, I also convert each beverage category to ethanol-equivalent liters, assuming alcohol contents of 40 percent for spirits, 5 percent for beer, and 12 percent for wine. This conversion allows the spirits response to be distinguished from the change in total alcohol content purchased after accounting for substitution across beverage categories.

I estimate a separate state-month RC-SDiD design for the purchasing outcomes, treating Washington beginning in June 2012 and constructing the purchasing counterfactual using observations through May 2012. Donor and time weights are learned from a standardized pre-policy purchasing index that combines spirits ethanol liters per household, total ethanol liters per household, and spirits purchase incidence. I apply the same one-treated-state inferential logic described in Section 3.4, including fully refitted placebo-state assignments. The Homescan analytic panel contains the contiguous states and District of Columbia; Alaska and Hawaii are not observed. With Washington as the treated state, this leaves 48 eligible untreated jurisdictions, all of which are used as placebo assignments. For the household-heterogeneity analysis, I hold the all-household donor and time weights fixed across demographic groups. Differences across groups therefore reflect differences in estimated purchasing responses rather than differences in the synthetic comparison used for each group.

Homescan measures purchases recorded by households, not alcohol consumed by particular individuals. It excludes most on-premise purchases and cannot identify who purchased or consumed a product, whether it was consumed during the month in which it was purchased, or whether a household member was pregnant. Household-head characteristics therefore provide evidence about the incidence of the *market* response, not a direct measure of pregnancy-relevant ethanol exposure. I supplement Homescan with descriptive evidence from NSDUH and BRFSS on self-reported adult drinking. Because these surveys differ in their samples, periods, and definitions of alcohol use, I report them separately as corroborating evidence rather than treating them as additional causal estimates.

3.6 Local Access Within Washington

The statewide analysis estimates the effect of the full post-I-1183 environment. I complement that design with a within-Washington analysis that asks whether birth outcomes changed more in places where privatization generated larger increases in short-distance spirits access. I-1183 took effect throughout Washington at the same time, but counties differed substantially in the retail infrastructure available to respond to the reform. Counties with more grocery and big-box capacity already in place before privatization were positioned to experience larger increases in convenient spirits access once those retailers became eligible to sell spirits.

The within-Washington design therefore does not compare treated and untreated counties. Instead, it exploits differences across Washington's 39 counties in their predetermined exposure to the retail-access consequences of a common statewide reform. The identifying variation comes from whether counties with greater pre-policy grocery coverage experienced larger post-I-1183 changes in access and birth outcomes. The monthly panel improves measurement and permits examination of differential trends around the reform, but the underlying exposure variation is across counties.

A direct comparison of counties with large and small realized retailer expansions would be difficult to interpret causally. Post-privatization entry is endogenous: retailers may enter where alcohol demand, income, population, development, or other determinants of birth outcomes are changing. I instead use grocery infrastructure measured before I-1183 to predict where the reform generated larger expansions in spirits access. The basic empirical question is whether counties predisposed to receive a larger access expansion experienced both larger increases in access and different changes in birth outcomes after June 2012.

I measure the spirits-retailer network using the complete June 2011 WSLCB roster of 328 civilian state and contract stores and a post-privatization universe of 1,348 off-premise spirits locations. Historical license records do not permit a reliable reconstruction of monthly entry and exit during 2011–2014, so I treat these data as pre- and post-policy network snapshots rather than constructing a retailer-entry event study. The design therefore exploits the size of the network change associated with privatization rather than the precise month in which each retailer entered. Appendix B reports the retailer and geocoding audit.

For each Census block group, I calculate distance from its fixed 2010 population centroid to the spirits-retailer network and aggregate these measures to the county using fixed 2010 population weights. The primary access measure is the share of the county population living in block groups whose centroid lies within one mile of at least one spirits retailer. Thus, an increase from 40 to 50 percent coverage means that population centers containing an additional 10 percent of the county's fixed population are brought within one mile of a retailer. It does not imply that a particular mother gained a retailer within one mile of her residence. I also examine coverage within 0.5, 3, 5, and 10 miles, nearby-retailer counts, nearest-retailer distance, and smooth gravity measures. Holding population locations and weights fixed ensures that changes in measured access reflect changes in the retailer network rather than contemporaneous population movement.

To isolate policy-induced differences in access, I measure pre-existing grocery retail infrastructure using USDA SNAP retailer records from before I-1183. I geocode qualifying SNAP-authorized grocery retailers and use their locations to construct the share of each county's fixed 2010 population with a retailer within five miles. This measure captures retail capacity that was already in place before privatization and therefore could not have been chosen in response to I-1183. Grocery and big-box stores generally could not sell spirits before the reform, but I-1183 made much of this existing retail capacity eligible to enter the spirits market. Counties with greater pre-policy grocery coverage were therefore positioned to experience larger increases in spirits access once privatization took effect.

The instrument and treatment use different distance thresholds by design. Five-mile pre-policy grocery coverage measures the stock of nearby retail capacity available to respond to privatization, whereas one-mile spirits coverage captures the more local convenience margin generated by that response. The excluded exposure measure is therefore the post-I-1183 indicator interacted with standardized pre-policy five-mile grocery coverage. I estimate the first stage

$$SpiritsAccess_{ct} = \pi \left(Post_t \times PreGroceryAccess_c \right) + \alpha_c + \delta_t + u_{ct}, \tag{7}$$

where $SpiritsAccess_{ct}$ is one-mile county population coverage, α_c are county fixed effects, and δ_t are calendar-month fixed effects. The coefficient π asks whether counties with more pre-existing grocery infrastructure experienced larger increases in very local spirits access after I-1183. County fixed effects absorb persistent differences across counties, while month fixed effects absorb shocks common to Washington as a whole.

I also test whether the measured access change is sensitive to the geographic aggregation and treatment mapping used to construct it. First, I reconstruct one-mile coverage using populated 2010 Census-block internal points rather than block-group population centroids. Second, I measure the share of the fixed 2010 population that was outside one mile of a spirits retailer before privatization but brought within one mile afterward. Third, I construct the population-weighted change in nearest-retailer distance, allowing the nearest retailer to lie across county boundaries. As a separate instrument-mapping sensitivity, I reconstruct pre-policy SNAP grocery coverage using the same Census-block geography. These alternatives were defined as measurement checks rather than selected according to their relationship with birth outcomes; the preferred specification remains block-group one-mile spirits coverage instrumented with block-group five-mile pre-policy grocery coverage.

I first estimate the reduced form by replacing the first-stage outcome with a county birth outcome. This specification asks whether birth outcomes changed more after I-1183 in counties with greater pre-policy grocery exposure. I treat this reduced form as the most transparent within-Washington estimand: it measures whether places more exposed to the retail-access consequences of the reform experienced different changes in birth outcomes, without yet requiring the stronger assumption that those differences operate exclusively through the particular spirits-access measure used in the first stage.

I then use two-stage least squares to express that policy-exposure response in units of instrument-predicted spirits access:

$$Y_{ct} = \beta \widehat{SpiritsAccess}_{ct} + \alpha_c + \delta_t + \varepsilon_{ct}, \tag{8}$$

where $\widehat{SpiritsAccess}_{ct}$ is the change in one-mile coverage predicted by the interaction between the post period and pre-policy grocery access. I scale β so that it represents the change in the birth outcome associated with a 10-percentage-point increase in policy-induced county population coverage. Outcome regressions use the relevant county-month birth denominator as weights and cluster standard errors by county. Although the regressions are estimated using county-month observations, the differential policy exposure varies across 39 counties rather than across independent county-month policy assignments.

The IV interpretation requires a stronger assumption than the reduced-form policy-exposure comparison. Specifically, the interaction between pre-policy grocery infrastructure and the post-I-1183 period must affect birth outcomes only through the policy-induced change in spirits access captured by the first stage, conditional on county and month fixed effects. This exclusion restriction would fail if pre-policy grocery coverage also predicts post-2012 changes in urbanization, economic conditions, demographic composition, health care access, or other determinants of birth outcomes. A separate concern is treatment mapping: cross-county shopping and spatial spillovers may cause county residence to measure the relevant retail environment imperfectly. I assess these concerns using balance and differential-pretrend tests, placebo treatment dates, county-specific trends, leave-one-county-out estimates, and alternative access measures in Appendix B.

Under the exclusion and treatment-mapping assumptions, 2SLS rescales the reduced-form county response into the change in birth outcomes associated with a policy-induced increase in short-distance spirits access. The treatment is county population coverage, not whether a particular mother has a retailer within one mile of her residence, and the design does not identify maternal alcohol purchasing, consumption, or a prenatal dose response. Because access is continuous and responses may differ across counties, I do not interpret the coefficient as a conventional individual-level "LATE." Rather, it summarizes the county-level birth-outcome response to the variation in access induced by pre-existing retail capacity. The within-Washington design therefore asks whether I-1183 generated larger access expansions in places predisposed to receive them and whether birth outcomes changed systematically with those expansions.

4 Results

The empirical sequence follows the incidence framework. I first establish the market response to I-1183. I then ask whether the statewide birth-health reduced form shows a corresponding deterioration, whether changes in the observed birth pool help explain the estimates, and whether counties with larger policy-induced access changes experience different birth outcomes.

4.1 Market First Stage: Household Alcohol Purchases

I-1183 produced a large response in household spirits purchasing. NielsenIQ Homescan records imply an increase of 0.117 beverage liters per projected household per month, equal to 35.0 percent of Washington's pre-policy mean. Expressed in ethanol, spirits purchases increase by 0.047 liters per projected household per month. Washington's fully refitted placebo-state rank is 0.061, placing the spirits response near the upper tail of the comparison distribution. The reform therefore generated a substantial market first stage in the product category it directly liberalized.

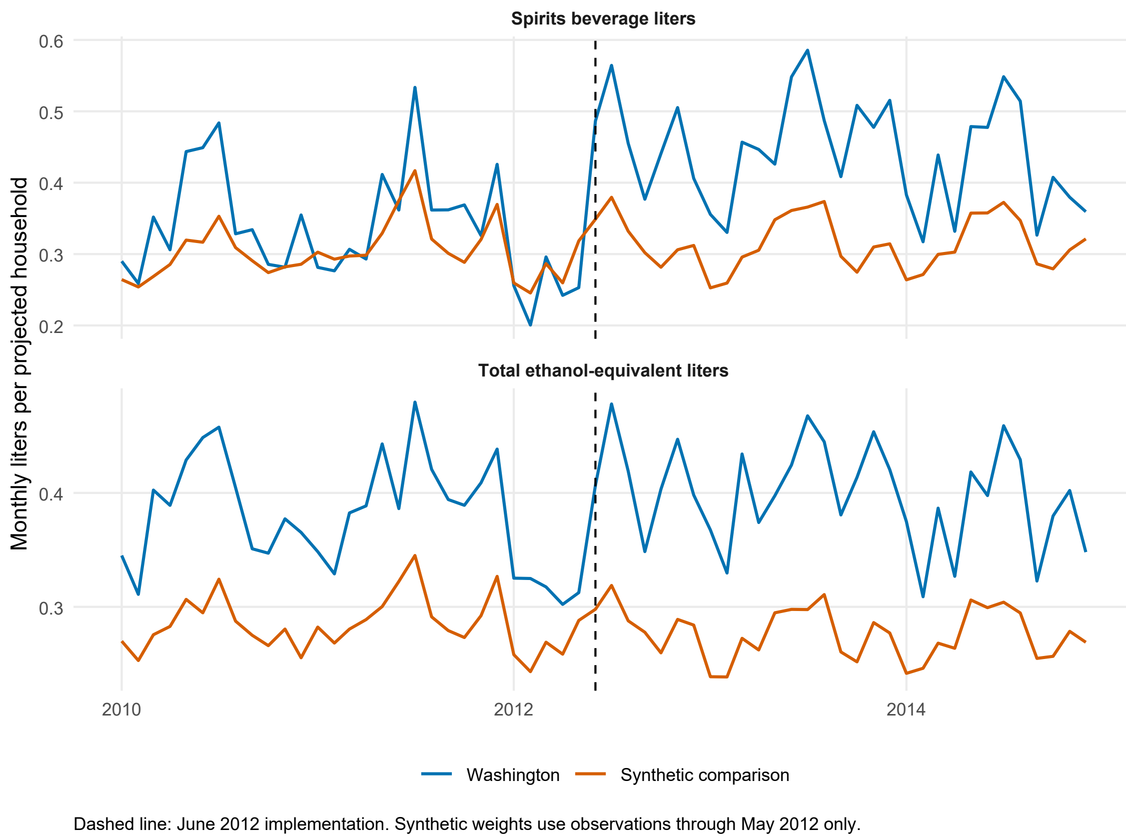

Figure 2: Washington and Synthetic Homescan Purchasing Paths

Notes: Panels show monthly projected purchases per household from January 2010 through December 2014 for Washington and its synthetic comparison. Donor and time weights are learned using observations through May 2012; Washington is treated beginning in June 2012. Spirits are measured in beverage liters and total alcohol in ethanol-equivalent liters. Homescan records household retail purchases and excludes most on-premise purchases; it does not identify the purchaser, consumer, or timing of consumption.

The increase in spirits does not translate one-for-one into total ethanol purchased. Figure 2 shows a pronounced post-June 2012 divergence in spirits purchases but a smaller separation in total ethanol. Spirits and beer purchases increase while wine purchases decline. Summing across beverage categories, total ethanol-equivalent purchases rise by 0.035 liters per projected household per month, or 9.2 percent of the pre-policy mean. The corresponding placebo-state rank is 0.286. Thus, the clearest market response is an increase in spirits purchasing; substitution across beverage categories substantially attenuates the implied change in total ethanol purchased. Appendix Table A.15 reports the category-specific estimates and inferential diagnostics.

Independent evidence supports the conclusion that privatization increased spirits purchasing. Using administrative sales data, Seo (2019) reports a 5.97 percent increase in observed liquor quantity, or 4.89 percent per capita, between the year before and the year after privatization. Seo's separate 16 percent estimate isolates the value of one-stop shopping in a structural counterfactual and should not be interpreted as the observed statewide increase. Although the administrative and Homescan measures differ in population coverage and empirical design, both indicate a meaningful expansion in spirits purchasing following I-1183.

The household incidence of that response is less supportive of a large pregnancy-relevant first stage. Appendix Table A.16 applies the common all-household counterfactual across demographic groups. The largest estimated increase in total ethanol occurs among households whose female head is age 45 or older. Estimates for younger households are smaller, and the under-35 estimate is negative, although pre-policy fit is poor for every beverage outcome in that group. The Homescan demographic fields describe household heads rather than the purchaser or consumer and do not identify pregnancy. These results therefore provide little evidence that the aggregate purchasing response was concentrated among households in prime childbearing ages, but they cannot establish that alcohol exposure during pregnancy was unchanged.

Self-reported drinking evidence provides no clearer pregnancy-relevant first stage. BRFSS shows an increase in any past-month drinking but little change in binge or heavy drinking, while published NSDUH estimates show declines in past-month and binge drinking. Differences in sampling frames, periods, and outcome definitions prevent a direct reconciliation of these estimates, and neither source identifies drinking during pregnancy. I therefore treat them as descriptive corroboration rather than additional causal estimates and report them in Appendix A.4.

Taken together, the evidence establishes a large market response to I-1183 but a much weaker mapping from that response to the exposure relevant for birth outcomes. Spirits purchasing rose sharply, while the increase in total ethanol was smaller and the available demographic and survey evidence does not show that the response was concentrated among people whose alcohol use could affect a pregnancy.

4.2 Statewide Effects on Health at Birth

The statewide point estimates are more consistent with reduced fetal growth than with shorter gestation. Among observed live births, average birth weight falls by 10.28 grams, low birth weight rises by 0.274 percentage points, SGA rises by 0.357 percentage points, and LGA falls by 0.396 percentage points. Together, these estimates describe a shift toward lower fetal growth across several margins of the birth-weight distribution. Relative to Washington's pre-policy means, the estimated changes in low birth weight, SGA, and LGA are 4.4, 4.8, and −3.1 percent, respectively. Table 1 reports the complete primary outcome family.

Table 1: Statewide Effects on Health at Birth

Outcome	Estimate	95% interval	WA pre mean	Effect rank	Fit rank
Birth weight and fetal growth					
Birth weight	-10.28	[-25.76, 5.20]	3364.01	0.235	0.157
Low birth weight	0.274	[-0.134, 0.682]	6.175	0.216	0.176
Very low birth weight	0.047	[-0.226, 0.320]	1.023	0.588	0.569
SGA	0.357	[-0.268, 0.982]	7.420	0.196	0.059
LGA	-0.396	[-1.055, 0.263]	12.909	0.314	0.118
Term LBW	0.111	[-0.213, 0.436]	2.293	0.373	0.176
Gestational duration					
Preterm	-0.089	[-0.914, 0.737]	9.874	0.804	0.608
Very preterm	0.017	[-0.316, 0.351]	1.456	0.882	0.235
Early term	0.145	[-1.686, 1.975]	22.751	0.882	0.510
Gestation	0.014	[-0.078, 0.105]	38.887	0.765	0.686
Prespecified family index					
Adverse fetal-growth/gestation index	0.019	[-0.289, 0.327]	–	0.902	0.902

Notes: Estimates are intention-to-treat effects of Washington's post-I-1183 environment among observed live births. All outcomes use the common donor and time weights learned from the pre-policy fetal-growth and gestational-duration index. Binary outcomes are in percentage points, birth weight is in grams, gestation is in weeks, and the index is measured in donor-state pre-period standard deviations. The 95 percent intervals use Algorithm 4 full-refit placebo variance across the 50 non-Washington assignments. Effect ranks compare the absolute value of Washington's estimate with the corresponding fully refitted placebo effects; fit ranks compare Washington's post/pre RMSPE ratio with the placebo distribution. These ranks are diagnostic comparisons rather than exact randomization-inference p-values because I-1183 was not randomly assigned and Washington and the placebo assignments use different donor pools.

The point estimates are adverse, but the one-treated-state inference is not decisive. Every Algorithm 4 interval in Table 1 includes zero. The two-sided placebo effect ranks are 0.235 for birth weight, 0.216 for low birth weight, 0.196 for SGA, and 0.314 for LGA, placing none of these estimates far in the tail of the corresponding placebo distribution. The intervals are also wide enough to include both economically meaningful adverse effects and small effects in the opposite direction. Thus, although the fetal-growth estimates move in a common adverse direction, the design does not distinguish them from the variation generated when the same procedure is applied to untreated states.

Gestational-duration outcomes do not reinforce the adverse fetal-growth pattern. If the decline in birth weight were driven primarily by earlier delivery, I would expect shorter gestation and more preterm birth. Instead, preterm birth falls by 0.089 percentage points, early-term birth rises by 0.145 percentage points, and average gestation rises by 0.014 weeks; all three estimates are imprecise. Together with the increase in SGA and decline in LGA, this pattern is more consistent with a shift in fetal growth conditional on gestational duration than with a shift toward earlier delivery. The prespecified fetal-growth and gestational-duration index is correspondingly close to zero at 0.019 standard deviations, with a 95 percent interval of $[-0.289, 0.327]$. Condition-at-birth outcomes are also mixed and are reported in Appendix Table A.6.

Family-level inference asks a broader question than the coefficient-by-coefficient intervals: whether the collection of related birth-health estimates is unusual when considered together. I define the fetal-growth family to include the six prespecified birth-weight outcomes reported in the first panel of Table 1: birth weight, low birth weight, very low birth weight, SGA, LGA, and term low birth weight. These outcomes measure overlapping but distinct features of fetal growth, so several moderate movements in the same direction could be informative even if no single coefficient is precisely estimated. Joint inference also guards against drawing conclusions from whichever outcome happens to produce the largest individual estimate.

For this family, I calculate two full-refit statistics for Washington and for every placebo assignment. The maximum statistic, M, asks whether any one standardized effect is unusually large, while the quadratic statistic, Q, asks whether the collection of standardized effects is unusually large overall. Washington's M statistic has a placebo rank of 0.529 and its Q statistic a rank of 0.412. Thus, neither the largest individual movement nor the combined pattern across the six fetal-growth outcomes is unusual relative to the fully refitted placebo states. Appendix Table A.8 reports the construction and results. The statewide evidence is therefore consistent with an adverse fetal-growth pattern in the point estimates, but the family-level evidence does not distinguish that pattern from the variation generated by the same design in untreated states.

Timing sensitivity checks address uncertainty in how pregnancies are aligned with the June 2012 implementation of I-1183. Appendix Table A.13 examines this timing uncertainty by replacing the single post-policy indicator with three fixed calendar windows corresponding to months 1–3, 4–6, and 7–9 after estimated fertilization. None of the window-specific Algorithm 4 intervals excludes zero. Jointly, the three coefficients on the prespecified health index are also unremarkable relative to the fully refitted placebo assignments, with a placebo rank of 1.000. A separate one-month shift in the exposure mapping likewise changes individual coefficients without producing a clearer pattern. These exercises show that the conclusions do not become stronger under alternative plausible mappings of pregnancies to the reform.

Appendix Table A.14 examines whether the statewide results are sensitive to the expiration of Washington's temporary beer-tax increase in June 2013. Because that change altered the relative price environment after I-1183, later cohorts are exposed to a different alcohol-policy regime than earlier post-reform cohorts. The estimated effect on birth weight is -10.28 grams in the full sample, -10.45 grams through June 2013 fertilization cohorts, and −9.12 grams in the stricter sample whose fixed nine-month gestational windows end before the tax expires. The similarity of these estimates suggests that the headline birth-weight result is not driven by the later policy regime. The broader outcome family remains imprecisely estimated in each specification, however, and the strictest comparison contains only five fully exposed post-I-1183 fertilization cohorts. I therefore treat this exercise as a policy-regime sensitivity rather than as a competing headline estimate.

4.3 Birth-Weight Distribution

The fetal-growth estimates in Section 4.2 do not reveal whether the apparent shift is concentrated around the conventional low-birth-weight threshold or reflects broader movement across the distribution. This distinction matters because birth-weight ranges differ in their medical and longer-run consequences (Almond et al., 2004; Black et al., 2005). Figure 3 therefore reports treatment effects for mutually exclusive 500-gram bins that exhaust the observed birth-weight distribution.

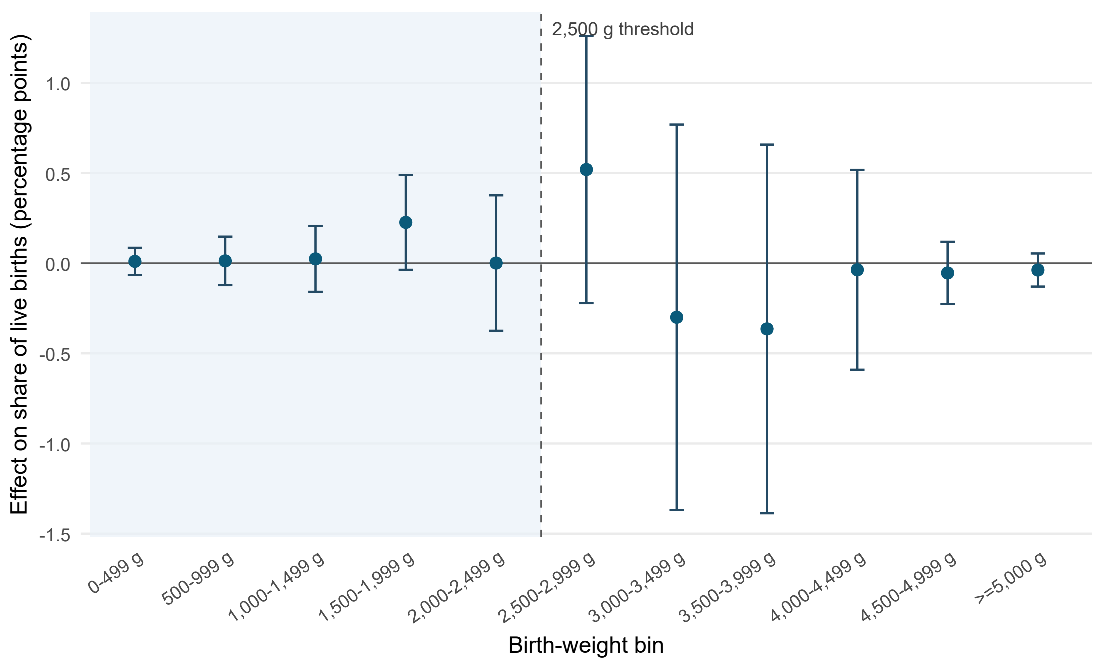

Figure 3: RC-SDiD Effects Across the Birth-Weight Distribution

Notes: Points are percentage-point RC-SDiD estimates for mutually exclusive 500-gram bins among observed live births in January 2010–December 2014 estimated-fertilization cohorts. The design uses the preferred common donor and time weights and treats Washington beginning in June 2012. Horizontal bars are 95 percent Algorithm 4 full-refit placebo-variance intervals based on 50 non-Washington assignments. The vertical line marks zero; the divider marks the conventional 2,500-gram low-birth-weight threshold.

The point estimates do not show a simple crossing of the 2,500-gram threshold. The largest positive estimate is a 0.519-percentage-point increase in births weighing 2,500–2,999 grams, while the 1,500–1,999 gram category rises by 0.226 percentage points. Births in the 3,000–3,499 and 3,500–3,999 gram ranges fall by 0.300 and 0.365 percentage points. Taken at face value, the pattern is more consistent with movement from the middle and upper-middle portions of the distribution toward lower and borderline-low birth weights than with a discrete increase concentrated immediately below 2,500 grams.

Every bin-specific Algorithm 4 interval includes zero, and the complete distribution is not jointly unusual relative to the fully refitted placebo states. The maximum standardized effect has rank 0.451 and the quadratic statistic has rank 0.490. Thus, neither one exceptionally large bin movement nor the distribution-wide pattern is distinguishable from the variation generated by placebo assignments. Figure 3 is therefore useful for locating the point-estimate pattern and for any later distributional valuation, but it does not identify a particular birth-weight range as being affected with statistical precision.

4.4 Pregnancy Accounting and Composition

The statewide health estimates condition on an observed live birth. They can therefore change either because I-1183 affected health within pregnancies that continued to delivery or because it changed which pregnancies ultimately appeared in the live-birth sample (Fertig and Watson, 2009). If the fetal-growth pattern were driven primarily by a substantial change in the observed birth pool, I would expect corresponding movement in recorded pregnancy totals, fetal survival, or the composition of mothers giving birth. Table 2 examines these margins using the same statewide counterfactual and one-treated-state inference as the main health analysis.

The pregnancy-accounting measures provide little evidence of a large shift in the pregnancies recorded by NVSS. Live births and reportable pregnancies each decline by 1.1 per 100,000 population, while reportable fetal deaths rise by 0.03 per 100,000 and fetal deaths per 1,000 reportable pregnancies rise by 0.33. All four Algorithm 4 intervals include zero, and the estimates do not reveal a coherent change in either the number of recorded pregnancies or later fetal survival.

Maternal composition is similarly stable. The estimated changes in maternal age, marital status, race, ethnicity, and father-information reporting do not form a consistent pattern, and each is imprecisely estimated. These characteristics are treated as outcomes rather than controls: the purpose is to ask whether privatization changed the composition of the observed birth pool, not to condition away a potentially endogenous response to the reform.

Taken together, the observed selection margins provide little support for a simple expla-

Table 2: Pregnancy Accounting and Maternal Composition

Outcome	Estimate	95% CI	% change	Effect-rank
Panel A. Pregnancy accounting and later fetal selection				
Reportable fetal deaths / 100,000	0.03	[-0.14, 0.19]	4.1%	0.647
Live births / 100,000	-1.1	[-5.3, 3.1]	-1.1%	0.451
Reportable pregnancies / 100,000	-1.1	[-5.3, 3.1]	-1.0%	0.451
Fetal deaths / 1,000 reportable pregnancies	0.33	[-1.11, 1.77]	5.3%	0.569
Panel B. Composition among observed live births				
Teen mother (percentage points)	0.25	[-0.61, 1.10]	4.1%	0.608
Maternal age (years)	0.04	[-0.13, 0.21]	0.1%	0.667
Married mother (percentage points)	0.29	[-1.11, 1.70]	0.4%	0.667
Black mother (percentage points)	-0.08	[-1.09, 0.92]	-1.7%	0.882
Hispanic mother (percentage points)	-0.31	[-1.53, 0.91]	-1.7%	0.569
Father information present (percentage points)	-12.85	[-61.42, 35.73]	-14.1%	0.451

Notes: Panel A reports population-normalized live births, reportable fetal deaths, their sum as reportable pregnancies, and fetal deaths per 1,000 reportable pregnancies. Reportable fetal deaths satisfy the NCHS tabulation criterion and have valid gestational age for assignment to estimated-fertilization cohorts. Panel B reports maternal characteristics among observed live births; these variables are treated as outcomes used to assess changes in the observed birth pool rather than as controls in the primary health specification. Reportable pregnancies do not include abortions, early pregnancy losses, or conceptions that never enter the natality or fetal-death records. Valid estimated-fertilization timing is available for 48.2 percent of audited fetal-death records, and NCHS changed its preferred gestational-age standard beginning in 2014; the fetal-death estimates are therefore interpreted as supporting evidence. All estimates use the preferred common RC-SDiD donor and time weights. Confidence intervals use Algorithm 4 full-refit placebo variance across the 50 non-Washington assignments. Effect ranks are separate two-sided placebo diagnostics and are not exact randomization-inference p-values.

nation in which a large change in the recorded pregnancy or maternal pool generates the adverse fetal-growth point estimates. They cannot, however, rule out selection more broadly. NVSS does not observe induced abortions, early pregnancy losses, or conceptions that never enter the natality or fetal-death files. Estimated fertilization timing is also available for only 48.2 percent of the audited fetal-death records, and residence at delivery does not identify migration during pregnancy. The statewide health estimates should therefore continue to be interpreted as effects among observed live births, combining any prenatal-health response with selection on margins that the available vital-statistics data cannot fully observe.

4.5 Within-Washington Retail Access

The within-Washington analysis first asks whether predetermined retail capacity predicted where I-1183 generated larger increases in short-distance spirits access. Figure 4 shows a clear first stage at very short distances. One standard deviation greater pre-policy grocery exposure predicts a 9.4-percentage-point larger increase in population coverage within 0.5 miles and a 12.45-percentage-point larger increase within one mile. The corresponding squared clustered first-stage t-statistics are 18.27 and 27.63. At longer distances, however, the coverage response disappears. Five-mile coverage was already 93.1 percent before privatization, leaving little scope for further expansion, and the first stages at three, five, and ten miles are close to zero. I therefore use one-mile population coverage as the preferred access margin.

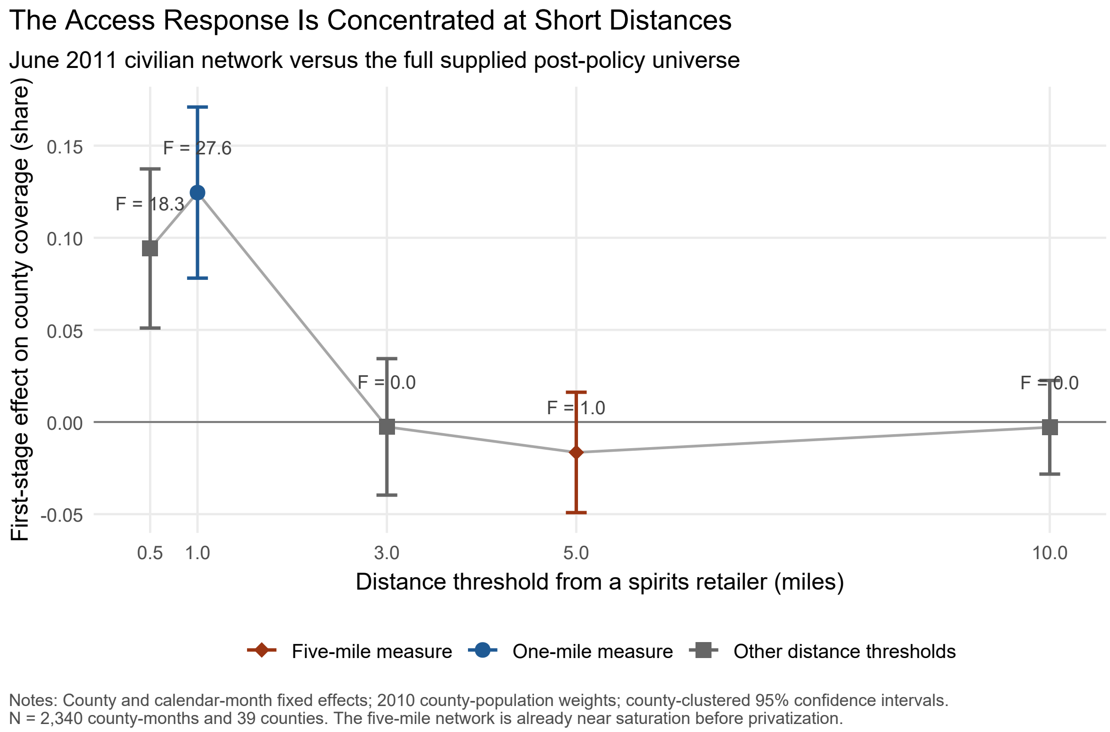

Figure 4: Within-Washington First Stage Across Access Measures

Notes: Points report coefficients on the interaction between the post-I-1183 period and standardized pre-policy grocery coverage in county and calendar-month fixed-effects regressions for 39 Washington counties and 60 months. Outcomes compare fixed June 2011 and supplied post-privatization retailer networks using fixed 2010 block-group population centroids. Regressions use county population weights and cluster standard errors by county. Bars are 95 percent confidence intervals; labels report the squared clustered t-statistic for the excluded interaction.

The one-mile relationship is not driven by a single county. Leave-one-county-out estimates range from 11.03 to 13.30 percentage points. The first stage also remains positive without population weights, although it weakens substantially, indicating that the population-weighted relationship is strongest in counties containing more Washington residents. Taken together, these results show that I-1183 changed the geography of spirits retailing primarily by expanding very short-distance population coverage in places with substantial pre-existing grocery infrastructure, rather than by expanding access uniformly at all distances.

The preferred one-mile first stage is also robust to alternative ways of measuring local access. Reconstructing one-mile coverage from populated 2010 Census blocks rather than block-group centroids produces a similar first stage and nearly identical IV rescalings. Defining treatment instead as the share of the fixed population newly brought within one mile produces an even stronger first stage and similar health estimates. By contrast,

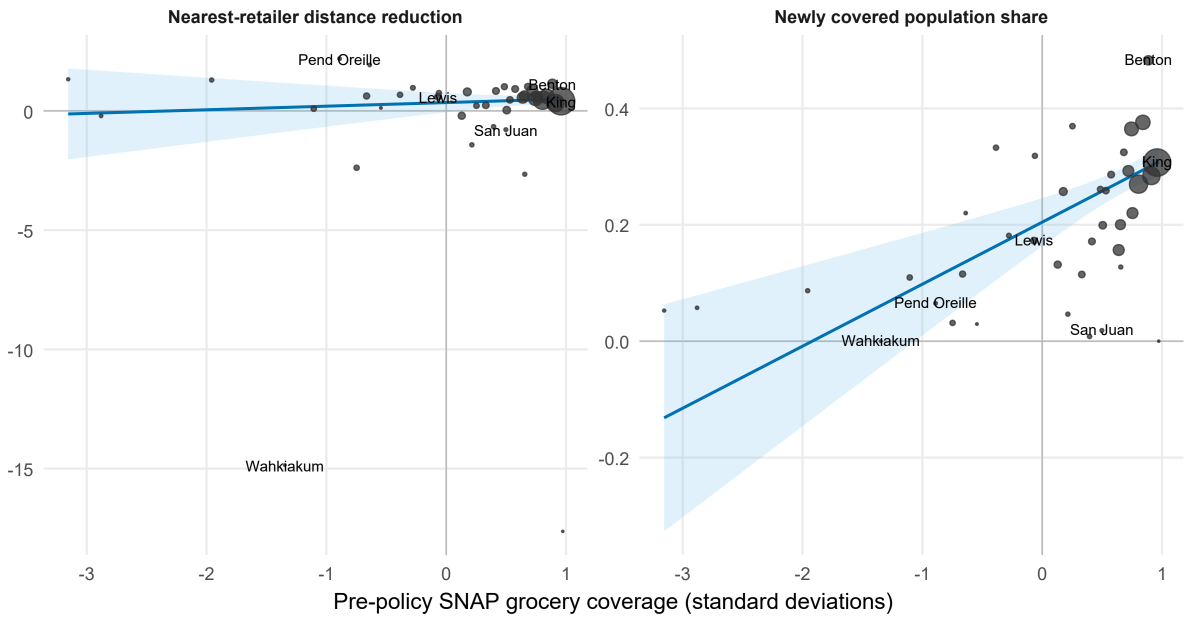

Figure 5: Within-Washington First Stages for Alternative Access Changes

Notes: The left panel reports the population-weighted reduction in mean nearest-retailer distance, defined as pre-policy minus post-policy distance in miles. The right panel reports the fixed 2010 population share newly covered within one mile. Both compare fixed June 2011 and supplied post-privatization retailer networks. Horizontal axes show standardized pre-policy SNAP grocery coverage within five miles. Points are the 39 Washington counties, sized by 2010 population; lines and shaded bands show population-weighted linear fits and 95 percent confidence intervals. Distances permit the nearest retailer to lie across a county boundary.

the population-weighted reduction in nearest-retailer distance is only weakly predicted by pre-policy grocery infrastructure and does not support a precise IV scaling. Appendix B reports these measurement comparisons, including Anderson–Rubin inference for the weakly identified distance specification.

I next ask whether birth outcomes changed along the same predetermined policy-exposure gradient. The reduced form is the most transparent version of this comparison. A one-standard-deviation increase in pre-policy grocery exposure is associated with a 17.62-gram decline in birth weight after I-1183 and a 0.632-percentage-point increase in SGA. Other outcomes are less consistent: LGA falls by 0.380 percentage points, while preterm birth falls by 0.254 percentage points and gestational duration changes little. The broader outcome family therefore does not reproduce a uniform adverse health response.

Table 3 also reports the corresponding 2SLS estimates, which rescale these reduced-form differences by the policy-induced change in one-mile access.

Table 3: Within-Washington Policy Exposure and One-Mile Birth-Outcome Estimates

Outcome	Reduced form	IV per 10 pp access	First-stage F
Birth weight (grams)	-17.62 (9.07)	-11.02 (6.70)	16.6
Small for gestational age (pp)	0.632 (0.292)	0.395 (0.223)	16.6
Large for gestational age (pp)	-0.380 (0.447)	-0.237 (0.299)	16.6
Preterm birth (pp)	-0.254 (0.346)	-0.159 (0.214)	16.6
Gestation (weeks)	-0.027 (0.029)	-0.017 (0.021)	16.6
Low APGAR5 (pp)	0.453 (0.364)	0.283 (0.218)	16.6

Notes: The reduced form is the county-month coefficient on the post-I-1183 indicator interacted with standardized pre-policy grocery access. The IV column instruments the fixed June-2011-to-post-policy change in one-mile county population coverage with that interaction and scales estimates to a 10-percentage-point instrument-induced coverage increase. Standard errors clustered by county are in parentheses. Rows span the prespecified outcome families; Appendix B reports all 12 outcomes and both timing constructions. These are assumption-dependent county-level policy-exposure/IV estimates, not individual proximity effects or independently established causal prenatal effects.

For birth weight, a 10-percentage-point instrument-predicted increase in one-mile population coverage corresponds to a decline of 11.02 grams. The analogous SGA estimate is an increase of 0.395 percentage points. The outcome-sample first-stage statistic is approximately 16.6, smaller than the 27.63 statistic in the full access panel because the IV regressions use outcome-specific birth samples and weights. These magnitudes are consistent with the adverse fetal-growth pattern in the statewide point estimates, but the remaining outcomes are mixed and the IV interpretation requires substantially stronger assumptions than the statewide design.

Those assumptions are consequential in the data. Pre-policy grocery exposure is related to county population, urbanization, baseline health, and some differential outcome trends.

Table 4: Within-Washington Birth Outcomes under Alternative Access Measurements

Outcome	Reduced form	Wild p	Canonical one-mile IV	Block one-mile IV	Newly covered IV	Distance-reduction IV
Birth weight (grams)	-17.620 (9.074)	0.158	-11.020 [-24.609, 2.569]	-10.699 [-23.874, 2.476]	-12.870 [-28.186, 2.447]	-75.431 [-323.663, 172.801]
Low birth weight (pp)	0.128 (0.273)	0.658	0.080 [-0.281, 0.441]	0.077 [-0.272, 0.427]	0.093 [-0.323, 0.509]	0.546 [-2.667, 3.759]
Small for gestational age (pp)	0.632 (0.292)	0.076	0.395 [-0.058, 0.848]	0.384 [-0.037, 0.804]	0.461 [-0.013, 0.936]	2.695 [-5.761, 11.150]
Term low birth weight (pp)	0.161 (0.174)	0.383	0.102 [-0.133, 0.336]	0.099 [-0.123, 0.320]	0.119 [-0.141, 0.379]	0.697 [-1.882, 3.275]
Gestation (weeks)	-0.027 (0.029)	0.367	-0.017 [-0.059, 0.025]	-0.016 [-0.056, 0.023]	-0.020 [-0.066, 0.026]	-0.116 [-0.605, 0.373]

Notes: The reduced form is reported first with its county-clustered standard error in parentheses and the prespecified 9,999-draw restricted wild-cluster p-value. IV cells report conventional 2SLS estimates with 95 percent county-clustered confidence intervals in brackets. Binary outcomes are percentage points; one-mile and newly-covered IV effects are scaled to a 10-percentage-point access increase; distance reduction is per mile. All four columns use the preferred block-group SNAP instrument, so the reduced form is identical and only the first-stage scaling changes. The distance-reduction mapping is weakly identified; its conventional interval is not sufficient for inference. The complete prespecified 12-outcome family and Anderson–Rubin sets appear in Table B.9. These county-level Appendix-B sensitivities are not individual proximity or prenatal-drinking effects.

The reduced-form birth-weight and SGA estimates have wild-cluster p-values of 0.160 and 0.072, respectively, and specifications allowing county-specific trends yield p-values of 0.105 and 0.881. Across the broader outcome family, the apparent health relationships are not robust to the complete set of pretrend, placebo-date, county-trend, multiplicity, and small-cluster diagnostics reported in Appendix B. The contrast between the strong access first stage and the substantially weaker health evidence is therefore important: predetermined retail capacity clearly predicts where the reform expanded very local access, but the same design does not provide comparably strong evidence that those access changes caused worse birth health.

I therefore interpret the within-Washington results primarily as evidence about the access channel of I-1183 rather than as independent confirmation of a causal prenatal-health mechanism. The reduced form shows how county birth outcomes changed with predetermined exposure to the reform, while 2SLS expresses that relationship in units of policy-induced one-mile access under the additional exclusion and treatment-mapping assumptions. Neither estimand identifies whether a particular mother gained a nearby retailer, where she purchased alcohol, whether she drank during pregnancy, or a biological dose response to prenatal alcohol exposure. Appendix B reports the complete outcome family, alternative access constructions, geocoding and weighting checks, balance and pretrend diagnostics, placebo dates, county-specific trends, influential-county analyses, multiplicity adjustments, wild-cluster inference, and weak-IV-robust inference where relevant.

5 From Reduced-Form Effects to Welfare-Relevant Sufficient Statistics

The empirical analysis identifies two policy responses to I-1183 that are useful for evaluating the welfare consequences of alcohol-market liberalization: a large change in household alcohol purchasing and a set of downstream birth-health reduced forms. These estimates answer different parts of the policy problem. The purchasing results describe how the alcohol market responded to the reform, while the health estimates describe how outcomes among observed live births changed under the same policy. Neither object, by itself, provides a welfare calculation.

I use a sufficient-statistics approach to organize what can be learned from these reduced-form estimates and what additional information is required for welfare analysis. The motivation follows Chetty (2009), who emphasizes that many welfare questions can be answered without recovering every primitive of a fully specified structural model. Instead, economic theory can identify a smaller set of empirically recoverable objects that are sufficient for the policy question. This approach is particularly useful here because I-1183 provides quasi-experimental variation in both alcohol purchasing and downstream health outcomes while leaving several components of the behavioral and market environment only partially observed.

For the corrective-tax application, I build first on Griffith et al. (2019). Their framework emphasizes that optimal alcohol taxes depend on the relationship between heterogeneous marginal external harms and consumers' substitution across differentiated alcohol products. The present setting adds an important supply-side complication. I-1183 changed the competitive environment itself, so the mapping from a statutory tax to consumer prices cannot be treated as fixed. Drawing on Weyl and Fabinger (2013), I therefore incorporate market power and equilibrium tax pass-through into the broader sufficient-statistics problem. Finally, following the logic of Allcott et al. (2019), I distinguish harms borne by others from private harms that consumers may fail to internalize.

The purpose of this section is not to report an optimal alcohol tax. It is to show how the reduced-form evidence estimated in this paper can be translated into one welfare-relevant input and how that input fits into the broader corrective-policy problem. I first construct the birth-related marginal external-cost input in three steps. I map the birth-health estimates into economically valued consequences, distinguish total health harm from the portion that constitutes an external cost, and scale that external cost by the policy-induced change in alcohol purchasing. I then show how this empirical object enters the broader optimal-tax problem.

This distinction is important because a reduced-form health effect does not itself reveal the economic cost of the reform, and an aggregate increase in alcohol purchasing does not identify the exposure margin responsible for that effect. Moreover, I-1183 changed several margins simultaneously: retail access, shopping convenience, product assortment, private pricing, statutory fees, and market structure. The statewide reduced form therefore identifies the effect of moving from the pre-I-1183 regime to the post-I-1183 regime, not the effect of a marginal change in a single alcohol tax and not the biological effect of an additional liter of prenatal ethanol exposure.

Even so, the reform provides useful inputs to the sufficient-statistics exercise. The birth-health estimates can be translated into a policy-induced change in health costs, while the Homescan estimates measure the associated change in alcohol purchasing. Moving from those objects to corrective tax design additionally requires information on substitution across products and, because the post-I-1183 market is imperfectly competitive, the equilibrium markups and tax pass-through generated by that market structure.

5.1 Birth-Health Effects as a Welfare Input

The first step is to translate the reduced-form health estimates into outcomes that can be economically valued. Mortality has a relatively direct mapping: an estimated change in deaths can be combined with a value of statistical life. Birth health is less straightforward. There is no natural welfare interpretation of assigning a constant dollar value to each gram of birth weight, particularly because the medical and longer-run consequences of low birth weight are highly nonlinear (Almond et al., 2004; Black et al., 2005).

I therefore use the estimated birth-weight distribution rather than the mean birth-weight effect as the primary basis for valuation. Let $h \in \mathcal{H}$ index the mutually exclusive birth-weight ranges reported in Figure 3, and let $\widehat{\tau}_h$ denote the estimated I-1183 effect on the share of observed Washington births in range h. If N_{WA}^{post} denotes the number of observed Washington births in the relevant post-policy cohorts, the corresponding estimated change in the number of births in each range is

$$\Delta\widehat{N}_h = \widehat{\tau}_h N_{WA}^{post}. \tag{9}$$

The economic valuation attaches consequences to these changes in health states rather than to grams of birth weight themselves. Let C_h denote the incremental resource cost associated with a birth in range h, and let L_h^Q denote the associated loss in quality-adjusted life years relative to a common reference state. The valued birth-health consequence is

$$\Delta\widehat{C}^{birth} = \sum_{h\in\mathcal{H}} \Delta\widehat{N}_h \left(C_h + V_Q L_h^Q\right), \tag{10}$$

where V_Q is the monetary value assigned to one QALY.

The two components capture different consequences of poorer health at birth. C_h measures incremental resource use, such as additional medical care, while $V_Q L_h^Q$ values morbidity and reductions in health or functioning not captured by those expenditures. Fiscal consequences can also be reported separately when relevant, but a transfer is not itself a social resource cost. Likewise, resource-cost and QALY components should be constructed so that the same consequence is not valued twice.

Using mutually exclusive birth-weight ranges also avoids double counting across the paper's overlapping health outcomes. Mean birth weight, low birth weight, SGA, and the birth-weight bins describe related features of the same underlying distribution and should not all be monetized separately. Measures such as SGA remain useful for interpreting the reduced-form evidence, but the baseline valuation uses one exhaustive set of birth-health states. Additional outcomes enter only when they represent distinct consequences. Infant mortality, for example, can be valued separately using a VSL; fetal mortality requires additional assumptions and is better treated as a separate sensitivity.

5.2 From Health Harm to External Cost

The valuation in Section 5.1 measures the economic consequences associated with the estimated change in birth health. Those consequences are not automatically external costs. Prospective parents may internalize some expected effects of alcohol use on a future child through altruism or other preferences, while other consequences may be borne by the child, insurers, taxpayers, or other parties. Incomplete information, alcohol use before pregnancy recognition, and other behavioral frictions may further prevent the full consequences of alcohol use from entering the private decision.

Let $\theta \in [0, 1]$ denote the share of the valued birth-health consequence that is not internalized in the relevant alcohol-use decision. The corresponding birth-related external cost is

$$\Delta\widehat{EC}^{birth} = \theta\Delta\widehat{C}^{birth} = \theta \sum_{h\in\mathcal{H}} \Delta\widehat{N}_h \left(C_h + V_Q L_h^Q\right). \tag{11}$$

I do not estimate θ. Instead, the valued birth-health consequence can be reported first, with the external-cost calculation evaluated under alternative assumptions about internalization. This separation matters because the marginal external-cost object that enters corrective tax design refers to harm not already incorporated into private decisions, rather than to total health harm generated by alcohol consumption.

Private internalities are conceptually separate. If consumers fail to account fully for harms they themselves bear because of present bias, addiction, imperfect information, or related behavioral frictions, those uninternalized private harms provide an additional corrective motive (Allcott et al., 2019). External costs and internalities therefore enter the broader corrective-policy problem as distinct objects.

5.3 The Quantity Denominator and Behavioral Incidence

The final step in constructing the birth-related sufficient statistic is to scale the estimated external cost by the behavioral response induced by I-1183. Because the birth-health mechanism operates through alcohol exposure rather than spirits consumption specifically, I use the change in total ethanol-equivalent purchasing as the baseline quantity response. This measure incorporates changes across spirits, beer, and wine rather than treating the increase in the directly liberalized product as the relevant first stage.

The distinction matters empirically. I-1183 increased spirits purchases by 0.047 ethanol-equivalent liters per projected household per month, while offsetting changes in beer and wine leave an estimated increase in total ethanol-equivalent purchases of 0.035 liters. Let $\Delta\widehat{A}^E$ denote the aggregate policy-induced change in ethanol-equivalent purchases over the same horizon used to construct the birth-health cost estimate. A Wald-style estimate of the birth-related marginal external cost is then

$$\widehat{MEC}^{birth} = \frac{\Delta\widehat{EC}^{birth}}{\Delta\widehat{A}^E}. \tag{12}$$

Equation (12) expresses the estimated birth-related external cost associated with I-1183 per additional ethanol-equivalent liter purchased. Under the maintained mapping from the observed purchasing response to alcohol exposure, this provides an empirical estimate of the birth-health component of the marginal external-cost object used in the sufficient-statistics framework below.

The interpretation depends on behavioral incidence. Homescan measures household purchases rather than individual consumption, while the biologically relevant first stage for birth health is alcohol exposure among people and periods in which drinking can affect a pregnancy. The household-demographic and survey evidence in Section 4.1 provides partial information about this margin but does not directly identify prenatal alcohol exposure. Equation (12) should therefore be interpreted as a policy-induced marginal external-cost estimate based on the observed aggregate purchasing response, not as a biological dose-response estimate for prenatal ethanol consumption.

This distinction also helps interpret the reduced-form results. A large aggregate market response need not produce a large birth-health effect if the additional alcohol purchasing occurs primarily outside the pregnancy-relevant margin. Conversely, a smaller aggregate response could generate substantial birth-related harm if it is concentrated among consumers and periods with high pregnancy-related marginal harm. The incidence of the behavioral response therefore determines how the aggregate market first stage maps into the birth-related marginal external cost.

5.4 Embedding the Evidence in an Optimal-Tax Condition

Estimating a birth-related marginal external cost is not sufficient to determine the optimal alcohol tax. The alcohol market contains many differentiated products, consumers substitute across them, and the consumers whose demand responds most strongly to a tax need not be those with the highest marginal external harms. Griffith et al. (2019) show that these features matter for corrective alcohol-tax design: when marginal externalities and demand responses are heterogeneous, optimal product-level tax rates depend on the relationship between the two, not simply on average alcohol demand or average external harm.

Their benchmark deliberately abstracts from equilibrium supply-side pricing by assuming complete tax pass-through. That simplification is less appropriate for Washington. I-1183 replaced state retail price setting with private retailing and induced substantial entry, changing the competitive environment in which alcohol taxes are transmitted to consumers. Under imperfect competition, pass-through is an equilibrium outcome rather than a mechanical property of the tax. As emphasized by Weyl and Fabinger (2013), it depends on the shape of demand and the conduct of firms. I therefore extend the Griffith et al. corrective-tax framework by allowing the existing price-cost wedge and the equilibrium pass-through of alternative tax instruments to reflect the post-privatization market structure.

Let $j \in \mathcal{J}$ index alcohol products or beverage categories, q_{ij} consumer i's demand for product j, p_j its consumer price, c_j its marginal resource cost, and z_j its ethanol content. Let $E_i(Z_i)$ denote the external cost generated by consumer i as a function of total ethanol consumption Z_i, so that $E_i'(Z_i)$ is marginal external harm per unit of ethanol. For any tax instrument t_k, the welfare effect of a marginal tax change can be written

$$\frac{\partial W}{\partial t_k} = \sum_i \sum_{j \in \mathcal{J}} [p_j - c_j - z_j E_i'(Z_i)] \frac{\partial q_{ij}}{\partial t_k}. \tag{13}$$

At an interior optimum,

$$\sum_i \sum_{j \in \mathcal{J}} \left[p_j - c_j - z_j E_i'(Z_i)\right] \frac{\partial q_{ij}}{\partial t_k} = 0 \qquad \forall k. \tag{14}$$

Equation (14) preserves the central insight of Griffith et al. (2019): marginal external harm must be weighted by the behavioral response induced by the tax. The birth-related marginal external cost estimated above provides empirical information about one component of the marginal external-harm schedule; traffic injuries, crime, other health consequences, and additional external harms would be required for a complete alcohol externality measure. Heterogeneity also remains important. A tax is more effective at correcting a particular externality when it disproportionately reduces the ethanol consumption of consumers with high marginal external harm.

The first supply-side extension is that the existing price-cost wedge need not equal the statutory tax wedge. Because $p_j - c_j$ is the difference between the consumer price and marginal resource cost, it includes both statutory taxes and equilibrium markups. Under imperfect competition, equilibrium markups create an additional wedge between consumer prices and marginal resource costs. Corrective policy therefore operates in a market with a pre-existing distortion generated jointly by taxation and market power.

The second extension concerns how a change in the statutory tax reaches consumers. Tax pass-through determines how a change in t_k produces the behavioral response appearing in Equation (14). Define

$$\rho_{mk} = \frac{\partial p_m}{\partial t_k} \tag{15}$$

as the equilibrium pass-through of tax instrument k into the consumer price of product m. The demand response can then be decomposed as

$$\frac{\partial q_{ij}}{\partial t_k} = \sum_{m \in \mathcal{J}} \frac{\partial q_{ij}}{\partial p_m} \rho_{mk}. \tag{16}$$

This decomposition separates two distinct empirical objects. The demand derivatives describe consumers' own- and cross-product substitution in response to prices. The pass-through terms describe how firms translate a particular tax change into those prices. Under imperfect competition, the latter depend on the competitive environment and the shape of demand (Weyl and Fabinger, 2013). A change in market structure can therefore alter the effect of a tax even when consumer preferences and the statutory tax change are held fixed.

That point is particularly important for I-1183. Privatization did not simply change alcohol availability while leaving the supply side fixed. It moved spirits retailing from a state-controlled system to a market with extensive private entry and decentralized pricing.

The relevant post-I-1183 tax counterfactual must therefore account for the market structure through which the tax is passed to consumers. In particular, the pass-through associated with a given specific or ad valorem tax need not be the same as it would have been under the pre-privatization retail regime. The policy problem consequently requires both the price-cost wedge generated by the new competitive environment and the equilibrium pass-through of each tax instrument within that environment.

The familiar Pigouvian rule is recovered as a special case. Suppose the market is competitive so that the only wedge between the consumer price and marginal resource cost is the statutory tax, and suppose marginal external harm per unit of ethanol is homogeneous across consumers on the relevant margin. If policy can independently implement the required wedge on each beverage, Equation (14) reduces to

$$\omega_j = z_j \bar{e}, \tag{17}$$

where ω_j is the statutory wedge on product j and $\bar{e}$ is marginal external harm per unit of ethanol. In this competitive benchmark, the optimal statutory wedge is pinned down by marginal external harm. Tax pass-through instead governs how the statutory change is divided between consumer and producer prices and how quantities respond. Once market power is present, the existing markup and equilibrium pass-through both become relevant for evaluating corrective tax changes.

The Wald-style estimate in Equation (12) supplies empirical information about the birth-health component of $E_i'(Z_i)$. It does not by itself complete Equation (14). The remaining objects are the other components of marginal external harm, consumers' own- and cross-price demand responses, the existing tax and markup wedges, and the equilibrium pass-through of the available tax instruments. Appendix C develops this supply-side extension formally and describes how these objects can be combined to evaluate corrective tax counterfactuals in the post-I-1183 market.

Finally, Equation (14) is written in terms of external harms. If consumers also fail to internalize private harms because of present bias, addiction, imperfect information, or related behavioral frictions, the corresponding internality enters as an additional component of the corrective wedge rather than as part of $E_i'(Z_i)$. The sufficient-statistics exercise therefore places the reduced-form evidence within a broader corrective-policy problem. I-1183 provides a policy-induced change in alcohol purchasing and a downstream birth-health response that can be valued and combined to estimate one component of marginal external harm. Corrective tax design additionally requires the remaining alcohol-related external harms, consumers' substitution across differentiated products, the price-cost wedges generated by existing taxes and market power, and the equilibrium pass-through of the available tax instruments. In Washington, the latter objects are particularly important because privatization changed the competitive environment through which taxes are transmitted to consumers. Appendix C develops this mapping formally.

6 Conclusion

A large literature shows that alcohol demand responds to prices, legal restrictions, and retail availability. Less is known about what "access" means when an entire market is liberalized and about how the resulting demand response translates into a particular population-level harm. Liberalization can reduce the generalized cost of acquiring a product through several margins at once: shorter travel, more sellers, broader assortment, longer hours, and the ability to combine purchases with other shopping. Together, these margins can generate a substantial demand response even without a large change in any single component of access. Understanding the effects of liberalization therefore requires knowing not just whether access expands, but how it expands and who responds to that change. Those two margins determine whether a large market response translates into measurable changes in birth health.

Washington's I-1183 provides a setting in which to trace both links. Privatization sharply expanded spirits retailing, but the access response was concentrated at very short distances. One-mile population coverage increased substantially where pre-existing grocery infrastructure was greatest, while five-mile coverage was already nearly saturated. Consumers also responded strongly to the new market environment, with recorded household spirits purchases rising by 35 percent. Yet substitution across beverages reduced the increase in total ethanol-equivalent purchasing to 9.2 percent, and the available household evidence provides little indication that this response was concentrated among households in prime childbearing ages. The reform therefore generated a large change in the spirits market without producing an equally large change in either total ethanol purchasing or the exposure margin most relevant for birth health.

The birth-health evidence is more muted. Statewide point estimates suggest some movement toward lower fetal growth among observed live births, including lower average birth weight, higher low-birth-weight and SGA rates, and lower LGA rates. At the same time, every primary sampling interval includes zero, the prespecified fetal-growth and gestational-duration index is near zero, and neither the fetal-growth family nor the complete birth-weight distribution is jointly unusual relative to the fully refitted placebo states. Pregnancy accounting and maternal composition also show no large, coherent shift in the observed birth pool. Taken together, the adverse point estimates leave open the possibility of modest or concentrated effects, but the evidence does not establish a broad deterioration in birth health following privatization.

The within-Washington health analysis reinforces this interpretation. Pre-I-1183 grocery infrastructure strongly predicts where very local spirits access expanded after privatization. While birth weight and SGA move adversely along that policy-exposure gradient, the broader outcome family is mixed and the health estimates are sensitive to differential-trend, placebo, multiplicity, and small-cluster diagnostics. The local analysis therefore establishes a clear retail-access response to privatization, but does not provide comparably strong evidence of a causal effect on birth health.

These results have two implications for evaluating alcohol-market liberalization. First, broad measures of availability may miss the margin on which consumers actually experience a reform. In Washington, the important change was not whether spirits were available somewhere within several miles, but whether they became available very nearby and through retailers already integrated into routine shopping. Second, a large increase in spirits purchasing does not necessarily imply an equally large increase in alcohol exposure. Consumers can substitute across beer, wine, and spirits, so the change in total ethanol can be much smaller than the change in spirits alone. Even then, the health consequences depend on who accounts for that additional consumption. For corrective policy, the relevant object is therefore the relationship between policy-induced changes in alcohol consumption and the marginal harms associated with the consumers generating those changes.

There are three primary limitations. First, Homescan measures household purchases rather than individual consumption, and neither the statewide nor within-Washington design directly observes alcohol use during pregnancy. Second, the vital-statistics analysis conditions on observed live births and therefore cannot capture selection through abortion or early pregnancy loss. Third, I-1183 changed prices, convenience, assortment, entry, and market structure simultaneously, so the statewide estimates identify the effect of the reform as a bundle rather than the causal effect of any single margin of access. These limitations constrain how directly the estimates can be mapped from market behavior to prenatal exposure and birth health, while also pointing to important directions for future research.

Future work should consider both where the behavioral response is strongest and how access is measured. The household-purchasing evidence points toward larger responses among older households, suggesting that health and social outcomes concentrated at older ages may be especially informative for studying the consequences of privatization. Future work should also examine whether consumers with higher baseline alcohol purchasing respond differently to expanded convenience and retail access. Better measurement is equally important. Linking finer measures of retail access to individual purchasing or consumption would help distinguish proximity, one-stop shopping, assortment, and other dimensions of access more directly.

More generally, the Washington experience shows why the incidence of a policy response is central to its welfare consequences. Liberalization can substantially alter markets and consumer behavior without producing equally large changes in every downstream outcome. Understanding those consequences requires tracing the response from the form of access that actually changes, through substitution and the consumers who respond, to the harms associated with their marginal consumption. That is the link between market liberalization and social cost that aggregate demand alone cannot provide.

References

A. Ahammer, S. Bauernschuster, M. Halla, and H. Lachenmaier. Minimum legal drinking age and the social gradient in binge drinking. *Journal of Health Economics*, 81:102571, Jan. 2022. ISSN 0167-6296. doi: 10.1016/j.jhealeco.2021.102571. URL https://www.sciencedirect.com/science/article/pii/S0167629621001569.

H. Allcott, B. B. Lockwood, and D. Taubinsky. Regressive Sin Taxes, with an Application to the Optimal Soda Tax*. *The Quarterly Journal of Economics*, 134(3):1557–1626, Aug. 2019. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qjz017. URL https://academic.oup.com/qje/article/134/3/1557/5499049.

D. Almond, K. Y. Chay, and D. S. Lee. The Costs of Low Birth Weight, June 2004. Number: 10552 Series: Working Paper Series tex.archiveprefix: National Bureau of Economic Research tex.eprint: 10552.

American Academy of Pediatrics Committee on Fetus and Newborn and American College of Obstetricians and Gynecologists Committee on Obstetric Practice. The apgar score. *Pediatrics*, 136(4):819–822, 2015. doi: 10.1542/peds.2015-2651. URL https://doi.org/10.1542/peds.2015-2651.

I. M. Aris, K. P. Kleinman, M. B. Belfort, A. Kaimal, and E. Oken. A 2017 US reference for singleton birth weight percentiles using obstetric estimates of gestation. *Pediatrics*, 144(1):e20190076, 2019. doi: 10.1542/peds.2019-0076. URL https://doi.org/10.1542/peds.2019-0076.

D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, 2021. doi: 10.1257/aer.20190159.

S. B. L. Barnett, N. B. Coe, J. R. Harris, and A. Basu. Washington's privatization of liquor: Effects on household alcohol purchases from Initiative 1183. *Addiction (Abingdon, England)*, 115(4):681–689, Apr. 2020. ISSN 0965-2140. doi: 10.1111/add.14875. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8509083/.

G. S. Becker and R. J. Barro. A Reformulation of the Economic Theory of Fertility. *The Quarterly Journal of Economics*, 103(1):1–25, 1988. ISSN 0033-5533. doi: 10.2307/1882640. URL https://www.jstor.org/stable/1882640.

S. E. Black, P. J. Devereux, and K. Salvanes. From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes, Nov. 2005. Number: 11796 Series: Working Paper Series tex.archiveprefix: National Bureau of Economic Research tex.eprint: 11796.

K. Buckles, D. Hungerman, and S. Lugauer. Is fertility a leading economic indicator? *The Economic Journal*, 131(634):541–565, 2021. doi: 10.1093/ej/ueaa068. URL `https://doi.org/10.1093/ej/ueaa068`.

C. Carpenter and C. Dobkin. The Minimum Legal Drinking Age and Public Health. *Journal of Economic Perspectives*, 25(2):133–156, June 2011. ISSN 0895-3309. doi: 10.1257/jep.25.2.133. URL `https://www.aeaweb.org/articles?id=10.1257/jep.25.2.133`.

R. Chetty. Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics*, 1(Volume 1, 2009): 451–488, Sept. 2009. ISSN 1941-1383, 1941-1391. doi: 10.1146/annurev.economics.050708.142910. URL `https://www.annualreviews.org/content/journals/10.1146/annurev.economics.050708.142910`.

G. Cil. Effects of posted point-of-sale warnings on alcohol consumption during pregnancy and on birth outcomes. *Journal of Health Economics*, 53:131–155, May 2017. ISSN 0167-6296. doi: 10.1016/j.jhealeco.2017.03.004. URL `https://www.sciencedirect.com/science/article/pii/S0167629617302448`.

I. Cintina. The effect of minimum drinking age laws on pregnancy, fertility, and alcohol consumption. *Review of Economics of the Household*, 13(4):1003–1022, 2015. doi: 10.1007/s11150-014-9271-8. URL `https://doi.org/10.1007/s11150-014-9271-8`.

L. Einav, E. Leibtag, and A. Nevo. Recording discrepancies in nielsen homescan data: Are they present and do they matter? *Quantitative Marketing and Economics*, 8 (2):207–239, 2010. doi: 10.1007/s11129-009-9073-0. URL `https://doi.org/10.1007/s11129-009-9073-0`.

A. R. Fertig and T. Watson. Minimum drinking age laws and infant health outcomes. *Journal of Health Economics*, 28(3):737–747, May 2009. ISSN 0167-6296. doi: 10.1016/j.jhealeco.2009.02.006. URL `https://www.sciencedirect.com/science/article/pii/S0167629609000319`.

R. Griffith, M. O'Connell, and K. Smith. Tax design in the alcohol market. *Journal of Public Economics*, 172:20–35, Apr. 2019. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2018.12.005. URL `https://www.sciencedirect.com/science/article/pii/S0047272718302329`.

R. A. Hahn, J. C. Middleton, R. Elder, R. Brewer, J. Fielding, T. S. Naimi, T. L. Toomey, S. Chattopadhyay, B. Lawrence, and C. A. Campbell. Effects of Alcohol Retail Privatization on Excessive Alcohol Consumption and Related Harms. *American Journal of Preventive Medicine*, 42(4):418–427, Apr. 2012. ISSN 07493797. doi: 10.1016/j.amepre.2012.01.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0749379712000256.

G. Illanes and S. Moshary. Market Structure and Product Assortment: Evidence from a Natural Experiment in Liquor Licensure, Apr. 2020. URL https://www.nber.org/papers/w27016.

W. C. Kerr, E. Williams, and T. K. Greenfield. Analysis of Price Changes in Washington Following the 2012 Liquor Privatization. *Alcohol and Alcoholism (Oxford, Oxfordshire)*, 50(6):654–660, Nov. 2015. ISSN 0735-0414. doi: 10.1093/alcalc/agv067. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4608622/.

W. C. Kerr, E. Williams, Y. Ye, M. S. Subbaraman, and T. K. Greenfield. Survey estimates of changes in alcohol use patterns following the 2012 privatization of the washington liquor monopoly. *Alcohol and Alcoholism*, 53(4):470–476, 2018. doi: 10.1093/alcalc/agy004. URL https://doi.org/10.1093/alcalc/agy004.

Kilts Center for Marketing. NielsenIQ consumer panel dataset, 2026. URL https://www.chicagobooth.edu/research/kilts/datasets/nielseniq-consumer-panel-dataset.

J. Luukkonen, L. Junna, H. Remes, and P. Martikainen. The association of lowered alcohol prices with birth outcomes and abortions: A population-based natural experiment. *Addiction (Abingdon, England)*, 118(5):836–844, 2023. ISSN 1360-0443. doi: 10.1111/add.16119. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/add.16119. tex.copyright: © 2023 Society for the Study of Addiction.

J. A. Martin, M. J. K. Osterman, S. E. Kirmeyer, and E. C. W. Gregory. Measuring gestational age in vital statistics data: Transitioning to the obstetric estimate. Technical Report 5, National Center for Health Statistics, Hyattsville, MD, 2015. URL https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_05.pdf. Series: National vital statistics reports Volume: 64.

Y. Morin. Synthetic difference in differences for repeated cross-sectional data, 2024. URL https://arxiv.org/abs/2409.20199. arXiv: 2409.20199 [econ.EM].

National Alcohol Beverage Control Association. Structure of U.S. alcohol regulation, 2026. URL https://www.nabca.org/structure-of-us-alcohol-regulation.

National Institute on Alcohol Abuse and Alcoholism. Understanding fetal alcohol spectrum disorders, Dec. 2025. URL https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/understanding-fetal-alcohol-spectrum-disorders.

J. P. Nilsson. Alcohol Availability, Prenatal Conditions, and Long-Term Economic Outcomes. *Journal of Political Economy*, 125(4):1149–1207, Aug. 2017. ISSN 0022-3808. doi: 10.1086/692694. URL https://www.journals.uchicago.edu/doi/10.1086/692694.

C. Noelke, Y.-H. Chen, T. L. Osypuk, and D. Acevedo-Garcia. Economic downturns and inequities in birth outcomes: Evidence from 149 million US births. *American Journal of Epidemiology*, 188(6):1092–1100, 2019. doi: 10.1093/aje/kwz042. URL https://doi.org/10.1093/aje/kwz042.

Office for National Statistics. User guide to conception statistics, 2025. URL https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/methodologies/userguidetoconceptionstatistics.

A. Z. Phillips, H. P. Rodriguez, W. C. Kerr, and J. A. Ahern. Washington's liquor license system and alcohol-related adverse health outcomes. *Addiction (Abingdon, England)*, 116(5):1043–1053, 2021. doi: 10.1111/add.15234. URL https://doi.org/10.1111/add.15234.

G. Picone, J. MacDougald, F. Sloan, A. Platt, and S. Kertesz. The effects of residential proximity to bars on alcohol consumption. *International Journal of Health Care Finance and Economics*, 10(4):347–367, Dec. 2010. ISSN 1573-6962. doi: 10.1007/s10754-010-9084-0. URL https://doi.org/10.1007/s10754-010-9084-0.

J. Roth, P. H. C. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023. doi: 10.1016/j.jeconom.2023.03.008. URL https://doi.org/10.1016/j.jeconom.2023.03.008.

P. H. C. Sant'Anna and J. Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020. doi: 10.1016/j.jeconom.2020.06.003. URL https://doi.org/10.1016/j.jeconom.2020.06.003.

K. Seim and J. Waldfogel. Public Monopoly and Economic Efficiency: Evidence from the Pennsylvania Liquor Control Board's Entry Decisions. *American Economic Review*, 103(2):831–862, Apr. 2013. ISSN 0002-8282. doi: 10.1257/aer.103.2.831. URL https://www.aeaweb.org/articles?id=10.1257/aer.103.2.831.

B. Seo. Firm Scope and the Value of One-Stop Shopping in Washington State's Deregulated Liquor Market, Apr. 2019. URL https://papers.ssrn.com/abstract=2863094.

S. Shakory, B. T. Smith, R. Jansen, B. Reel, and E. Hobin. The impacts of policies controlling the spatial availability of take-away alcohol on consumption and harms: A systematic narrative review. *Addiction (Abingdon, England)*, 120(10):1951–1969, 2025. ISSN 1360-0443. doi: 10.1111/add.70070. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/add.70070. tex.copyright: © 2025 The Author(s). Addiction published by John Wiley & Sons Ltd on behalf of Society for the Study of Addiction.

L. Sun, E. Ben-Michael, and A. Feller. Using Multiple Outcomes to Improve the Synthetic Control Method. *The Review of Economics and Statistics*, 2025.

L. P. Tabb, L. Ballester, and A. Grubesic. The spatio-temporal relationship between alcohol outlets and violence before and after privatization: a natural experiment, seattle, WA 2010–2013. *Spatial and Spatio-temporal Epidemiology*, 19:115–124, 2016. doi: 10.1016/j.sste.2016.08.003. URL https://doi.org/10.1016/j.sste.2016.08.003.

W. Tian, S. Lee, and V. Panchenko. Synthetic Controls with Multiple Outcomes. *The Econometrics Journal*, page utag005, Mar. 2026. ISSN 1368-4221. doi: 10.1093/ectj/utag005.

B. Trolldal. An investigation of the effect of privatization of retail sales of alcohol on consumption and traffic accidents in Alberta, Canada. *Addiction (Abingdon, England)*, 100(5):662–671, 2005. ISSN 1360-0443. doi: 10.1111/j.1360-0443.2005.01049.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1360-0443.2005.01049.x.

Washington Secretary of State. November 8, 2011 general election voters' pamphlet: Initiative measure 1183, 2011. URL https://www.sos.wa.gov/sites/default/files/2022-05/Voters%27%2520Pamphlet%25202011.pdf.

Washington State Liquor and Cannabis Board. Gov. Inslee: Keeping marijuana away from teens top focus for state leaders, 2014. URL https://lcb.wa.gov/pressreleases/gov-inslee-keeping-cannabis-away-teens-top-focus-state-leaders.

E. G. Weyl and M. Fabinger. Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*, 121(3):528–583, 2013. doi: 10.1086/670401. URL https://doi.org/10.1086/670401.

E. Williams, W. C. Kerr, and S. B. L. Barnett. Price Changes in Washington Following the 2012 Liquor Privatization: An Update Through 2016 With Comparisons to

California, Idaho, and Oregon. *Alcoholism: Clinical and Experimental Research*, 44 (2):501–510, Feb. 2020. ISSN 0145-6008, 1530-0277. doi: 10.1111/acer.14255. URL `https://onlinelibrary.wiley.com/doi/10.1111/acer.14255`.

N. Zhang. Alcohol Taxes and Birth Outcomes. *International Journal of Environmental Research and Public Health*, 7(5):1901–1912, May 2010. ISSN 1661-7827. doi: 10.3390/ijerph7051901. URL `https://pmc.ncbi.nlm.nih.gov/articles/PMC2898024/`.

A Appendix A: Statewide Design and Supporting Evidence

This appendix documents the statewide sample, counterfactual construction, inference checks, and supporting purchasing and selection evidence.

A.1 Data Construction

Table A.1 reports the live-birth sample flow and outcome-specific denominators. Table A.2 audits gestational age and confirms that estimated fertilization uses no fallback. Table A.3 replaces the study-specific fetal-growth thresholds with the published 2017 U.S. reference.

Table A.1: NVSS Natality Sample Construction and Outcome Availability

Step or analytic denominator	Records	Share of final sample
Records in cleaned natality files	40,070,651	–
Valid maternal residence state	39,984,305	–
Valid recorded birth year and month	39,984,305	–
Valid gestation for fertilization timing	39,942,931	–
Estimated fertilization in analysis window	19,805,760	100.00%
Valid birth weight	19,790,195	99.92%
Valid combined gestation	19,805,760	100.00%
Valid five-minute APGAR score	19,727,200	99.60%
Singleton eligible for SGA/LGA	19,105,585	96.46%
Congenital-anomaly items observed	17,966,733	90.71%

Notes: The cleaned files contain the deliveries needed to construct estimated-fertilization cohorts from January 2010 through December 2014, including some deliveries occurring in 2015. Valid gestation means that NCHS's edited combined gestation field (`COMBGEST`) is between 17 and 47 completed weeks; unknown, missing, and out-of-range values are treated as missing. Estimated fertilization month equals the fifteenth of the reported birth month minus seven times completed gestational weeks plus 11 days, and the resulting date is assigned to its calendar month. The 11-day adjustment combines the conventional 14-day LMP-to-fertilization interval with a three-day midpoint correction for gestation recorded in completed weeks. No gestational-length fallback is assigned. Rows preceding the estimated-fertilization-window restriction describe the broader source pool and therefore are not reported as shares of the final sample. Subsequent shares use the 19,805,760-birth final sample. Outcome denominators vary with item availability; SGA/LGA additionally require a singleton birth, observed infant sex, valid birth weight, and valid gestation.

A.2 Counterfactual Selection and Fit

Washington is the only treated jurisdiction, so the central design choice is the untreated path. I selected the weighting target using only pre-policy holdouts. Table A.4 compares seven

Table A.2: Gestational-Age Measurement and Fertilization-Timing Audit

Birth year	Records	Combined observed	Obstetric substitution	NCHS imputed	Fallback assigned
2010	1,143,267	100.00%	5.96%	3.56%	0.00%
2011	3,948,744	100.00%	5.82%	4.27%	0.00%
2012	3,948,761	100.00%	5.38%	3.02%	0.00%
2013	3,928,501	100.00%	5.46%	2.69%	0.00%
2014	3,984,826	100.00%	5.15%	2.43%	0.00%
2015	2,851,661	100.00%	5.13%	2.45%	0.00%

Notes: The sample is restricted by estimated fertilization month and maternal residence, so the 2010 and 2015 birth-year rows cover only deliveries associated with cohorts inside the analysis window. Combined observed is the share with a valid edited combined gestational age. Obstetric substitution is the NCHS flag indicating use of the clinical or obstetric estimate in the combined measure where that flag is available. NCHS imputed records carry the gestational-age imputation flag. The primary timing rule assigns no fallback: records without valid edited gestation are excluded before entering the estimated-fertilization-month sample.

Table A.3: SGA and LGA Using a Published US Growth Reference

Outcome	Effect (placebo SE)	Rank p	HAC SE	% change
SGA, published US reference	0.572 (0.352)	0.137	0.092	6.35
LGA, published US reference	-0.377 (0.422)	0.373	0.136	-2.52

Notes: Estimates reuse the preferred three-outcome donor and time weights. Binary effects are percentage points. Parentheses contain the placebo-variance sampling standard error from full-refit non-Washington assignments. Rank p is the separate two-sided placebo-rank diagnostic. HAC SE is the fixed-weight monthly-gap Newey–West sensitivity with six lags.

weighting approaches across the final 4, 6, and 9 pre-policy months. The preferred birthweight, preterm, and gestational-duration index minimizes mean scaled holdout RMSPE and has the smallest worst-case value. Actual post-I-1183 outcomes did not enter the selection.

Table A.4: Blinded Pre-Policy Holdout Comparison

Weighting target	Mean	Median	Worst	vs. all-state
Scalar BW/preterm/gestation index	**0.343**	**0.340**	**0.617**	**9.0%**
Common-vector adverse-tail	0.349	0.341	0.657	7.3%
Scalar continuous index	0.350	0.313	0.866	7.0%
Scalar adverse-tail index	0.359	0.342	0.708	4.6%
Common-vector BW/preterm/gestation	0.362	0.374	0.643	3.9%
Unweighted all-state DiD	0.376	0.399	0.730	0.0%
Outcome-specific RC-SDiD weights	0.384	0.390	1.082	-2.2%

Notes: Lower RMSPE is better. The final 4, 6, and 9 pre-policy months are separately treated as pseudo-post periods. The comparison covers 14 outcomes across all three splits (42 outcome-split cells). Each holdout RMSPE is divided by donor-state first-difference volatility so outcomes with larger units do not dominate. Improvement is relative to the unweighted all-state comparison. The selected target (bold) minimizes mean scaled holdout RMSPE and also has the smallest worst-case value. Washington's actual post-I-1183 outcomes were not used in selection.

Figure A.1 plots raw Washington levels, the common-index donor series, and outcome-specific donor series. Persistent level gaps are visible. RC-SDiD includes unit fixed effects and identifies changes relative to the weighted pre-policy comparison; it does not require identical raw levels.

Table A.5 reports the ten largest donor weights. Positive weight is spread across all 50 donors; the effective donor sample size is 43.5 and the largest donor weight is 4.1 percent.

Table A.5: Largest Donor Weights in the Preferred RC-SDiD Design

State	Weight (%)	Cumulative (%)
ND	3.4	3.4
NE	3.2	6.6
FL	3.1	9.7
ID	3.0	12.7
OR	2.8	15.5
WI	2.7	18.1
NH	2.6	20.8
IA	2.6	23.4
WY	2.5	25.9
AZ	2.4	28.3

Notes: Nonnegative donor weights sum to one and are learned from Washington's pre-treatment path of the preferred three-outcome index. Washington is excluded from the donor pool. All 50 eligible donors receive positive weight; the donor effective sample size is reported in the text.

A.3 Inference, Robustness, and Timing

Table A.6 reports the secondary condition-at-birth family; the signs are mixed and all Algorithm 4 intervals include zero. Table A.7 keeps three uncertainty summaries distinct:

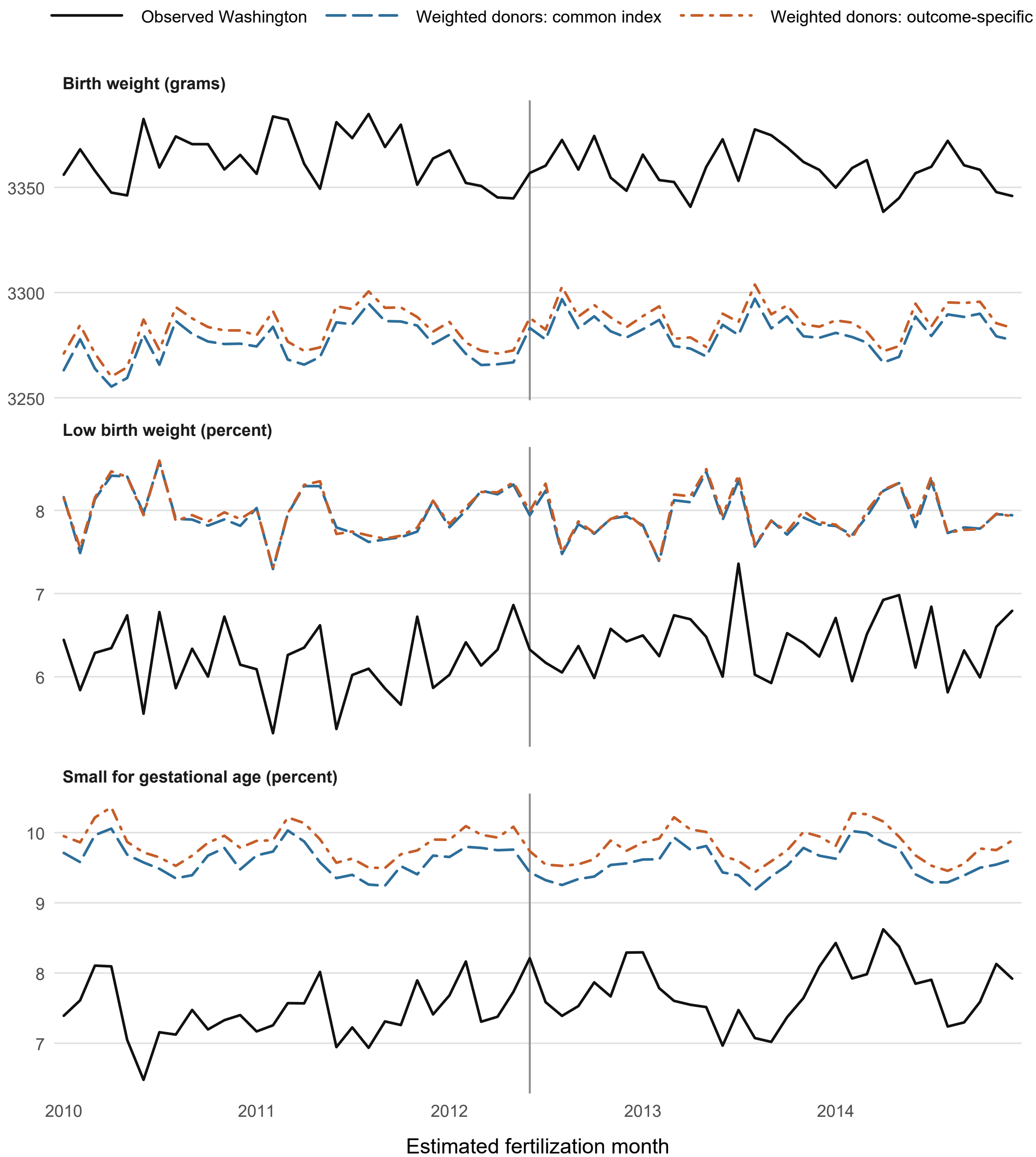

Figure A.1: Observed Washington and Weighted-Donor Outcome Levels

Notes: Panels show unsmoothed monthly levels for observed live births by estimated fertilization month from January 2010 through December 2014. Binary outcomes are percentages. The common-index donor series uses weights learned from the pre-policy birth-weight, preterm, and gestational-duration index; the outcome-specific series uses weights learned separately from each pre-policy outcome path. The vertical line marks I-1183 implementation in June 2012.

Algorithm 4 intervals use dispersion across 50 full refits, effect ranks compare Washington with the refitted placebo assignments, and HAC(6) uses monthly variation conditional on the selected counterfactual. The five-times-pre-MSPE screen is a fit diagnostic and does not change Washington's estimate or primary interval.

Table A.6: RC-SDiD Effects on Condition at Birth

Outcome	Estimate	95% CI	WA pre-policy mean	% change
APGAR5	0.030	[-0.171, 0.232]	8.821	0.34%
Low APGAR5	-0.167	[-1.512, 1.178]	1.886	-8.85%
Severe low APGAR5	-0.118	[-0.541, 0.305]	0.485	-24.33%
Any anomaly	0.128	[-1.211, 1.467]	0.452	28.40%

Notes: Estimates use donor and time weights learned from the selected birth-weight, preterm, and gestational-duration index. The 95% confidence interval is the Algorithm 4 full-refit placebo-variance sampling interval based on 50 non-Washington assignments. Binary outcomes and their pre-policy means are in percentage points; continuous outcomes remain in their stated units. No significance stars are shown. The interval relies on learning Washington's noise dispersion from control jurisdictions and should be read separately from placebo-rank and HAC diagnostics.

Table A.7: Placebo-Fit and HAC Sensitivity Diagnostics

Outcome	Effect-rank p	Post/pre RMSPE	RMSPE-rank p	5×-fit p	Screen N	HAC(6) SE
Birth weight and fetal growth						
Birth weight	0.235	1.43	0.157	0.178	44	1.713
Low birth weight	0.216	1.26	0.176	0.184	37	0.059
Very low birth weight	0.588	0.98	0.569	0.595	41	0.033
SGA	0.196	1.53	0.059	0.081	36	0.105
LGA	0.314	1.41	0.118	0.133	44	0.135
Term LBW	0.373	1.33	0.176	0.258	30	0.047
Gestational duration						
Preterm	0.804	1.08	0.608	0.562	31	0.115
Very preterm	0.882	1.22	0.235	0.306	35	0.044
Early term	0.882	1.12	0.510	0.471	33	0.152
Gestation	0.765	1.13	0.686	0.711	37	0.015
Condition at birth						
APGAR5	0.627	2.92	0.333	0.481	26	0.005
Low APGAR5	0.824	1.43	0.588	0.686	34	0.067
Severe low APGAR5	0.471	1.52	0.353	0.385	38	0.040
Any anomaly	0.510	0.73	0.569	0.684	37	0.053
Prespecified family index						
Prespecified family index	0.902	0.98	0.902	0.865	36	–

Notes: Effect-rank p is the separate two-sided Washington-excluded placebo-rank diagnostic. Post/pre RMSPE uses each jurisdiction's full-refit observed-minus-synthetic path after subtracting its SDID-time-weighted pre-period gap. RMSPE-rank p compares Washington with all 50 placebos. The fixed 5× sensitivity retains placebo jurisdictions whose pre-period MSPE is no more than five times Washington's; Screen N reports surviving placebo jurisdictions. HAC(6) SE is the fixed-counterfactual six-lag Newey–West gap-series sensitivity and is not the primary SDID sampling interval.

Table A.8 assesses whether related coefficients are jointly unusual using the maximum absolute standardized effect and the sum of squared standardized effects. Because Washington and placebo assignments use asymmetric donor pools, these are diagnostics rather than exact permutation tests or conventional simultaneous confidence bands.

Table A.8: Joint Full-Refit Placebo Diagnostics

Outcome set	Statistic	Components	Washington	Exceeding	Rank
Birth weight and fetal growth	M	6	1.248	26	0.529
Birth weight and fetal growth	Q	6	6.132	20	0.412
Exhaustive 500-gram birth-weight bins	M	11	1.658	22	0.451
Exhaustive 500-gram birth-weight bins	Q	11	6.608	24	0.490

Notes: The birth-weight and fetal-growth family contains the six prespecified outcomes; the distribution contains all eleven mutually exclusive 500-gram bins. For each assignment, M is the maximum absolute standardized effect and Q is the sum of squared standardized effects. Washington is standardized using all 50 non-Washington full refits; each placebo assignment uses the other 49 placebo assignments. The rank is $(1 + \#\{T_j \geq T_{WA}\})/51$. These are joint full-refit placebo diagnostics, not exact permutation tests or conventional simultaneous confidence bands, because Washington and placebo assignments use asymmetric donor pools.

Table A.9 varies only the Newey–West bandwidth for the fixed Washington-minus-weighted-donor monthly gap. Table A.10 instead varies the weighting target. These conditional and counterfactual-construction checks are sensitivities, not competing headline estimates.

Table A.9: Fixed-Counterfactual HAC Bandwidth Sensitivity

Outcome	Effect	HAC(3)	HAC(6)	HAC(12)
Birth weight and fetal growth				
Birth weight	-10.28	1.68 [0.000]	1.71 [0.000]	1.53 [0.000]
Low birth weight	0.274	0.088 [0.002]	0.059 [0.000]	0.047 [0.000]
Very low birth weight	0.047	0.041 [0.255]	0.033 [0.159]	0.027 [0.078]
SGA	0.357	0.115 [0.002]	0.105 [0.001]	0.087 [0.000]
LGA	-0.396	0.132 [0.003]	0.135 [0.003]	0.127 [0.002]
Term LBW	0.111	0.054 [0.041]	0.047 [0.018]	0.036 [0.002]
Gestational duration				
Preterm	-0.089	0.112 [0.428]	0.115 [0.443]	0.089 [0.321]
Very preterm	0.017	0.049 [0.724]	0.044 [0.694]	0.039 [0.651]
Early term	0.145	0.175 [0.410]	0.152 [0.341]	0.116 [0.215]
Gestation	0.014	0.016 [0.387]	0.015 [0.372]	0.012 [0.255]
Condition at birth				
APGAR5	0.030	0.005 [0.000]	0.005 [0.000]	0.004 [0.000]
Low APGAR5	-0.167	0.070 [0.017]	0.067 [0.013]	0.053 [0.002]
Severe low APGAR5	-0.118	0.037 [0.002]	0.040 [0.003]	0.035 [0.001]
Any anomaly	0.128	0.062 [0.040]	0.053 [0.016]	0.048 [0.007]
Prespecified family index				
Prespecified family index	0.019	0.050 [0.706]	0.050 [0.703]	0.038 [0.617]

Notes: The Effect column reproduces the preferred RC-SDiD estimate for Washington beginning in June 2012 using estimated-fertilization cohorts from January 2010 through December 2014. Each HAC cell reports the Newey–West standard error followed by its two-sided normal-approximation p-value in brackets for bandwidths of 3, 6, or 12 months. The calculation holds the selected Washington-minus-weighted-donor monthly gap and time weights fixed: 10 positive-weight pre-policy months and 31 post-policy months. Binary outcomes are percentage points; birth weight is grams; gestation is weeks; the family index is in donor-state pre-period standard deviations. These are time-series sensitivities conditional on the fitted donor path and weights. They do not replace the Algorithm 4 full-refit placebo-variance sampling intervals or the family-level full-refit diagnostics.

Table A.10: Health Estimates Under Alternative RC-SDiD Weight Targets

Outcome	Preferred index	Single outcome	Adverse-tail index	Continuous index	Common-vector tail
Birth weight	-10.28	-9.10	-7.65	-8.12	-6.90
Low birth weight	0.274	0.267	0.128	0.224	0.195
Very low birth weight	0.047	0.053	-0.026	-0.043	-0.064
SGA	0.357	0.379	0.425	0.481	0.336
LGA	-0.396	-0.424	-0.553	-0.497	-0.445
Term LBW	0.111	0.092	0.178	0.188	0.129
Preterm	-0.089	-0.100	-0.112	-0.030	0.034
Very preterm	0.017	0.004	-0.024	-0.013	-0.063
Early term	0.145	0.350	0.388	0.337	0.206
Gestation	0.014	-0.011	0.015	0.009	0.011
APGAR5	0.030	0.024	0.031	0.029	0.027
Low APGAR5	-0.167	-0.124	-0.219	-0.217	-0.144
Severe low APGAR5	-0.118	-0.109	-0.115	-0.139	-0.119
Any anomaly	0.128	0.012	0.128	0.163	0.176

Notes: The preferred index averages donor-pre-period standardized negative birth weight, preterm birth, and negative gestational length. The adverse-tail index uses SGA, preterm birth, and low five-minute Apgar; the continuous index uses negative birth weight, negative gestational length, and negative five-minute Apgar. Single-outcome columns refit the complete scalar design for each row. The common-vector column is a labeled multiple-outcome extension rather than the Morin scalar estimator. Binary outcomes are percentage-point effects.

Table A.11 ends with June 2014 estimated-fertilization cohorts, before recreational cannabis retail commenced. Table A.12 drops each of the five largest donors and refits the design.

Table A.11: Sensitivity to Excluding Conception Cohorts After Cannabis Retail Sales Began

Outcome	Through Dec. 2014	Through June 2014
Birth weight (grams)	-10.28 (1.71)	-9.80 (1.78)
Low birth weight (pp)	0.274 (0.059)	0.315 (0.066)
Very low birth weight (pp)	0.047 (0.033)	0.043 (0.034)
Small for gestational age (pp)	0.357 (0.105)	0.356 (0.138)
Large for gestational age (pp)	-0.396 (0.135)	-0.344 (0.142)
Term low birth weight (pp)	0.111 (0.047)	0.137 (0.053)
Preterm birth (pp)	-0.089 (0.115)	-0.137 (0.121)
Very preterm birth (pp)	0.017 (0.044)	0.009 (0.051)
Early term birth (pp)	0.145 (0.152)	0.190 (0.169)
Gestation (weeks)	0.014 (0.015)	0.014 (0.017)
APGAR5 score	0.030 (0.005)	0.034 (0.005)
Low APGAR5 (pp)	-0.167 (0.067)	-0.196 (0.069)
Severe low APGAR5 (pp)	-0.118 (0.040)	-0.137 (0.038)
Any congenital anomaly (pp)	0.128 (0.053)	0.123 (0.067)

Notes: Both columns use the preferred three-outcome RC-SDiD weights and treatment beginning in June 2012. The second column ends with June 2014 conception cohorts, before recreational cannabis retail sales began in Washington in July 2014. Parentheses contain fixed-weight six-lag Newey–West HAC sensitivity standard errors for the monthly Washington-minus-synthetic gap; no placebo refits are reported for this window sensitivity.

Table A.12: Leave-One-Top-Donor Sensitivity

Outcome	Estimate range	Refitted designs
Birth weight	[-10.67, -10.22]	5
Low birth weight	[0.264, 0.284]	5
Very low birth weight	[0.045, 0.054]	5
SGA	[0.352, 0.371]	5
LGA	[-0.418, -0.381]	5
Term LBW	[0.109, 0.117]	5
Preterm	[-0.097, -0.059]	5
Very preterm	[0.018, 0.020]	5
Early term	[0.109, 0.159]	5
Gestation	[0.013, 0.015]	5
APGAR5	[0.031, 0.032]	5
Low APGAR5	[-0.195, -0.165]	5
Severe low APGAR5	[-0.121, -0.110]	5
Any anomaly	[0.129, 0.137]	5

Notes: Each design removes one of the five largest positive-weight donors and then relearns both donor and time weights. Binary-outcome ranges are in percentage points.

Table A.13 reports the prespecified fixed-calendar exposure check. Although the nine monthly indicators are algebraically identified, adjacent indicators have variance-inflation factors as high as 37.5, so the table reports the more stable three-window specification and a separate one-month exposure-map sensitivity. Table A.14 compares the preferred full period with the June 2013 temporary-beer-tax endpoint and the stricter fixed-calendar endpoint. These are timing and policy-regime sensitivities, not biological mechanism estimates or alternative headline results.

A.4 Purchasing and Selection Evidence

The two Homescan tables report beverage-category and household-group estimates. Category quantities reconcile exactly to total ethanol-equivalent purchases, and the category table reports both HAC and full-refit placebo comparisons. Household demographics describe household heads rather than purchasers or consumers; the under-35 group has weak pre-policy fit and is interpreted cautiously.

The adult-drinking evidence is descriptive. NSDUH uses published two-year Washington estimates for adults age 18 and older, while BRFSS uses 2011–2014 respondent-level public-use data with survey weights, strata, and primary sampling units. Their mixed patterns are not pooled because sampling frames, timing, and outcome definitions differ; neither source identifies drinking during pregnancy.

The fetal analysis is limited to records meeting NCHS's tabulation-inclusion rule for stated or presumed gestation of at least 20 weeks. Supporting estimated-fertilization-cohort diagnostics contain 119,209 reportable fetal deaths with valid gestation, but valid conception timing covers only 48.2449 percent of audited fetal records. The combined denominator

Table A.13: Health Outcomes by Fixed Gestational-Calendar Policy Exposure

Outcome	Months 1–3	Months 4–6	Months 7–9
Birth weight and growth			
Birth weight	-6.0 [-27.0, 15.1]	1.6 [-25.8, 28.9]	-6.1 [-27.7, 15.4]
Low birth weight	0.38 [-0.45, 1.21]	-0.16 [-1.30, 0.98]	-0.02 [-0.94, 0.91]
Very low birth weight	-0.15 [-0.56, 0.26]	0.02 [-0.58, 0.62]	0.20 [-0.23, 0.63]
Small for gestational age	0.47 [-0.57, 1.50]	-0.42 [-1.69, 0.85]	0.25 [-0.94, 1.43]
Large for gestational age	-0.63 [-2.00, 0.73]	0.39 [-1.75, 2.54]	-0.07 [-1.29, 1.15]
Term low birth weight	0.24 [-0.32, 0.80]	-0.25 [-0.98, 0.49]	0.12 [-0.44, 0.69]
Gestational duration			
Preterm	-0.18 [-1.49, 1.12]	0.16 [-1.61, 1.93]	-0.08 [-1.39, 1.23]
Very preterm	-0.12 [-0.58, 0.34]	0.03 [-0.75, 0.81]	0.14 [-0.46, 0.73]
Early term	-0.35 [-2.43, 1.74]	0.39 [-1.74, 2.52]	0.19 [-1.73, 2.10]
Gestation	0.04 [-0.09, 0.17]	-0.03 [-0.17, 0.11]	-0.00 [-0.10, 0.09]
Condition at birth			
Apgar score	0.03 [-0.16, 0.23]	0.01 [-0.09, 0.10]	-0.01 [-0.09, 0.06]
Low Apgar	-0.20 [-1.56, 1.15]	-0.03 [-0.93, 0.86]	0.13 [-0.63, 0.88]
Severe low Apgar	-0.22 [-0.69, 0.25]	0.05 [-0.35, 0.46]	0.12 [-0.26, 0.49]
Any congenital anomaly	0.06 [-0.88, 1.00]	0.05 [-1.77, 1.87]	0.08 [-1.84, 2.00]
Index			
Prespecified family index	-0.06 [-0.54, 0.41]	0.06 [-0.54, 0.67]	0.02 [-0.39, 0.44]

Notes: Each coefficient is the change associated with moving a fixed three-month calendar window from no overlap to full overlap with I-1183, conditional on the other two windows. Cohorts are assigned using the paper's estimated-fertilization rule; windows do not use realized gestational length after assignment. Estimates use the preferred aggregate panel and donor universe and the prespecified birth-weight, preterm, and gestation index, with weights retrained only on cohorts before the earliest overlap (October 2011). Brackets contain 95% one-treated-state placebo-variance sampling intervals from 50 complete state refits. Binary outcomes are percentage-point effects. No coefficient-level interval excludes zero. The index's three-coefficient empirical full-refit placebo rank is 1.000. This is a joint diagnostic, not the sampling interval or an exact randomization p-value. Nine monthly indicators are not estimated jointly because their maximum variance-inflation factor is 37.4. Effect ranks, fit ranks, and six-lag conditional HAC sensitivities are kept distinct. These are timing sensitivities, not trimester-specific biological effects.

Table A.14: Full-Period and Temporary Beer-Tax-Regime Robustness

Panel A: Homescan monthly ethanol-equivalent purchases per projected household

Outcome	Full period	Through June 2013	Full placebo rank	Truncated placebo rank
Spirits ethanol	0.047 (0.008)	0.055 (0.005)	0.061	0.082
Beer ethanol	0.009 (0.002)	0.007 (0.001)	0.653	0.592
Wine ethanol	-0.020 (0.010)	-0.008 (0.006)	0.163	0.286
Total ethanol	0.035 (0.018)	0.054 (0.011)	0.286	0.143

Panel B: Health outcomes by estimated-fertilization cohort

Outcome	Through Dec. 2014	Through June 2013	Fully pre-expiration through Oct. 2012
Birth weight and growth			
Birth weight	-10.3 [-25.8, 5.2]	-10.5 [-23.9, 3.0]	-9.1 [-28.7, 10.5]
Low birth weight	0.27 [-0.13, 0.68]	0.33 [-0.15, 0.81]	0.14 [-0.62, 0.91]
Very low birth weight	0.05 [-0.23, 0.32]	0.03 [-0.27, 0.33]	0.02 [-0.28, 0.32]
Small for gestational age	0.36 [-0.27, 0.98]	0.32 [-0.27, 0.91]	0.53 [-0.20, 1.27]
Large for gestational age	-0.40 [-1.06, 0.26]	-0.20 [-0.80, 0.41]	-0.52 [-1.36, 0.32]
Term low birth weight	0.11 [-0.21, 0.44]	0.11 [-0.32, 0.54]	0.20 [-0.45, 0.85]
Gestational duration			
Preterm	-0.09 [-0.91, 0.74]	-0.07 [-0.86, 0.72]	-0.49 [-1.67, 0.69]
Very preterm	0.02 [-0.32, 0.35]	-0.02 [-0.39, 0.35]	-0.18 [-0.65, 0.28]
Early term	0.14 [-1.69, 1.97]	0.34 [-1.03, 1.72]	0.33 [-1.33, 2.00]
Gestation	0.01 [-0.08, 0.10]	0.00 [-0.08, 0.08]	0.04 [-0.06, 0.15]
Condition at birth			
Apgar score	0.03 [-0.17, 0.23]	0.03 [-0.14, 0.20]	0.03 [-0.12, 0.19]
Low Apgar	-0.17 [-1.51, 1.18]	-0.23 [-1.40, 0.95]	-0.27 [-1.40, 0.85]
Severe low Apgar	-0.12 [-0.54, 0.31]	-0.15 [-0.52, 0.22]	-0.15 [-0.50, 0.20]
Any congenital anomaly	0.13 [-1.21, 1.47]	0.15 [-1.52, 1.82]	0.11 [-1.57, 1.79]
Index			
Prespecified family index	0.02 [-0.29, 0.33]	0.04 [-0.25, 0.33]	-0.11 [-0.51, 0.30]

Notes: Washington's temporary \$15.50-per-barrel beer-tax increment applied through June 30, 2013. Panel A reports RC-SDiD estimates with six-lag conditional Newey–West standard errors in parentheses; placebo effect ranks are distinct diagnostics. Panel B reports RC-SDiD estimates with 95% one-treated-state placebo-variance sampling intervals in brackets; binary outcomes are percentage-point effects. Every truncated design preserves the baseline donor universe, outcome definitions, prespecified weight target, and full-refit inference hierarchy. A June 2013 estimated-fertilization cutoff does not ensure that all gestational months preceded tax expiration. The strict column ends with October 2012 cohorts, whose fixed ninth gestational month is June 2013, leaving only five fully exposed cohort months and correspondingly wide intervals.

contains observed live births and reportable fetal deaths; it is not all conceptions and excludes abortions and early losses.

Table A.15: Homescan Changes in Alcohol Purchases by Product Category

Category	Beverage liters	Ethanol liters	% of pre-mean	HAC p	Placebo p
Spirits	0.117***	0.047***	35.0	< 0.001	0.061
	(0.021)	(0.008)			
Beer	0.172***	0.009***	8.3	< 0.001	0.653
	(0.047)	(0.002)			
Wine	-0.170**	-0.020**	-14.4	0.044	0.163
	(0.084)	(0.010)			
Total alcohol	–	0.035*	9.2	0.053	0.286
		(0.018)			

Notes: Entries are RC-SDiD changes in monthly projected purchases per household. Newey–West HAC standard errors with six monthly lags are in parentheses. Ethanol-equivalent liters assume 40% ABV for spirits, 5% for beer, and 12% for wine. The total-alcohol estimate is the exact sum of the three category estimates. Washington is treated beginning June 2012; state and time weights use only data through May 2012. Placebo p-values compare Washington with donor-state placebo estimates. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.16: Homescan Ethanol-Equivalent Purchasing Effects by Household Characteristics

Household group	Spirits	Beer	Wine	Total alcohol	Total %
Panel A: Female-head age					
Under 35	-0.006	0.005	-0.072***	-0.073***	-20.3
	(0.019)	(0.018)	(0.027)	(0.027)	
35–44	0.052**	0.020***	-0.032**	0.040*	16.3
	(0.021)	(0.008)	(0.013)	(0.023)	
45+	0.078***	0.026***	0.011	0.114***	35.6
	(0.008)	(0.007)	(0.010)	(0.020)	
Panel B: Household-head composition					
Female and male heads	0.055***	0.015***	0.000	0.070***	19.4
	(0.011)	(0.006)	(0.009)	(0.016)	
Female head only	0.059***	0.030***	-0.035*	0.053*	23.5
	(0.012)	(0.009)	(0.019)	(0.032)	
Male head only	0.009	-0.035***	-0.052***	-0.078	-12.7
	(0.030)	(0.010)	(0.020)	(0.054)	

Notes: Entries are RC-SDiD changes in monthly ethanol-equivalent liters purchased per projected household; Newey–West HAC standard errors with six monthly lags are in parentheses. Every row uses the common all-household Homescan state and time weights estimated through May 2012, and each total is the exact sum of spirits, beer, and wine. Age describes the female household head, not the purchaser. The standalone under-25 Washington path has 48 months and is pooled with ages 25–34. The composition panel describes household heads, not purchaser sex. Normalized pre-fit exceeds 0.5 for every Under-35 path, spirits and wine among ages 35–44, and wine among female-head-only and male-head-only households; those cells are exploratory. The smallest Washington group-month contains 51 households. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B Appendix B: Within-Washington Retail Access and Birth Outcomes

This appendix supports Subsection 4.5. The within-Washington design has a strong and stable first stage for short-distance population coverage. The corresponding birth-outcome estimates are less robust and do not survive the full trend, placebo, small-cluster, and multiplicity diagnostic battery. I therefore treat the design as evidence on the local-access channel rather than as independent confirmation of a causal prenatal-health effect.

B.1 Retail-Access Data, Measurement, and First Stage

The pre-policy spirits network is the complete WSLCB roster effective June 28, 2011: 166 state stores and 162 contract stores, excluding military and tribal outlets. All 328 civilian baseline locations are geocoded. The supplied post-privatization universe contains 1,348 unique off-premise spirits locations, of which 1,261 are geocoded. Table B.1 reports the network and geocoding audit.

Table B.1: Within-Washington Spirits Retailer Geocoding Audit

Retailer universe	Addresses	Matched	Exact	Non-exact	No match	Match rate
Full supplied post-policy universe	1348	1261	1067	194	87	93.5%
June 2011 state/contract roster	328	328	202	126	0	100.0%

Notes: The pre-policy civilian network includes all 328 state and contract outlets on the WSLCB roster effective June 28, 2011; military and tribal outlets are excluded from the general-population measure. The post-policy network is the complete supplied off-premise location universe rather than a dated active-license subset. Historical approval dates are incomplete, so these networks identify a fixed pre/post change and not monthly retailer entry. The post universe contains 1,348 addresses before matching, close to the 1,373 post-privatization locations reported by Seo (2019).

The preferred measure assigns fixed 2010 block-group population to population centroids and calculates the county share within one mile of at least one spirits retailer. Measurement sensitivities use populated Census-block internal points, the share newly brought within one mile, and the population-weighted reduction in nearest-retailer distance. The distance measure permits the nearest retailer to lie across a county boundary.

The preferred block-group and Census-block one-mile measures are very similar across counties. Nearest-retailer distance is less closely tied to one-mile coverage and represents a different intensive margin. These alternatives were defined as measurement sensitivities rather than selected according to birth outcomes.

The excluded instrument is standardized pre-policy five-mile SNAP grocery coverage interacted with the post-I-1183 period. Figure 4 in the main text shows that the access

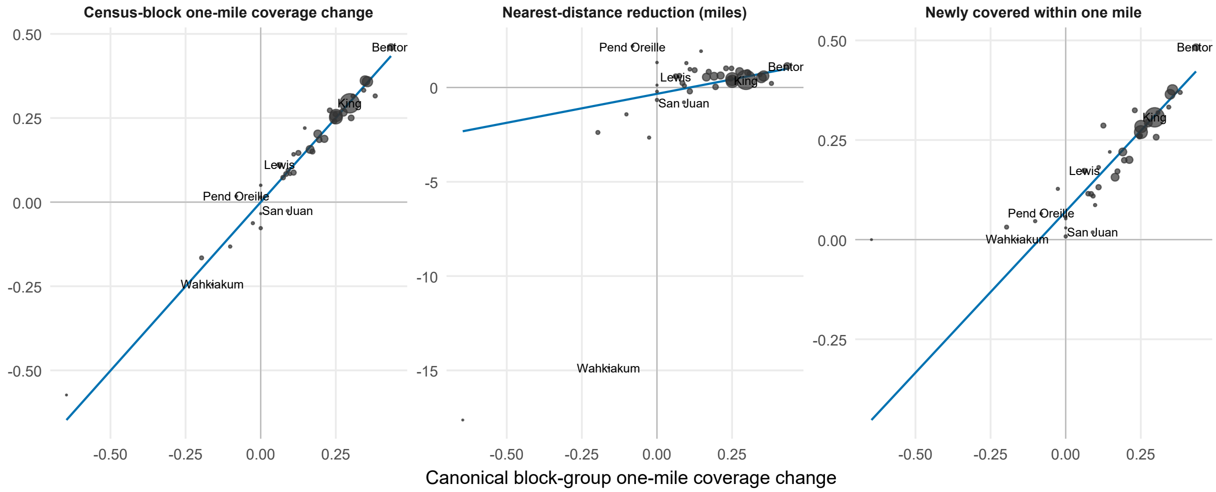

Figure B.1: County-Level Comparisons of Alternative Retail-Access Measurements
Notes: Panels compare the preferred block-group-centroid one-mile coverage change with Census-block one-mile coverage, the population-weighted reduction in mean nearest-retailer distance, and the fixed 2010 population share newly covered within one mile. Points are Washington counties, sized by 2010 population. Labeled counties were selected using prespecified discrepancy and influence thresholds before outcome estimation. Measures compare the fixed June 2011 and supplied post-privatization retailer networks; distance calculations permit cross-county nearest retailers.

response is concentrated at short distances. One-mile population coverage has a coefficient of 0.1245 (standard error 0.0237; squared clustered t-statistic = 27.63), whereas five-mile coverage was already nearly saturated and has essentially no first stage.

Table B.3 shows nearly the same one-mile first stage using Census blocks and a strong relationship for newly covered population. Mean nearest-distance reduction is weakly predicted under either instrument geography.

Leave-one-county-out one-mile estimates range from 0.1103 to 0.1330. Geocoding screens retain a positive relationship. Removing population weights leaves the coefficient positive but substantially weakens the first stage, so the evidence supports a short-distance population-coverage interpretation rather than a general expansion at all spatial scales.

B.2 Identification Diagnostics

A strong first stage establishes relevance but not the exclusion restriction required for a causal access interpretation. Pre-policy grocery infrastructure predicts population, urbanization, grocery density, several baseline health levels, and some differential pre-policy outcome paths. Cross-county shopping and other spatial spillovers also make county of residence an imperfect mapping from retail access to individual exposure.

The diagnostic battery includes linear and quarterly differential-pretrend tests, seven

Table B.2: First Stage: Pre-Policy Grocery Access and Post-I-1183 Spirits Access

Outcome	Coef.	SE	F-stat.	Obs.
Retailers per 100,000 residents	0.841	(1.041)	0.65	2340
Inverse nearest-retailer distance	0.044***	(0.010)	17.92	2340
Population share within 0.5 miles	0.094***	(0.022)	18.27	2340
Population share within 1 mile	0.125***	(0.024)	27.63	2340
Population share within 3 miles	-0.003	(0.019)	0.02	2340
Population share within 5 miles	-0.017	(0.017)	0.98	2340
Population share within 10 miles	-0.003	(0.013)	0.05	2340
Mean retailers within 0.5 miles	0.285***	(0.103)	7.58	2340
Mean retailers within 1 mile	1.112***	(0.337)	10.91	2340
Mean retailers within 3 miles	7.494***	(1.924)	15.17	2340
Mean retailers within 5 miles	18.037***	(4.993)	13.05	2340
Mean retailers within 10 miles	54.854***	(17.919)	9.37	2340
Smooth access (0.5-mile half-life)	0.981***	(0.282)	12.15	2340
Smooth access (1-mile half-life)	3.285***	(0.913)	12.94	2340
Smooth access (3-mile half-life)	19.155***	(5.980)	10.26	2340

Notes: Each row reports a county-month fixed-effects regression of the fixed June-2011-to-post-policy spirits-access change on the interaction of the post period with standardized pre-policy SNAP grocery coverage within five miles. Regressions include county and month fixed effects, use county population weights, and cluster standard errors by county. The reported statistic is the squared clustered t-statistic for the single excluded instrument. Cumulative coverage and nearby-retailer counts use 0.5, 1, 3, 5, and 10 mile thresholds. Smooth indices sum exp(-log(2) times distance divided by the stated half-life) across retailers, so a retailer's contribution halves at 0.5, 1, or 3 miles. Because access is observed as two snapshots, a retailer-access event-study would be mechanical and is not reported.

Table B.3: Within-Washington Access Measurement and First Stages

Treatment mapping	Instrument mapping	Coef.	Clustered SE	Wald/effective F	Obs.
Canonical BG one-mile	BG SNAP	0.125	0.024	27.63	2340
Census-block one-mile	BG SNAP	0.124	0.022	31.50	2340
Nearest-distance reduction	BG SNAP	0.151	0.194	0.61	2340
Newly covered within one mile	BG SNAP	0.107	0.015	50.03	2340
Block-based instrument sensitivity					
Canonical BG one-mile	Block SNAP	0.122	0.021	33.38	2340
Census-block one-mile	Block SNAP	0.122	0.020	37.13	2340
Nearest-distance reduction	Block SNAP	0.228	0.188	1.47	2340
Newly covered within one mile	Block SNAP	0.100	0.014	50.68	2340

Notes: Each row is a county-month first-stage regression with county and month fixed effects, county population weights, and county-clustered standard errors (39 counties, 60 months). The instrument is standardized pre-policy SNAP grocery coverage within five miles interacted with the post-I-1183 indicator. BG uses 2010 block-group population centroids; Block uses populated 2010 Census-block internal points. Positive nearest-distance reduction means the population-weighted mean nearest-retailer distance fell. With one excluded instrument, the reported cluster-robust effective-F analogue is numerically the squared clustered first-stage t statistic shown as the Wald/effective F; it is not interpreted through a universal threshold. The distance-reduction first stage is weak under either instrument mapping.

Table B.4: Within-Washington Pretrend and Placebo Diagnostics

Outcome	Status	Linear pretrend p-value	Minimum placebo p-value
Any congenital anomaly (pp)	estimated	0.173	0.143
Birth weight (grams)	estimated	0.810	0.132
Large for gestational age (pp)	estimated	0.377	0.114
Low birth weight (pp)	estimated	0.278	0.109
Small for gestational age (pp)	estimated	0.312	0.229
Term low birth weight (pp)	estimated	0.337	0.034
Very low birth weight (pp)	estimated	0.858	0.335
APGAR5 score	estimated	0.172	0.061
Low APGAR5 (pp)	estimated	0.047	0.006
Gestation (weeks)	estimated	0.981	0.550
Preterm birth (pp)	estimated	0.742	0.136
Very preterm birth (pp)	estimated	0.353	0.092

Notes: The linear pretrend diagnostic estimates whether pre-policy outcome trends vary with the pre-policy grocery-access instrument. Placebo p-values are the minimum across seven quarterly-spaced placebo dates from July 2010 through January 2012, using only pre-I-1183 months. Supporting diagnostics also report 9,999-draw wild-cluster p-values, Holm adjustments, and quarterly event-study plots; the conventional p-values in this compact table should be interpreted together with that broader evidence.

placebo dates, leave-one-county-out estimates, county-specific trends, family-wise multiplicity adjustment, and 9,999-draw wild-cluster inference. Several quarterly joint pretrend tests reject after adjustment, and no principal health outcome survives the combined battery. The reduced form is therefore the primary within-Washington comparison; the IV estimates are stronger-assumption rescalings.

B.3 Reduced Form, IV, and Measurement Robustness

The reduced form asks whether birth outcomes changed more after June 2012 in counties with greater pre-policy grocery exposure. Table B.5 reports the complete 12-outcome family under preferred estimated-fertilization timing and a delivery-minus-nine-month sensitivity. Under primary timing, one standard deviation greater pre-policy grocery exposure is associated with a 17.62-gram birth-weight decline and a 0.632-percentage-point SGA increase; restricted wild-cluster p-values are 0.158 and 0.076. The remaining outcomes are mixed, and county-specific-trend p-values are 0.105 for birth weight and 0.881 for SGA.

Table B.5: Within-Washington Reduced-Form Estimates Across Timing Constructions

Outcome	Estimated fertilization	Delivery minus nine months
Birth weight (grams)	-17.620 (9.074)	-19.002 (9.353)
Low birth weight (pp)	0.128 (0.273)	0.241 (0.309)
Very low birth weight (pp)	0.005 (0.046)	-0.037 (0.065)
Small for gestational age (pp)	0.632 (0.292)	0.506 (0.295)
Large for gestational age (pp)	-0.380 (0.447)	-0.457 (0.488)
Term low birth weight (pp)	0.161 (0.174)	0.313 (0.169)
Preterm birth (pp)	-0.254 (0.346)	-0.159 (0.290)
Very preterm birth (pp)	-0.118 (0.076)	-0.194 (0.077)
Gestation (weeks)	-0.027 (0.029)	-0.028 (0.027)
APGAR5 score	-0.060 (0.053)	-0.062 (0.054)
Low APGAR5 (pp)	0.453 (0.364)	0.587 (0.380)
Any congenital anomaly (pp)	0.041 (0.076)	0.093 (0.087)

Notes: Entries are county-month reduced-form coefficients on the post-I-1183 indicator interacted with standardized pre-policy grocery access; county-clustered standard errors are in parentheses. Models include county and calendar-month fixed effects and use outcome-specific birth-denominator weights. Binary outcomes are in percentage points. The estimates are county-level policy-exposure associations; conventional significance stars are omitted because the interpretation is governed by the complete pretrend, placebo, multiplicity, wild-cluster, county-trend, and influence battery.

Table B.6 rescales the reduced form by instrument-induced one-mile access. A 10-percentage-point increase corresponds to a decline of 11.02 grams in birth weight and an increase of 0.395 percentage points in SGA. Outcome-sample first-stage statistics are approximately 16.4–16.6, rather than 27.63 in the full access panel, because the health models use outcome-specific samples and birth-denominator weights.

Table B.6: Within-Washington One-Mile IV Estimates Across Timing Constructions

	Estimated fertilization		Delivery minus nine months	
Outcome	IV per 10 pp (SE)	F	IV per 10 pp (SE)	F
Birth weight (grams)	-11.020 (6.700)	16.6	-17.060 (10.063)	16.3
Low birth weight (pp)	0.080 (0.178)	16.6	0.216 (0.297)	16.3
Very low birth weight (pp)	0.003 (0.029)	16.6	-0.033 (0.056)	16.3
Small for gestational age (pp)	0.395 (0.223)	16.6	0.453 (0.304)	16.3
Large for gestational age (pp)	-0.237 (0.299)	16.6	-0.409 (0.471)	16.3
Term low birth weight (pp)	0.102 (0.116)	16.4	0.284 (0.174)	16.0
Preterm birth (pp)	-0.159 (0.214)	16.6	-0.143 (0.256)	16.3
Very preterm birth (pp)	-0.074 (0.042)	16.6	-0.175 (0.075)	16.3
Gestation (weeks)	-0.017 (0.021)	16.6	-0.025 (0.027)	16.3
APGAR5 score	-0.038 (0.032)	16.6	-0.056 (0.047)	16.3
Low APGAR5 (pp)	0.283 (0.218)	16.6	0.527 (0.327)	16.3
Any congenital anomaly (pp)	0.025 (0.049)	16.6	0.083 (0.080)	16.3

Notes: The endogenous variable is the county population share whose fixed 2010 block-group population centroid lies within one mile of a spirits retailer. The instrument is the post-I-1183 indicator interacted with standardized pre-policy grocery access. IV coefficients are scaled to a 10-percentage-point instrument-induced coverage increase; county-clustered standard errors are in parentheses. These assumption-dependent county-level rescalings do not identify individual proximity, maternal drinking, a prenatal dose response, or the effect of adding one store.

Table B.7: Within-Washington IV Estimates for Short-Distance Access Measures

Access mapping	Min. F	Birth weight	Gestation	LBW	SGA	Term LBW
Within 0.5 miles	15.7	-14.36 (6.50)	-0.02 (0.03)	0.104 (0.221)	0.516 (0.224)	0.133 (0.146)
Within 1 mile	16.4	-11.02 (6.70)	-0.02 (0.02)	0.080 (0.178)	0.395 (0.223)	0.102 (0.116)

Notes: Each IV estimate is scaled to a 10 percentage point increase in the listed county population-coverage measure. Binary outcomes are in percentage points; birth weight is in grams and gestation in weeks. The instrument is the post-I-1183 interaction with standardized pre-policy grocery access. The minimum first-stage statistic is across the five displayed outcome samples. Three-, five-, and ten-mile IV estimates are not reported because their first stages are too weak to support useful IV rescaling. These are assumption-dependent county-level rescalings, not confirmed health effects.

Table B.8 independently varies treatment and instrument geography. Holding the instrument fixed, changing treatment geography rescales the same reduced form; changing instrument geography changes both the reduced form and first stage. The four one-mile specifications produce similar health rescalings, but this is a measurement check rather than a model-selection exercise.

Table B.8: Four-Cell One-Mile Measurement Decomposition

Treatment	Instrument	FS coef. (SE)	FS F	Birth weight	Low birth weight	SGA	Gestation
BG	BG SNAP	0.160 (0.039)	16.63	-11.020 [-24.609, 2.569]	0.080 [-0.281, 0.441]	0.395 [-0.058, 0.848]	-0.017 [-0.059, 0.025]
Block	BG SNAP	0.165 (0.035)	21.72	-10.699 [-23.874, 2.476]	0.077 [-0.272, 0.427]	0.384 [-0.037, 0.804]	-0.016 [-0.056, 0.023]
BG	Block SNAP	0.151 (0.032)	21.87	-9.238 [-21.797, 3.321]	0.062 [-0.291, 0.415]	0.341 [-0.069, 0.752]	-0.006 [-0.039, 0.028]
Block	Block SNAP	0.154 (0.031)	25.40	-9.065 [-21.491, 3.361]	0.061 [-0.285, 0.406]	0.335 [-0.055, 0.725]	-0.005 [-0.038, 0.027]

Notes: The four cells independently vary the one-mile treatment geography and five-mile SNAP-instrument geography between block-group (BG) centroids and populated Census-block internal points. First stages use the birth-weight estimation sample and weights; F is the squared clustered first-stage t statistic. Health cells report 2SLS estimates with 95 percent clustered confidence intervals and are scaled to a 10-percentage-point access increase (birth weight in grams, binary outcomes in percentage points, gestation in weeks). Changing only treatment measurement rescales a fixed reduced form; changing the instrument changes the reduced form and first stage. The table is a prespecified measurement decomposition, not a model-selection exercise. Full 12-outcome results for all eight generated cells remain in the versioned CSV.

Table B.9 extends the treatment comparison to the complete prespecified outcome family. Census-block one-mile and newly-covered-population results are close to the preferred one-mile estimates. Nearest-distance reduction has a weak first stage, so conventional 2SLS scaling is unstable and Anderson–Rubin sets are frequently unbounded or disjoint.

Table B.9: Complete Within-Washington Measurement-Robustness Family

Outcome	Treatment mapping	Reduced form	Wild p	FS F	2SLS [95% CI]	AR 95% set
Birth weight (grams)	Canonical BG one-mile	-17.620 (9.074)	0.158	16.63	-11.020 [-24.609, 2.569]	[-32.403, 0.453]
	Census-block one-mile	-17.620 (9.074)	0.158	21.72	-10.699 [-23.874, 2.476]	[-29.942, 0.429]
	Newly covered	-17.620 (9.074)	0.158	26.70	-12.870 [-28.186, 2.447]	[-33.566, 0.531]
	Distance reduction	-17.620 (9.074)	0.158	0.61	-75.431 [-323.663, 172.801]	(-Inf, 1.415] ∪ [63.267, Inf)
Low birth weight (pp)	Canonical BG one-mile	0.128 (0.273)	0.658	16.63	0.080 [-0.281, 0.441]	[-0.250, 0.605]
	Census-block one-mile	0.128 (0.273)	0.658	21.72	0.077 [-0.272, 0.427]	[-0.240, 0.552]
	Newly covered	0.128 (0.273)	0.658	26.70	0.093 [-0.323, 0.509]	[-0.298, 0.617]
	Distance reduction	0.128 (0.273)	0.658	0.61	0.546 [-2.667, 3.759]	$(-\infty, \infty)$
Very low birth weight (pp)	Canonical BG one-mile	0.005 (0.046)	0.957	16.63	0.003 [-0.056, 0.062]	[-0.050, 0.091]
	Census-block one-mile	0.005 (0.046)	0.957	21.72	0.003 [-0.055, 0.060]	[-0.047, 0.084]
	Newly covered	0.005 (0.046)	0.957	26.70	0.003 [-0.066, 0.072]	[-0.056, 0.098]
	Distance reduction	0.005 (0.046)	0.957	0.61	0.020 [-0.411, 0.450]	$(-\infty, \infty)$
Small for gestational age (pp)	Canonical BG one-mile	0.632 (0.292)	0.076	16.65	0.395 [-0.058, 0.848]	[0.022, 1.125]
	Census-block one-mile	0.632 (0.292)	0.076	21.78	0.384 [-0.037, 0.804]	[0.022, 0.987]
	Newly covered	0.632 (0.292)	0.076	26.89	0.461 [-0.013, 0.936]	[0.028, 1.075]
	Distance reduction	0.632 (0.292)	0.076	0.61	2.695 [-5.761, 11.150]	(-Inf, -2.045] ∪ [0.074, Inf)
Large for gestational age (pp)	Canonical BG one-mile	-0.380 (0.447)	0.486	16.65	-0.237 [-0.843, 0.368]	[-1.120, 0.314]
	Census-block one-mile	-0.380 (0.447)	0.486	21.78	-0.231 [-0.814, 0.353]	[-1.018, 0.302]
	Newly covered	-0.380 (0.447)	0.486	26.89	-0.277 [-0.960, 0.405]	[-1.106, 0.386]
	Distance reduction	-0.380 (0.447)	0.486	0.61	-1.620 [-8.857, 5.618]	$(-\infty, \infty)$
Term low birth weight (pp)	Canonical BG one-mile	0.161 (0.174)	0.383	16.41	0.102 [-0.133, 0.336]	[-0.122, 0.430]
	Census-block one-mile	0.161 (0.174)	0.383	21.64	0.099 [-0.123, 0.320]	[-0.122, 0.374]
	Newly covered	0.161 (0.174)	0.383	27.09	0.119 [-0.141, 0.379]	[-0.152, 0.413]
	Distance reduction	0.161 (0.174)	0.383	0.60	0.697 [-1.882, 3.275]	$(-\infty, \infty)$

Continued on the next page.

Table B.9: Complete Within-Washington Measurement-Robustness Family (continued)

Outcome	Treatment mapping	Reduced form	Wild p	FS F	2SLS [95% CI]	AR 95% set
Preterm birth (pp)	Canonical BG one-mile	-0.254 (0.346)	0.494	16.62	-0.159 [-0.592, 0.274]	[-0.662, 0.337]
	Census-block one-mile	-0.254 (0.346)	0.494	21.70	-0.154 [-0.569, 0.260]	[-0.593, 0.329]
	Newly covered	-0.254 (0.346)	0.494	26.67	-0.186 [-0.698, 0.326]	[-0.756, 0.358]
	Distance reduction	-0.254 (0.346)	0.494	0.61	-1.090 [-4.261, 2.080]	$(-\infty, \infty)$
Very preterm birth (pp)	Canonical BG one-mile	-0.118 (0.076)	0.166	16.62	-0.074 [-0.160, 0.012]	[-0.169, 0.030]
	Census-block one-mile	-0.118 (0.076)	0.166	21.70	-0.072 [-0.152, 0.009]	[-0.151, 0.029]
	Newly covered	-0.118 (0.076)	0.166	26.67	-0.086 [-0.188, 0.016]	[-0.190, 0.032]
	Distance reduction	-0.118 (0.076)	0.166	0.61	-0.506 [-1.631, 0.620]	$(-\infty, \infty)$
Gestation (weeks)	Canonical BG one-mile	-0.027 (0.029)	0.367	16.62	-0.017 [-0.059, 0.025]	[-0.086, 0.017]
	Census-block one-mile	-0.027 (0.029)	0.367	21.70	-0.016 [-0.056, 0.023]	[-0.076, 0.016]
	Newly covered	-0.027 (0.029)	0.367	26.67	-0.020 [-0.066, 0.026]	[-0.079, 0.022]
	Distance reduction	-0.027 (0.029)	0.367	0.61	-0.116 [-0.605, 0.373]	$(-\infty, \infty)$
APGAR5 score	Canonical BG one-mile	-0.060 (0.053)	0.308	16.62	-0.038 [-0.103, 0.028]	[-0.114, 0.037]
	Census-block one-mile	-0.060 (0.053)	0.308	21.70	-0.037 [-0.099, 0.026]	[-0.104, 0.036]
	Newly covered	-0.060 (0.053)	0.308	26.66	-0.044 [-0.119, 0.031]	[-0.120, 0.043]
	Distance reduction	-0.060 (0.053)	0.308	0.60	-0.259 [-1.008, 0.490]	$(-\infty, \infty)$
Low APGAR5 (pp)	Canonical BG one-mile	0.453 (0.364)	0.242	16.62	0.283 [-0.159, 0.725]	[-0.223, 0.796]
	Census-block one-mile	0.453 (0.364)	0.242	21.70	0.275 [-0.141, 0.691]	[-0.219, 0.707]
	Newly covered	0.453 (0.364)	0.242	26.66	0.331 [-0.169, 0.831]	[-0.259, 0.831]
	Distance reduction	0.453 (0.364)	0.242	0.60	1.942 [-3.487, 7.372]	$(-\infty, \infty)$
Any congenital anomaly (pp)	Canonical BG one-mile	0.041 (0.076)	0.664	16.64	0.025 [-0.073, 0.124]	[-0.076, 0.153]
	Census-block one-mile	0.041 (0.076)	0.664	21.73	0.025 [-0.071, 0.121]	[-0.071, 0.144]
	Newly covered	0.041 (0.076)	0.664	26.73	0.030 [-0.087, 0.146]	[-0.080, 0.177]
	Distance reduction	0.041 (0.076)	0.664	0.61	0.174 [-0.715, 1.063]	$(-\infty, \infty)$

Notes: This table reports the complete outcome family prespecified before estimating these health models. Reduced forms use the preferred block-group SNAP instrument and therefore repeat across the four treatment mappings within an outcome. Parentheses contain county-clustered standard errors. IV entries contain conventional 2SLS estimates and 95 percent clustered confidence intervals. Binary outcomes are percentage points. Share-based IV effects are per 10 percentage points; distance reduction is per mile. AR sets invert the county-cluster-robust Anderson–Rubin test with 38 cluster degrees of freedom and may be bounded, disjoint, or all real numbers. The weak distance-reduction first stage produces nonstandard or unbounded AR sets and does not support a precise causal scaling. No LIML is added because the model is just identified and LIML equals 2SLS.

B.4 Interpretation and Limitations

The evidence establishes a local-access response to I-1183: predetermined grocery infrastructure strongly predicts expansion in very short-distance population coverage, and the first stage is stable to finer population geography and the newly-covered treatment. It does not support an equally precise distance-based mapping.

The downstream health evidence is considerably weaker. Birth weight and SGA move adversely along the preferred policy-exposure gradient, but the broader family is mixed and the results are sensitive to trend, placebo, multiplicity, and small-cluster diagnostics. Alternative spatial measurements do not resolve these identification concerns. The IV estimates additionally require that pre-policy grocery infrastructure affect post-I-1183 birth outcomes only through instrument-induced spirits access.

Accordingly, this analysis documents a local retail-access channel without independently establishing a causal prenatal-health mechanism. Treatment is county population coverage, not individual residential proximity, and the design does not identify whether a particular mother gained a nearby retailer, where alcohol was purchased, whether alcohol consumption changed during pregnancy, or the biological effect of prenatal ethanol exposure.

C Sufficient-Statistics Derivation for Corrective Alcohol Taxation

This appendix develops the welfare condition used in Section 5. The purpose is to show formally how the reduced-form birth-health estimates enter a corrective-tax problem with differentiated products, heterogeneous external harms, imperfect competition, and incomplete tax pass-through. The derivation builds on the sufficient-statistics logic of Griffith et al. (2019) and the treatment of pass-through under imperfect competition in Weyl and Fabinger (2013).

The key distinction from the benchmark alcohol-tax problem is that I-1183 changed the market structure through which subsequent taxes are transmitted to consumers. The post-privatization equilibrium can therefore differ from the pre-privatization equilibrium in both its price-cost wedges and its pass-through of a given tax instrument.

C.1 Environment

There are N consumers and a set of differentiated alcohol products $j \in \mathcal{J}$. Consumer i purchases quantity q_{ij} of product j, and the vector of consumer prices is $p = (p_1, \ldots, p_J)'$.

Let

$$Q_j = \sum_{i=1}^{N} q_{ij} \tag{18}$$

denote aggregate demand for product j.

Product j contains z_j units of ethanol per unit of the product. Individual ethanol consumption is therefore

$$Z_i = \sum_{j \in \mathcal{J}} z_j q_{ij}. \tag{19}$$

Let $E_i(Z_i)$ denote external costs associated with consumer i's ethanol consumption. The derivative $E_i'(Z_i)$ is the marginal external harm generated by an additional unit of ethanol consumed by individual i. This object can vary across consumers. Such heterogeneity is central to the alcohol-tax problem because consumers with different marginal harms can respond differently to the same tax change (Griffith et al., 2019).

Let c_j denote the marginal resource cost of supplying product j. For tax accounting, let s_j denote the amount of the consumer price ultimately retained by private producers and retailers per unit sold, net of government taxes and fees. Define the statutory tax wedge by

$$\omega_j = p_j - s_j. \tag{20}$$

The total price-cost wedge can then be decomposed as

$$p_j - c_j = \underbrace{\omega_j}_{\text{statutory wedge}} + \underbrace{(s_j - c_j)}_{\text{equilibrium markup}}. \tag{21}$$

This decomposition is useful because the consumer price can exceed marginal resource cost for two conceptually different reasons: taxes and fees create a statutory wedge, while imperfect competition creates an equilibrium markup.

Specific and ad valorem taxes are both nested in this notation. For example, if product j faces a specific tax ℓ_j and an ad valorem tax v_j applied to the seller's pre-tax receipt, then

$$p_j = (1 + v_j)s_j + \ell_j, \tag{22}$$

or equivalently,

$$s_j = \frac{p_j - \ell_j}{1 + v_j}. \tag{23}$$

The precise statutory mapping can be modified to reflect the relevant tax base or fee structure without changing the welfare derivation below.

C.2 Social Welfare and the Tax First-Order Condition

Let $\mathbf{t} = (t_1, \dots, t_K)'$ denote the vector of available tax instruments. Consumer prices, quantities, tax revenue, and producer surplus are equilibrium functions of $\mathbf{t}$.

Aggregate social welfare is

$$W = CS + \Pi + R - \sum_{i=1}^{N} E_i(Z_i), \tag{24}$$

where CS is consumer surplus, Π is total private producer surplus, and R is government revenue. Government revenue is included because taxes and statutory fees are transfers within this partial equilibrium; the external costs are subtracted separately.

For a marginal change in tax instrument t_k, the envelope theorem gives

$$\frac{\partial CS}{\partial t_k} = -\sum_{j \in \mathcal{J}} Q_j \frac{\partial p_j}{\partial t_k}. \tag{25}$$

Total private producer surplus can be written locally as

$$\Pi = \sum_{j \in \mathcal{J}} (s_j - c_j) Q_j, \tag{26}$$

where c_j is interpreted as marginal resource cost at the equilibrium quantity. Its derivative is

$$\frac{\partial \Pi}{\partial t_k} = \sum_{j \in \mathcal{J}} Q_j \frac{\partial s_j}{\partial t_k} + \sum_{j \in \mathcal{J}} (s_j - c_j) \frac{\partial Q_j}{\partial t_k}. \tag{27}$$

Government revenue is

$$R = \sum_{j \in \mathcal{J}} \omega_j Q_j, \tag{28}$$

so

$$\frac{\partial R}{\partial t_k} = \sum_{j \in \mathcal{J}} Q_j \frac{\partial \omega_j}{\partial t_k} + \sum_{j \in \mathcal{J}} \omega_j \frac{\partial Q_j}{\partial t_k}. \tag{29}$$

Because $p_j = s_j + \omega_j$, the inframarginal transfer terms cancel:

$$-Q_j \frac{\partial p_j}{\partial t_k} + Q_j \frac{\partial s_j}{\partial t_k} + Q_j \frac{\partial \omega_j}{\partial t_k} = 0. \tag{30}$$

Combining Equations (25)–(30) therefore gives

$$\frac{\partial(CS+\Pi+R)}{\partial t_k} = \sum_{j\in\mathcal{J}} (p_j - c_j)\frac{\partial Q_j}{\partial t_k}. \tag{31}$$

The corresponding change in external cost is

$$\begin{aligned}\frac{\partial}{\partial t_k}\sum_{i=1}^{N} E_i(Z_i) &= \sum_{i=1}^{N} E_i'(Z_i)\frac{\partial Z_i}{\partial t_k} \\ &= \sum_{i=1}^{N}\sum_{j\in\mathcal{J}} z_j E_i'(Z_i)\frac{\partial q_{ij}}{\partial t_k}.\end{aligned} \tag{32}$$

Since $\partial Q_j/\partial t_k = \sum_i \partial q_{ij}/\partial t_k$, the welfare effect of a marginal change in tax instrument t_k is

$$\frac{\partial W}{\partial t_k} = \sum_{i=1}^{N}\sum_{j\in\mathcal{J}} \left[p_j - c_j - z_j E_i'(Z_i)\right]\frac{\partial q_{ij}}{\partial t_k}. \tag{33}$$

For an unconstrained interior optimum,

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{J}} \left[p_j - c_j - z_j E_i'(Z_i)\right]\frac{\partial q_{ij}}{\partial t_k} = 0 \qquad \forall k. \tag{34}$$

Equation (34) is the condition reported in Section 5.4. It has two economically distinct components. The first is the gap between the existing price-cost wedge and marginal external harm. The second is the behavioral response induced by the tax. Marginal external harm is therefore not sufficient for optimal tax design unless the tax-induced quantity response and the pre-existing price-cost wedge are also known.

C.3 Demand Substitution and Equilibrium Pass-Through

Consumers respond to consumer prices rather than directly to statutory tax instruments. Define the consumer-level demand derivative matrix

$$D_i(p) = \frac{\partial q_i(p)}{\partial p'}, \tag{35}$$

where the diagonal elements are own-price derivatives and the off-diagonal elements capture substitution or complementarity across products.

Define the equilibrium pass-through matrix

$$\rho(\mathbf{t}) = \frac{\partial p(\mathbf{t})}{\partial \mathbf{t}'}, \tag{36}$$

with element

$$\rho_{mk} = \frac{\partial p_m}{\partial t_k}. \tag{37}$$

The chain rule implies

$$\frac{\partial q_i}{\partial \mathbf{t}'} = D_i(p)\rho(\mathbf{t}), \tag{38}$$

or, componentwise,

$$\frac{\partial q_{ij}}{\partial t_k} = \sum_{m \in \mathcal{J}} \frac{\partial q_{ij}}{\partial p_m} \rho_{mk}. \tag{39}$$

Substituting Equation (39) into Equation (34) yields

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{J}} [p_j - c_j - z_j E_i'(Z_i)] \left(\sum_{m \in \mathcal{J}} \frac{\partial q_{ij}}{\partial p_m} \rho_{mk} \right) = 0 \qquad \forall k. \tag{40}$$

The sufficient statistics in this expression are therefore the marginal external-harm schedule, own- and cross-price demand derivatives, existing price-cost wedges, and instrument-specific pass-through.

The pass-through matrix is itself an equilibrium object under imperfect competition. To make this explicit, let $\mathcal{M}$ summarize the competitive environment, including the set of active firms and products, ownership structure, and mode or intensity of competition. Stack the firms' equilibrium pricing conditions as

$$F(p, \mathbf{t}; \mathcal{M}) = 0. \tag{41}$$

Provided the equilibrium is locally unique, the implicit function theorem gives

$$\rho(\mathbf{t}; \mathcal{M}) = -\left[\frac{\partial F}{\partial p'}\right]^{-1} \frac{\partial F}{\partial \mathbf{t}'}. \tag{42}$$

Equation (42) makes clear why pass-through cannot generally be treated as a mechanical feature of the tax. The pricing Jacobian depends on the shape and curvature of demand, marginal costs, and firms' strategic interaction. The mapping can also differ across tax instruments because specific and ad valorem taxes enter firms' pricing conditions differently (Weyl and Fabinger, 2013).

This point is particularly relevant for Washington. Let $\mathcal{M}^{pre}$ denote the state-controlled spirits market before I-1183 and $\mathcal{M}^{post}$ the private retail market after privatization. In general,

$$\rho(\mathbf{t}; \mathcal{M}^{pre}) \neq \rho(\mathbf{t}; \mathcal{M}^{post}). \tag{43}$$

The large entry response to I-1183 can therefore matter for subsequent tax policy even if consumer preferences are unchanged. The relevant pass-through for a post-I-1183 corrective-tax counterfactual is the pass-through generated by the post-privatization competitive environment.

Equation (43) should not be interpreted as assuming that a marginal tax change itself leaves entry irrelevant. The baseline sufficient-statistics condition is local to a given post-I-1183 equilibrium and holds the active set of firms and products fixed in a neighborhood of the tax change. If a sufficiently large tax change induces additional entry, exit, or product repositioning, the counterfactual must also incorporate fixed costs, changes in product variety, and the associated extensive-margin welfare effects. The role of I-1183 here is that it changes the equilibrium around which the subsequent tax counterfactual is evaluated.

C.4 The Competitive Pigouvian Benchmark

The familiar Pigouvian rule is nested as a special case. Under perfect competition, private sellers receive marginal resource cost,

$$s_j = c_j, \tag{44}$$

so that

$$p_j - c_j = \omega_j. \tag{45}$$

Equation (34) then becomes

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{J}} \left[\omega_j - z_j E_i'(Z_i)\right] \frac{\partial q_{ij}}{\partial t_k} = 0 \qquad \forall k. \tag{46}$$

Suppose further that marginal external harm per unit of ethanol is common across consumers on the relevant margin,

$$E_i'(Z_i) = \bar{e} \qquad \forall i, \tag{47}$$

and define the aggregate demand derivative matrix

$$D(p) = \sum_{i=1}^{N} D_i(p) = \frac{\partial Q}{\partial p'}. \tag{48}$$

In matrix notation, the tax first-order conditions are then

$$(\omega - z\bar{e})' \, D(p)\rho(\mathbf{t}) = 0. \tag{49}$$

If the available policy instruments span the relevant price changes so that $D(p)\rho(\mathbf{t})$ has full row rank, Equation (49) implies

$$\omega_j = z_j \bar{e} \qquad \forall j. \tag{50}$$

Thus, the familiar Pigouvian result does not require pass-through to equal one. It requires that market power create no additional price-cost wedge and that the available instruments be capable of implementing the relevant product-price wedges. Pass-through governs how statutory changes translate into equilibrium prices and quantities.

With imperfect competition,

$$p_j - c_j = \omega_j + \mu_j, \qquad \mu_j \equiv s_j - c_j, \tag{51}$$

and the markup μ_j enters the same welfare wedge as the statutory tax. With differentiated products and heterogeneous marginal harms, there is generally no component-by-component rule in which the statutory tax simply equals average external harm. The optimal instrument depends jointly on who responds, which products they substitute toward, existing markups, and equilibrium pass-through.

C.5 Mapping the Reduced-Form Estimates into Marginal External Cost

Section 5.1 constructs the birth-health component of the external-cost numerator from mutually exclusive birth-weight states. Let $h \in \mathcal{H}$ index those states and let $\widehat{\tau}_h$ denote the estimated I-1183 effect on the share of observed Washington live births in state h, expressed in proportion units. For N_{WA}^{post} observed post-policy births,

$$\Delta\widehat{N}_h = \widehat{\tau}_h N_{WA}^{post}. \tag{52}$$

Let C_h denote incremental resource cost, L_h^Q the morbidity QALY loss relative to a common reference state, and V_Q the value assigned to one QALY. The valued birth-health consequence is

$$\Delta\widehat{C}^{birth} = \sum_{h\in\mathcal{H}} \Delta\widehat{N}_h \left(C_h + V_Q L_h^Q\right). \tag{53}$$

Because the birth-weight categories are mutually exclusive, this construction avoids separately monetizing overlapping measures such as mean birth weight, low birth weight, and SGA. If the bins are estimated separately, their implied changes should also satisfy the accounting restriction associated with an exhaustive distribution before the valuation is applied.

Let θ denote the share of the valued consequence that is not internalized in the relevant private alcohol-use decision. Then

$$\Delta\widehat{EC}^{birth} = \theta\Delta\widehat{C}^{birth}. \tag{54}$$

The denominator is the policy-induced change in total ethanol-equivalent purchasing. If $\widehat{\delta}_A$ is the estimated change in ethanol-equivalent liters per projected household per month and H_t is the projected number of households represented in month t, then under a constant post-policy purchasing effect,

$$\Delta\widehat{A}^E = \widehat{\delta}_A \sum_{t\in\mathcal{T}^{post}} H_t. \tag{55}$$

If the purchasing effect varies over time, the corresponding construction is

$$\Delta\widehat{A}^E = \sum_{t\in\mathcal{T}^{post}} \widehat{\delta}_{A,t} H_t. \tag{56}$$

The birth-related marginal external-cost estimate used in the main text is therefore

$$\widehat{MEC}^{birth} = \frac{\Delta\widehat{EC}^{birth}}{\Delta\widehat{A}^E}. \tag{57}$$

This ratio has a useful connection to the marginal external-harm primitive in Equation (34). Suppose, for the moment, that the observed quantity denominator corresponds exactly to the policy-induced change in individual ethanol consumption, so that

$$\Delta A^E = \sum_i \Delta Z_i. \tag{58}$$

For a sufficiently local change,

$$\Delta EC^{birth} \approx \sum_i E_i^{\prime,birth}(Z_i)\Delta Z_i. \tag{59}$$

Dividing by the aggregate ethanol response gives

$$\frac{\Delta EC^{birth}}{\Delta A^E} \approx \frac{\sum_i E_i'^{,birth}(Z_i)\Delta Z_i}{\sum_i \Delta Z_i}. \tag{60}$$

Equation (60) shows why behavioral incidence matters. When individual responses have the same sign, the ratio is a response-weighted average of marginal birth-related external harms: consumers whose ethanol consumption responds more strongly receive greater weight. More generally, when some consumers increase and others decrease consumption, it is a signed response-weighted ratio.

The Homescan denominator does not satisfy the equality above by construction. It measures recorded household retail purchases rather than individual ethanol consumption, and it does not identify whether the purchasing response occurred among people or periods relevant to pregnancy. Equation (57) is therefore a policy-induced external-cost-per-purchased-ethanol statistic. Its interpretation as the local marginal external harm in Equation (34) requires additional assumptions linking recorded purchasing to the relevant consumption response and linking the discrete I-1183 reform to the margin affected by a counterfactual tax change. It should not be interpreted as a biological prenatal dose-response parameter.

C.6 Internalities

The preceding derivation treats corrective taxation as responding to external harms. Consumers may also fail to internalize private harms from alcohol use because of present bias, addiction, imperfect information, or related behavioral frictions (Allcott et al., 2019).

Let $B_i'(Z_i)$ denote a welfare-relevant marginal internality per unit of ethanol: the marginal private harm generated by additional consumption that is not incorporated into consumer i's decision. Under a behavioral-welfare specification in which this object is identified separately, the corrective-harm term becomes

$$E_i'(Z_i) + B_i'(Z_i), \tag{61}$$

and Equation (34) generalizes to

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{J}} \left[p_j - c_j - z_j \left(E_i'(Z_i) + B_i'(Z_i)\right)\right] \frac{\partial q_{ij}}{\partial t_k} = 0 \qquad \forall k. \tag{62}$$

The reduced-form birth-health analysis does not identify $B_i'(Z_i)$. The internality term is therefore conceptually distinct from the birth-related external-cost calculation and would require its own empirical or calibrated sufficient statistic.

C.7 Interpretation

The derivation separates three links that are easy to conflate. First, the reduced-form evidence can be used to construct a birth-related external-cost response and scale it by the reform-induced change in ethanol purchasing. Second, consumer demand determines which products and consumers respond when prices change. Third, the supply side determines how a statutory tax change reaches those prices.

I-1183 is particularly informative for the final distinction because it changed the competitive environment itself. The relevant post-privatization tax problem therefore cannot generally import a pass-through rate from the pre-I-1183 state-controlled market or assume that statutory taxes map one-for-one into consumer prices. The corrective-tax calculation instead requires the demand derivatives, price-cost wedges, and equilibrium pass-through generated by the post-I-1183 market structure, together with the full set of relevant external harms.