Glyphosate exposure and GM seed rollout unequally reduced perinatal health

Edward Rubin   Emmett Saulnier*

June 27, 2024

Abstract

The advent of herbicide-tolerant genetically modified (GM) crops spurred rapid and widespread use of the herbicide glyphosate (GLY) throughout US agriculture. In the two decades following GM-seed's introduction, the volume of GLY applied in the US increased by more than 750%. Despite its breadth and scale, science and policy remain unresolved regarding the effects of GLY on human health. We identify the causal effect of GLY exposure on perinatal health by combining (1) county-level variation in GLY use driven by (2) the timing of the GM technology and (3) differential geographic suitability for GM crops. Our results suggest the introduction of GM seeds and GLY significantly reduced average birthweight and gestational length. While we find effects throughout the birthweight distribution, low-weight births experienced the largest reductions: the effect for births in the lowest decile is 4.5 times larger than that of the highest decile. Together, these estimates suggest that GLY exposure caused previously undocumented and unequal health costs for rural US communities over the last 20 years.

Significance Statement

While the herbicide glyphosate (GLY) is the most commonly used herbicide globally, the effects of GLY exposure on human health and the environment remain unclear—particularly in more developed countries, where GLY exposure is often considered low. Using spatiotemporal variation in the adoption of GLY-resistant crops, we document significant adverse perinatal health effects due to increased GLY exposure in the rural United States. Further, historically disadvantaged groups disproportionately bear these health effects. These results conflict with current regulatory guidance, suggest regulations may be inefficiently set, and highlight the need to improve pesticide use and exposure monitoring.

*We thank Wes Austin, Mark Colas, Olivier Deschenes, Todd Doley, Eyal Frank, Benjamin Hansen, Frederik Noack, Glen Waddell, Eric Zou, and participants in the Exeter LEEP seminar, the WEAI annual conference, TWEEDS, the UC Santa Barbara Bren seminar, the University of Oregon Applied Micro Workshop, the AERE Summer Conference, and the Oregon State School of Public Policy. Errors are the authors'. Rubin: Assistant Professor of Economics, University of Oregon (edwardr@uoregon.edu). Saulnier: Doctoral candidate in the Department of Economics at the University of Oregon. (emmetts@uoregon.edu). Authors contributed equally. This project was supported in part by an appointment to the Research Participation Program at the Water Economics Center, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the US EPA; no official endorsement by the EPA should be inferred.

While the introduction of genetically modified (GM) crops profoundly transformed the agricultural landscape of the United States, our understanding of the implications for human health remains limited. At the heart of GM technology lies its resistance to the herbicide glyphosate (GLY), which enabled farmers to directly spray GLY onto GM crops, eliminating weeds while sparing the crops themselves. Recent research advertises the potential of this technology to improve farm productivity [1–3], but the pairing of GM seeds with GLY introduced a complex array of health externalities—with substantial uncertainty regarding the total effect. On the one hand, if GLY replaced more toxic herbicides common to non-GM cultivation, its introduction could yield positive health effects. However, the ensuing liberal application of GLY—enabled by the central innovation of herbicide resistance—has led to substantially higher volumes of chemicals sprayed, potentially worsening health [4–6].

In this study, we use spatiotemporal variation in GM crop adoption to quantify the sign and magnitude of the human health externality that resulted from the widespread adoption of GM crops and the rapid increase of GLY. We then utilize a novel application of machine learning to document heterogeneity in the effects of the GM weed management regime on perinatal health. Our results suggest that GLY has adverse effects on perinatal health, net of any benefits associated with reductions in the use of other herbicides. These adverse effects concentrate among the most at-risk births.

The US first approved GLY for agricultural use in 1974, and the chemical subsequently gained prominence as a broad-spectrum herbicide used extensively in agriculture. Its effectiveness in weed control and relatively low toxicity contributed to its widespread adoption, becoming a critical component of weed-management practices. Pairing GLY with GLY-resistant GM crops removed a natural limiter in GLY use—GLY kills non-genetically modified crops and targeted weeds. Relaxing this constraint resulted in dramatic increases in GLY application intensity. Before the release of GM seeds in 1996, US farmers applied 0.1 kg of GLY per hectare of cropland; since GM seeds, the intensity has risen to over 1.3 kg/hectare [7].

Meanwhile, in the EU, which never approved GM seeds, the GLY application rate remains near the US's pre-GM levels: approximately 0.2 kg/hectare [8]. While potentially affected by other factors, this US-EU gap in GLY intensity illustrates how GM technology enabled substantially higher levels GLY intensity than otherwise available.

Since its approval, US regulators have consistently affirmed that "there are no risks to human health from the current registered uses of glyphosate" [9]—despite a dearth of population-wide, causally founded studies. However, two recent studies document negative health impacts of GLY exposure in Brazil. Dias, Rocha, and Soares [10] and Skidmore, Sims, and Gibbs [11] find that GLY exposure—driven by the expansion of GM seeds and transported through rivers—increased infant mortality and pediatric cancer deaths in Brazil. These two studies offer the first large-scale, population-level, plausibly causal estimates of the health costs of GLY exposure. Our study complements these analyses by considering GLY impacts in a substantively different socioeconomic setting—the US's GDP per capita is approximately nine times greater than Brazil's [12]—with a potentially different exposure mechanism stemming from differences in intensity of GLY use, hydrology, geology, and meteorology. Brazil applies nearly twice the amount of GLY per hectare of cropland as the US [10]. Thus, while the US and Brazil overlap as leaders in GLY application, key differences between the contexts warrant new study in the US and abroad. Finally, given the Brazilian context, Dias, Rocha, and Soares [10] and Skidmore, Sims, and Gibbs [11] focus on effects driven by exposure through upstream GLY application; we find that local exposure drives negative health impacts in the US context.

Background and motivation: GLY exposure, health effects, and mechanisms

GLY and other herbicides pervade the environment, giving rise to multiple pathways of human exposure [13] including water [14], dust blown by the wind [15], aerial drift [16], direct contact [17], and food residue [18]. Following application, GLY exhibits a relatively short breakdown period, with a half-life ranging from 2 to 215 days [19]. Although a significant portion of the herbicide binds to the soil, reducing runoff, its high water solubility allows the unbound remnant to enter both surface and groundwater [20, 21]. A comprehensive study across US waterways from 2015–2017 revealed the presence of GLY or its degradate, aminomethylphosphonic acid (AMPA), in 90% of samples [14]. Additionally, wind-dispersed dust particles containing soil-bound herbicide residues can contribute to air pollution [15]. While food residue is suspected to be another source of population-wide exposure to GLY [22], we cannot capture its effect in this study.

Exposure The multiple exposure mechanisms—coupled with the breadth and volume of GLY application—have resulted in widespread detection of GLY in the urine and blood of US residents. The US Center for Disease Control (CDC) detected GLY in 81% of urine samples from a nationally representative cohort [23]. Multiple studies with pregnant women found GLY present in the urine of nearly every tested mother-to-be [24, 25]. This ubiquity of GLY exposure in the US population highlights the importance of understanding the impacts of GLY exposure at a national scale—particularly within populations likely exposed to higher levels. We focus on rural populations' exposures to local GLY sources—GLY exposure through dust, drift, direct contact, or water originating within the county of residence. We explore the potential effects of upstream spraying in Appendix Section C.10.

Health impacts A growing body of literature suggests that GLY has the potential to impact human health through a variety of biological mechanisms. Existing evidence typically comes from either laboratory studies on non-human animals or human-focused observational studies. While laboratory-based studies offer well-identified causal effects, they often suffer challenges of external validity. Previous observational studies [24–26] primarily focus on associations between self-reported exposure and health outcomes and typically avoid causal claims. Dias, Rocha, and Soares [10], Skidmore, Sims, and Gibbs [11], and Camacho and Mejia [27] are exceptions—employing quasi-experimental methods to make causal statements.

Lab studies have established links between GLY and congenital anomalies in rats, developmental issues in frogs and chickens, and endocrine disruption for male reproduction in mice [28–30]. Multiple additional studies link GLY exposure to endocrine disruption, which can affect developmental and reproductive health [31]. Research also documents GLY toxicity for placental cells—raising concerns for adverse effects in fetal development [32].

Consistent with lab-based concerns for the effect of GLY on development and reproductive health, several observational studies report associations between GLY exposure and miscarriage [26], gestational length [24], and birthweight [25]. Camacho and Mejia [27] also find that aerially applied GLY during the Colombian government's anti-coca campaign increased miscarriages and short-run medical consultations for dermatological and respiratory issues. Considering these established mechanisms and documented associations, we evaluate the evidence for a causal effect of GLY on perinatal health outcomes—particularly birthweight and gestational length.

Beyond GLY-specific studies, a growing body of literature documents adverse health effects associated with pesticide exposure, even at low doses [33–39]. One of the closest studies to the current paper—Larsen, Gaines, and Deschenes [33]—finds pesticide exposure increases adverse birth outcomes (weight, gestation, and abnormalities) for California mothers. The authors highlight that this effect is driven by the sub-population exposed to the highest levels of pesticides. These results and the growing body of literature motivate three points for our paper. First, birthweight and gestation are plausible outcomes to test for the health effects of GLY. Second, if GLY affects health, rural populations are likely the most impacted. Third, the effect of GLY may be heterogeneous—varying within the exposed population. These three observations lay the foundation of our analysis.

Empirical approach

Our estimation strategy isolates plausibly exogenous variation in county-level GM adoption and GLY exposure by combining (1) temporal variation in the commercial release of GM seeds in the US with (2) spatial variation in the suitability for growing the main crops for which GM seeds are available—corn, soy, and cotton. The first dimension of this strategy utilizes the arbitrary timing of the release of GM seeds. GM seeds became commercially available in the US in 1996, and farmers rapidly adopted GM seeds and intensified GLY applications in the following years. The second dimension of our approach uses the fact that these changes, on average, had larger impacts in places that were relatively more suitable for the crops—typically where the crops were already grown. This approach effectively leverages two comparisons to isolate GLY's effect on perinatal health outcomes: (1) *before* versus *after* GM-induced GLY expansion, which began in 1996, and (2) *areas more suitable* for GM crops versus *less-suitable areas*. The empirical approach is quite similar to the commonly applied difference-in-differences estimator; however, because our suitability measurement is non-binary, we implement estimation in a two-stage least squares (2SLS) framework [40]. After estimating the average effects from GM seed adoption and GLY, we use a novel machine-learning-based approach to document heterogeneous effects as a function of expected birthweight.

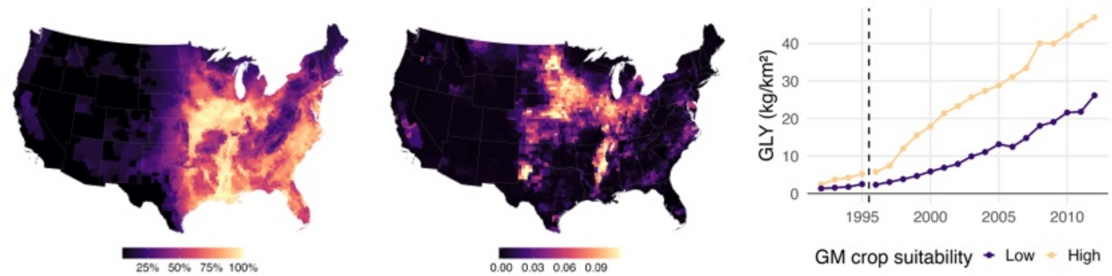
While this estimated effect of GM seeds and GLY on perinatal health highlights a critical consequence of herbicide application, it represents only part of the total potential externality associated with increased chemical usage. The previously discussed toxicology literature suggests several additional mechanisms and effects on human health that we do not measure here. Lab experiments [28–30] and recent observational work suggest additional ecological costs, such as biodiversity loss [41].

While we cannot address all of GLY's effects, using birthweight as our primary outcome offers several advantages. As discussed above, prior research establishes adverse effects of GLY on reproduction/development and finds GLY in the urine of nearly every expectant mother. Numerous studies then link birthweight to other later-life outcomes [42, 43]. Additionally, data on birthweight is widely available and reliable—the US has accurately and objectively measured the birthweight of nearly every child born for decades. Infants are also less prone to endogenous responses to health and environmental shocks. Similarly, infants have short histories over which they can accumulate exposure—simplifying exposure measurement. Birthweight thus provides both a key health outcome and a canary-in-a-coalmine-like indicator of GLY toxicity. Ample opportunities exist for future work to test additional costs and benefits of GLY and GM technology to more broadly assess welfare effects.

Data We measure perinatal health outcomes using the universe of individual-level birth data from the National Vital Statistics System (NVSS) during 1990–2013—using the restricted-access natality files that allow us to match each birth to the county of occurrence and the mother's county of residence [44]. These natality data include our primary perinatal health outcomes—birthweight and gestational length—and demographic and residence information from both parents and birth-location information. Our primary analyses focus on the 9 million births that occurred in rural US counties or involved mothers residing in rural counties—as defined by the US Department of Agriculture (USDA). We focus on this subset as it represents the births most likely to be impacted by the increase in GLY intensity and exposure induced by the release of GM seeds.

Using each infant's mother's county of residence, we match the birth to a measure of GLY exposure—the volume of annual, county-level estimates of GLY applications per square kilometer of total county area—from the USGS National Pesticide Synthesis Project spanning 1992–2017 [7]. Our measures of crop suitability—attainable yield estimates for corn, soy, and cotton—come from the Food and Agriculture Organization of the United Nations Global Agro-Ecological Zones modelling framework (FAO-GAEZ). These time-invariant predicted yields result from modeling crop responses to environmental conditions such as soil type and climate—holding management practices constant [45].

Figure 1 illustrates both the spatial and temporal variation that we use to identify the effect of GLY on infant health. Figure 1a maps the spatial variation in GM-crop suitability. This suitability strongly correlates with the increase in GLY intensity (kg/km^2) after the introduction of GM seed—shown in Figure 1b. Figure 1c highlights how the commercial



(a) GM crop suitability, GM attainable yield pctl.

(b) Increase in GLY, 1995–2012 (kg/km^2)

(c) GLY and suitability, 1992–2012 (kg/km^2)

Figure 1: GM-crop suitability predicts GLY increases after GM-seed introduction. (a) Percentile of attainable yield for GM crops equals the difference in attainable yield between high- and low-input scenarios from FAO and IIASA [45] for corn, soy, and cotton. We rescale each crop to be a national percentile, average over the three crops, and finally scaling again to be a national percentile. **(b)** Change in GLY censored at the 1st and 99th percentiles. **(c)** Total GLY applications for low and high GM crop suitability, defined as below or above the median attainable yield for GM.

release of GM seeds (dashed vertical line in 1996) drove marked increases in GLY use—particularly in counties with high (above median) attainable yield for GM crops.

Estimation Methodologically, we use two approaches. First, we show reduced-form results using an event study, where a county's "treatment" level is its percentile of attainable yield for GM crops (corn, soy, and cotton). Methods section A.1 describes this measure in depth. Thus, this suitability measure ranges from 0 to 1—*0* representing counties with the lowest attainable yield for GM crops; *1* denoting counties with the highest. Thus, one can interpret the event study's yearly coefficients as summarizing the difference between the average birthweight in high-GM-crop suitability counties relative to low-GM-crop suitability counties each year. Put together, the event study coefficients show how the birthweight difference between high- and low-suitability counties evolved through time.

We then recover GLY's causal effect on perinatal health outcomes using two-stage least squares (2SLS) [40]. In the first stage, we regress county-level GLY use on attainable yield interacted with year dummies (akin to the event study described above), along with fixed effects for county and year by month and controls for family demographics. We then regress birthweight on the predicted values of GLY from the first stage. This approach avoids bias from confounding effects by focusing on ‘good’ (exogenous) variation in the regressor of interest (GLY intensity). The first stage extracts good variation by projecting the regressor of interest on variables (instruments) unlikely to suffer from confounding—e.g., GM-crop suitability interacted with the arbitrary timing of GM technology’s rollout. The second stage uses these projections to estimate the causal effect of interest. The central assumption is that no unrelated phenomenon coincided with the rollout of GM technology *and* specifically affected birthweight in more GM-suitable counties. Under this assumption, known as the exclusion restriction, 2SLS provides consistent estimates of the average causal effect of GLY on perinatal health.

While 2SLS provides consistent estimates for our parameter of interest, our approach likely understates the actual magnitude of GLY’s effect for exposed individuals. This understatement is due to the “ecological fallacy,” which arises from the fact that we only assign an individual’s GLY exposure at the county level—masking any potential differences in exposure within counties [46].

Separating GLY’s direct and policy effects The interpretation of these causal effects requires some nuance. GLY-tolerant GM seeds allowed farmers to change their weed management practices—reducing their usage of non-GLY pesticides and mechanical tilling [6]. These other changes could potentially affect perinatal health, violating the exclusion restriction. Therefore, we present two effects in our results—a *policy effect* and a *GLY effect*.

The *policy effect* does not control for these other changes and thus captures the total (net) effect of the introduction of GM seeds. The policy effect is potentially the policy-relevant parameter since it represents the effects of GM seeds and GLY relative to alternative weed-management practices in place before the introduction of GM seeds.

Alternatively, by explicitly controlling for non-GLY pesticides and local economic conditions, the *GLY effect* provides an estimate of the direct causal effect of GLY on perinatal health. However, this direct GLY effect requires a stronger assumption related to the exclusion restriction discussed above: except for the pesticides and economic outcomes for which we are controlling, no other mechanisms affected perinatal that also controlled correlated with both suitability for GM crops *and* the timing of the GM-seed rollout.

Finally, GLY-based herbicides like Roundup typically contain other chemicals, e.g., surfactants that reduce surface surface tension. Our measured effects include impacts from other ingredients mixed with GLY in commercial herbicide formulations.

Heterogeneity In addition to estimating the *average* effect of GLY exposure (methods described above), we also estimate heterogeneous effects. Understanding heterogeneity in this setting is critical: health and policy implications can differ depending on *whom* GLY impacts and *how diffuse* the impacts are. Reductions in birthweight among infants already at risk for low birthweight may have more costly consequences than decreases in birthweight among higher-weight infants.

We estimate GLY's impact as a function of the infant's *expected* birthweight, which we estimate using a random forest trained to predict infants' expected birthweights. Specifically, this learning model predicts birthweight as a function of the infant's and parents' information—using data from all pre-1996 births, before the GLY-resistant seeds were widely available. (See Section A.3 for detailed methodology.) The resulting predictions enable us to test whether GLY-driven losses in birthweight came equally from all infants or whether they concentrated in low- or high-weight births. Using *predicted* birthweight allows us to avoid bias from splitting the sample on the outcome [47]. Subsequently, we estimate our reduced form and 2SLS results by predicted birthweight percentiles. Toward the goal of documenting shifts in the birthweight distribution, we also directly test whether GLY increased the probability of "low birthweight" (LBW, <2500g) or "very-low birthweight" (VLBW, <1500g) births—changing our outcome from birthweight to an indicator for low or very-low birthweight.

Results

Figure 2a demonstrates that the percentile of GM-crop suitability strongly predicts post-1996 increases in GLY application levels—effectively comparing GLY use in more- and less-suitable counties each year, relative to their difference in 1995. Before 1996, lower- and higher-GM-crop-suitable counties followed similar GLY trajectories: the event study remains relatively flat and close to zero. However, after the 1996 introduction of GM seeds, GLY intensity in counties with higher attainable yields for corn, soy, and cotton quickly outpaced GLY intensity in less-suitable counties. As GM-seed adoption accelerated in the late 1990s and early 2000s, this GLY intensity gap between high- and low-attainable yield counties widened. The event study confirms the strength of our instrument (percentile of GM-crop suitability) and illustrates the first stage of our 2SLS approach.

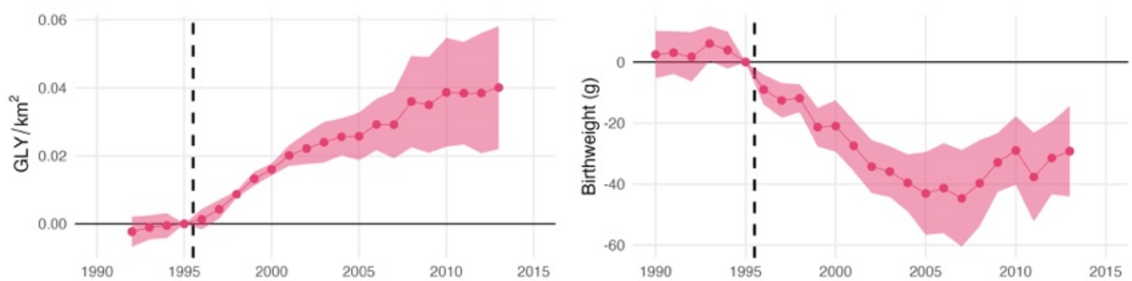
These first-stage findings align with a plausible mechanism: higher attainable yields fostered greater GM crop adoption, increasing GLY application. Extended Data Figure A1 shows applications of several prevalent herbicides—alachlor, cyanizine, fluazifop, and metolachlor—decreased following the release of GM seeds, suggesting farmers substituted away from these herbicides and toward GLY. Consequently, we control for these pesticides when estimating the *GLY effect* and omit them when estimating the *policy effect*.

Figure 2b depicts a similar event study for birthweight—plotting each year's average birthweight difference between births in high and low GM-crop suitability counties (relative to the 1995 difference) after controlling for maternal demographics. Prior to the introduction of GM crops, the birthweight gap between births in higher- and lower-attainable-yield counties remained stable. However, beginning in 1996—coinciding with the release of GLY-tolerant seeds and the intensification of GLY application—birthweights in higher-GM-suitability counties began declining relative to lower-yield counties. In 2005, a decade after the introduction of GLY-tolerant seeds, the average birthweight in the highest-yield county had fallen approximately 40 grams relative to the lowest-yield county.

The event study in Figure 2b also visualizes the reduced-form estimates of our 2SLS estimator: the effect of high attainable yield for GM crops on birthweight over time.

The flat trend prior to 1996 in the event study coefficients of Figure 2b also supports the parallel-trends assumption that underpins our empirical approach. In our context, this assumption requires that in the absence of the release of GLY-tolerant GM seeds, the difference in birthweight between higher- and lower-attainable-yield counties would have remained constant. Figure 2b suggests this assumption is plausible in our context: before the introduction of GM-tolerant seeds, counties followed similar trajectories independent of GM-crop suitability.

Figure A2 reproduces these same reduced-form event studies for several additional health outcomes. The event studies depict similar reductions in perinatal health: decreased gestation length; increased probabilities of LBW, VLBW, and preterm birth. We do not an effect on the probability of a C-section. Together, these results bear considerable evidence that after the release of GM seeds in 1996, perinatal health declined in high GM-attainable-yield counties relative to low-yield counties.



(a) *First stage,*
Effect of GM suitability on GLY intensity

(b) *Reduced form,*
Effect of GM suitability on birthweight

Figure 2: GM-seed introduction increased GLY intensity in GM-crop suitable areas; birthweight reductions were also higher in GM-suitable counties and match GM-seed timing. **(a)** Estimated event-study coefficients for the effect of local GM attainable yield percentile on GLY by year relative to 1995. Pesticide data only go back to 1992—there are no coefficients in 1990–1991. **(b)** Similar event study but with birthweight as outcome. Both regressions include county and year-by-month fixed effects and controls for parent demographics. Standard errors cluster by state and year. Demographic controls include mother's age, race, education, marital status, birth facility, resident status, previous births, fathers age and race, and sex of infant. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

The effect of GLY on perinatal health To estimate the main policy and GLY effects, we integrate the variation depicted in Figure 2 into a 2SLS framework. The instrument for GLY is the county's percentile of GM-crop suitability interacted with year indicators. This 2SLS approach effectively regresses individual health outcomes on the predicted level of GLY exposure, controlling for year, month, and parental demographics—where the prediction results from the first-stage estimates shown in Figure 2a (the expected GLY intensity for the mother's county of residence based upon the year and the county's suitability for GM crops). The GLY effect also controls for non-GLY pesticides and county-level unemployment. To aid interpretation, we report the estimated effect of GLY exposure at the 2012 mean level of GLY intensity (0.023 kg/km^2) rather than the actual coefficients from the regression (which represent a less intuitive effect: the effect of an additional kg/km^2 of GLY). Table A2 contains the regression coefficients and summary statistics.

Table 1 contains the 2SLS estimates of the policy effect and the GLY effect on perinatal health. At the mean level of GLY exposure in 2012, the joint introduction of GM seeds and use of GLY (the *policy effect*) led to a 23.3-gram reduction in birthweight, a 1-day reduction in gestation length, a 0.5-percentage-point increase in LBW, a 0.1-percentage-point increase in VLBW, and a 1.6-percentage-point increase in the probability of a preterm birth for the average rural US birth. Comparing the effects on these outcomes to each outcome's mean value reveals that these effects are indeed meaningful: the effects represent 0.7% of mean birthweight, 0.4% of mean gestation length, 6.3% of LBW, 8.6% of VLBW, and 7.7% of preterm probability. The effect on the probability of having a C-section is positive but not statistically significant and small in magnitude. Together, these results suggest that the GLY intensification induced by GM-seed technology—along with any other changes induced by this rollout—on net reduced perinatal health.

We estimate the direct effect of GLY—i.e., the effect of GLY after controlling for other pesticides and unemployment—is approximately 50% larger than the policy effect for each outcome, with another null effect on C-sections. These results indicate GLY exposure signifi-

	Policy effect		GLY effect		
Outcome (*unit*)	Estimate	Conf. Interval	Estimate	Conf. Interval	2012 Mean
Birthweight (*g*)	−23.3	[−39.9, −6.8]	−32.0	[−61.3, −2.8]	3,271.1
Gestation (*days*)	−1.08	[−1.66, −0.49]	−1.54	[−2.63, −0.45]	270.5
LBW (*%pt*)	0.51	[0.10, 0.92]	0.77	[0.09, 1.45]	8.0
VLBW (*%pt*)	0.12	[0.05, 0.19]	0.17	[0.03, 0.32]	1.4
Preterm (*%pt*)	1.59	[0.46, 2.72]	2.24	[0.23, 4.26]	20.7
C-section (*%pt*)	0.76	[−0.23, 1.75]	1.10	[−0.46, 2.67]	27.8

Table 1: Glyphosate's direct and policy effects reduced birth outcomes for the average rural birth. All reported estimates are the effect at the weighted mean of GLY in 2012, where we weight by total births. *Policy effect* is from 2SLS regression of perinatal health on GLY, controlling for family demographics. *GLY effect* is from 2SLS regression of perinatal health on GLY, controlling for other pesticides and unemployment. *LBW* and *VLBW* give the probabilities of low birthweight (<2,500g) and very low birthweight (<1,500g) in percentage points (0–100). See text for details. 95% confidence intervals calculated using standard errors clustered by year and state. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

cantly reduces perinatal health.

The fact that the direct GLY effect exceeds the policy effect suggests that reductions in non-GLY pesticides—likely substituted with GLY and GM seeds—generated health benefits. However, the negative sign of the estimated policy effect implies that these health benefits were smaller than the health costs imposed by the substantial increase in GLY. Unlike Dias, Rocha, and Soares [10] and Skidmore, Sims, and Gibbs [11], we do not find any significant effects of GLY sprayed upstream—Appendix C.10.2 contains the results of our water-based analysis.

Our estimate for the effect of GLY on birthweight is similar in magnitude to other studies on the impact of pollutants on birthweight. Chay and Greenstone [48] report an elasticity of birthweight to air pollution of 0.006; our GLY effect estimates imply an elasticity of 0.007. Currie, Greenstone, and Meckel [49] examines the effects of maternal proximity to fracking sites in Pennsylvania and estimates that living within 1 kilometer of a fracking well reduces birthweight by 39 grams—quite similar to our 2012-mean policy effect of 32 grams.

These findings are robust to a number of modeling alternatives. Figure A3 provides 2SLS estimates for the marginal effect of GLY on birthweight, where we vary the inclusion of controls, fixed effects, and the definition of attainable yield. Event studies for birthweight under these alternative models are also robust (Appendix C.4), and specification charts for other outcomes are in Appendix C.5. We also estimate OLS results without instruments in Appendix C.2 and a "shift-share" model with slightly different identifying assumptions in Appendix C.3. Finally, Figure A18 shows results where we estimate the model using different geographic subsets of the United States. In each case, the results are qualitatively unchanged. A common thread across specifications is that the point estimate increases in magnitude when we control for non-GLY pesticides. As discussed above, GLY replaced other potentially toxic pesticides. Whether one includes/excludes these pesticides in the analysis changes the interpretation of the estimand. Including non-GLY controls gives the direct health effect of GLY; excluding the non-GLY controls provides the policy effect—the net health effect of the introduction of GM seeds and GLY.

We also investigate various threats to identification. A primary concern for our approach is whether the release of GM seeds coincided with non-GLY socioeconomic effects which also affected birthweight—e.g., employment or income. Notably, there is no evidence that the introduction of GM seeds significantly affected average farm or non-farm income in the study counties. The event study for the unemployment rate does show a trend—suggesting the potential for bias if excluded (Figure A24). However, controlling for unemployment does not meaningfully change our results.

Heterogeneity in GLY's health impacts The results in the previous sections estimate the average effect of GLY within a heterogeneous population—where the effect of GLY may differ across individuals. Figure 3 shows the policy and GLY effects on birthweight vary by decile of expected birthweight. Section A.3 explains our prediction of birthweight. Both effects exhibit considerable heterogeneity that follows the same pattern: the magnitude of GLY's estimated effect on birthweight is largest in the lowest decile (i.e., for births with the lowest expected birthweight), and the effect declines as the percentile of predicted birthweight increases. Fig A7 shows first-stage and reduced form results for birthweight by predicted birthweight quintile. The policy effect in the first decile is over 4.5 times larger than in the tenth decile (16 times larger for the GLY effect). Accordingly, GLY's largest impact concentrates among the most vulnerable births: the estimated policy effect in the first decile at mean 2012 GLY exposure is a loss of 54 grams (85 grams for GLY effect), relative to just 12 grams in the tenth decile of predicted birthweight (5 grams for GLY Effect).

We also find significant evidence that GLY's effects vary with expected birthweights on several other outcomes—gestation length and the probabilities of low and very-low birthweight. Fig A6 illlustrates heterogeneity in the policy effect for all outcomes by predicted birthweight quintiles, deciles, and ventiles. As with birthweight, the most acute effects for gestation occur in the lowest percentiles, declining in magnitude with predicted birthweight percentile. This result again suggests that the most vulnerable infants bear the largest impacts. Unlike our birthweight results, where we find no significant effect among higher percentiles, we find statistically significant evidence that GLY reduces gestation length in every decile. At the 2012 mean level of GLY exposure, the policy effect on gestation ranges from -1.5 to -0.8 days for the average birth.

The effect of GLY on low and very-low birthweight is even more concentrated among the lowest predicted birthweights. The estimated policy effect at the 2012 mean level of GLY exposure is a 1.9pp increase in LBW and 0.9pp increase in VLBW—compared to an essentially zero ($<$0.1pp) effect among the tenth decile of predicted birthweights.

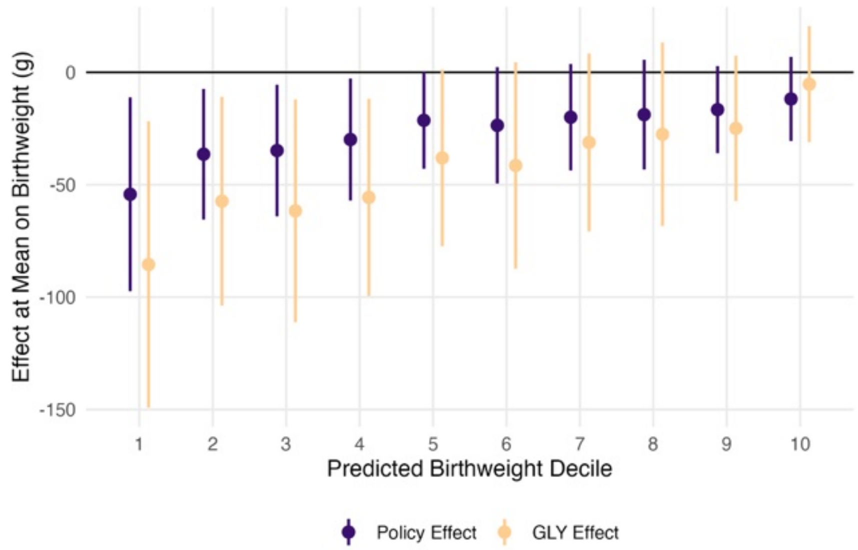


Figure 3: Birthweight losses due to GLY and GM are largest for births with the lowest expected birthweights. Estimates of the Policy and GLY Effects on birthweight by predicted birthweight deciles. Each estimate results from a separate regression. All regressions include controls for family demographics, county fixed effects, and year by month fixed effects. Sample restricted to births occurring in a rural county or to mothers residing in a rural county. Standard errors cluster by year and state. GLY instrumented with GM attainable yield percentile.

Discussion

This study finds significant evidence that GLY adversely affected births across several measures of perinatal health throughout the rural United States in the last twenty years. To our knowledge, these findings are the first quasi-experimental evidence of GLY's adverse health effects at a population scale in the US. However, they are consistent with a growing body of literature documenting GLY's negative impact on development and reproduction. Additionally, using a novel empirical approach, we find significant heterogeneity underlying the effect of GLY on perinatal health: GLY's adverse effects are largest for infants with the lowest expected birthweights.

Heterogeneity, mechanisms, and inequity Several mechanisms potentially explain the observed heterogeneity. While we cannot fully disentangle these mechanisms, we use the dimension along which we observe heterogeneity—predicted birthweight—to understand which attributes correlate with the most-affected births. As predicted birthweight results from a predictive model trained on the demographics of infants' parents, we observe which parental attributes correlate with lower predicted birthweight and, consequently, more adverse GLY effects. Figure 4 reveals that lower predicted birthweight infants are more likely to be female, Black, and/or children of unmarried parents. The predictive model places nearly all births to Black parents into the first quintile, where we detect the most adverse impacts of GLY on perinatal outcomes. If we estimate our main model separately for births to white and non-white mothers, we find that the policy effect is two times larger for births to non-white mothers relative to white mothers (GLY effect 2.6 times larger, see Fig A4). This finding links to an extensive environmental justice literature documenting disparities in pollutant exposure among Black households [46].

We observe similar effects on male and female births *within* birthweight quintiles (see Fig A8). Therefore, the increased effect among low-predicted birthweights is not driven by a larger effect among females. However, female births still bear a larger health burden since they make up a larger proportion of low-predicted births.

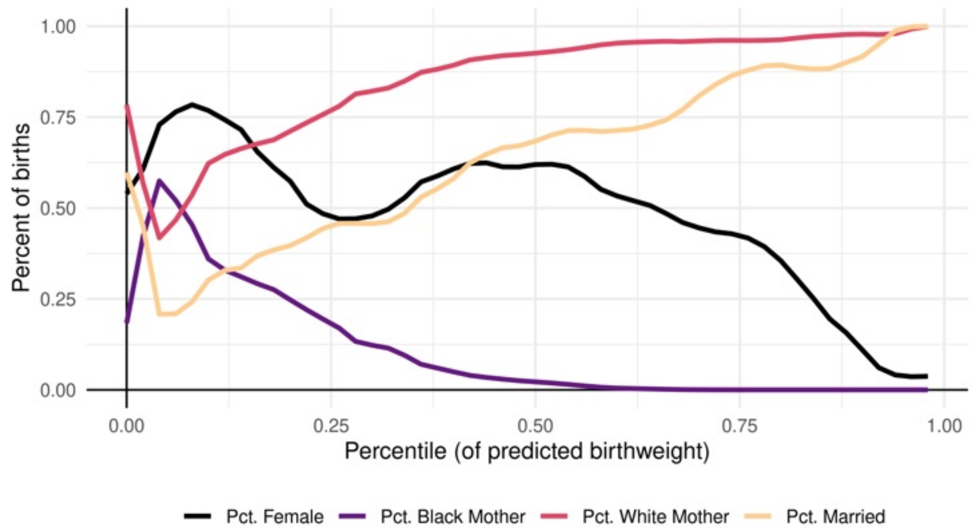


Figure 4: Female infants, children of Black and non-White parents, and children of unmarried parents have lower predicted birthweights. Each line represents the percent of births (*y*-axis) within the demographic group (line color) at each predicted birthweight percentile (*x*-axis) for births to mothers with rural residences. Predicted birthweight percentile is calculated relative to the distribution of rural-residence births in years prior to 1996. Averages cover two-percentile bins.

The fact that the adversity of GLY's effects correlates with race highlights a potentially serious issue for equity. Further, we find the most adverse effects among the lowest expected-weight births—potentially magnifying short-term healthcare costs [50], later-life outcomes/welfare [43], and epigenetic/intergenerational consequences [42]. Consequently, our results have potentially important implications for equity and justice in the US.

Unfortunately, data limitations restrict us from further testing potential mechanisms. While our empirical approach recovers the causal effect of GLY on perinatal health, it does not provide causal estimates for the drivers of heterogeneity. One potential mechanism for the heterogeneity is differential exposure. Because the natality and GLY data are only resolved at the county, we cannot test whether infants with lower predicted birthweights face higher levels of GLY relative to other infants in their county. Differential baseline health (or healthcare access) could also contribute to the observed heterogeneity. Further, the shape of GLY's damage function is unknown. If GLY has nonlinear effects on health—nonlinear in GLY exposure or due to interactions with other health risks/complications—differential exposure or heterogeneous non-GLY health risks could also produce these heterogeneous impacts. Finally, our examination of heterogeneity focuses on only one of many possible dimensions. Future work could contribute to many of these issues with more resolved data.

Broader considerations GM crops and the resulting GLY intensification profoundly changed agriculture. In this study, we quantify one health externality caused by these technological innovations, which reduced average birthweight by 23–32 grams at the average level of GLY exposure.

To put the estimated GLY health damages in perspective, we convert them to dollars. Waitzman, Jalali, and Grosse [51] estimate that a preterm birth costs an additional 82 thousand USD (2023 dollars) relative to a full-term birth. This estimate includes additional medical care following the birth, special education expenses, and lost labor market earnings later in the child's life. These estimates may omit additional costs from GLY. We combine our estimates from Table 1 on the increased probability of preterm birth and the number of total births in 2012. This calculation implies the economic costs of the policy effect were just over 750 million USD annually and nearly 1.1 billion USD annually for the direct GLY effect (both in 2023 dollars).

Our findings, combined with other recent work [10, 11, 27], challenge the prevailing regulatory position that GM crops and their associated agricultural practices are safe—and even beneficial—for health. Advocacy historically argued that GLY is less toxic than the herbicides it replaced. The mounting evidence of the negative health externalities associated with the rollout GM crops and the ensuing GLY intensification warrant new policy discussions about informed, efficient, and equitable regulation of these technologies. Efficient policy must carefully weigh these practices' economic benefits against the adverse health effects we identify and other human/ecological costs. Further, policymakers must consider the unequal burden GLY appears to levy.

Additional work is needed to better understand the benefits and costs of GM crops, GLY, and the specific exposure mechanisms underlying their effects. For instance, unlike Dias, Rocha, and Soares [10] and Skidmore, Sims, and Gibbs [11], we do not find evidence of health effects from upstream GLY use (see Appendix C.10). This difference could result from differences in water treatment, measurement error in upstream-GLY exposure, or other meteorologic/geologic/hydrologic factors.

Uncertainty around GLY and GM health impacts has not slowed the spread of these technologies. Neither has this uncertainty led to substantive monitoring that would enable regulators or researchers to precisely estimate damages, disentangle exposure mechanisms, or understand heterogeneity in GLY's damage function. Despite the relative dearth of data, consistent evidence is emerging that GM-spurred GLY intensification adversely affects health. Nevertheless, without further research, improved monitoring/data, and re-optimized policies, the public will likely continue to bear the health burden of GLY inefficiently and unequally.

References

[1] Zheng Xu, David A. Hennessy, Kavita Sardana, and GianCarlo Moschini. “The Realized Yield Effect of Genetically Engineered Crops: U.S. Maize and Soybean”. *Crop Science* 53.3 (2013), 735–745.

[2] Jayson L. Lusk, Jesse Tack, and Nathan P. Hendricks. *Heterogeneous Yield Impacts from Adoption of Genetically Engineered Corn and the Importance of Controlling for Weather*. Working Paper. 2017. National Bureau of Economic Research: 23519.

[3] Casper Worm Hansen and Asger Mose Wingender. “National and Global Impacts of Genetically Modified Crops”. *American Economic Review: Insights* 5.2 (2023), 224–240.

[4] Andrew R. Kniss. “Long-Term Trends in the Intensity and Relative Toxicity of Herbicide Use”. *Nature Communications* 8.1 (2017), 14865.

[5] Ralf Schulz, Sascha Bub, Lara L. Petschick, Sebastian Stehle, and Jakob Wolfram. “Applied Pesticide Toxicity Shifts toward Plants and Invertebrates, Even in GM Crops”. *Science* 372.6537 (2021), 81–84.

[6] Seungki Lee, GianCarlo Moschini, and Edward D. Perry. “Genetically Engineered Varieties and Applied Pesticide Toxicity in U.S. Maize and Soybeans: Heterogeneous and Evolving Impacts”. *Ecological Economics* 211 (2023), 107873.

[7] Nancy T. Baker and Wesley W. Stone. *Estimated Annual Agricultural Pesticide Use for Counties of the Conterminous United States, 2008-12*. Tech. rep. 907. U.S. Geological Survey, 2015.

[8] Clémentine Antier, Per Kudsk, Xavier Reboud, Lena Ulber, Philippe V. Baret, and Antoine Messéan. “Glyphosate Use in the European Agricultural Sector and a Framework for Its Further Monitoring”. *Sustainability* 12.14 (2020), 5682.

[9] U.S. Environmental Protection Agency. *Glyphosate Interim Registration Review Decision*. Docket ID: EPA-HQ-OPP-2009-0361. 2020.

[10] Mateus Dias, Rudi Rocha, and Rodrigo R Soares. "Down the River: Glyphosate Use in Agriculture and Birth Outcomes of Surrounding Populations". *The Review of Economic Studies* 90 (6 2023), 2943–2981.

[11] Marin Elisabeth Skidmore, Kaitlyn M. Sims, and Holly K. Gibbs. "Agricultural Intensification and Childhood Cancer in Brazil". *Proceedings of the National Academy of Sciences* 120.45 (2023), e2306003120.

[12] World Bank Open Data. *GDP per capita*.

[13] Christina Gillezeau, Maaike van Gerwen, Rachel M. Shaffer, Iemaan Rana, Luoping Zhang, Lianne Sheppard, and Emanuela Taioli. "The evidence of human exposure to glyphosate: a review". *Environmental Health* 18.1 (2019).

[14] Laura Medalie, Nancy T. Baker, Megan E. Shoda, Wesley W. Stone, Michael T. Meyer, Edward G. Stets, and Michaelah Wilson. "Influence of Land Use and Region on Glyphosate and Aminomethylphosphonic Acid in Streams in the USA". *Science of The Total Environment* 707 (2020), 136008.

[15] Nancy B. Ramirez Haberkon, Virginia C. Aparicio, and Mariano J. Mendez. "First Evidence of Glyphosate and Aminomethylphosphonic Acid (AMPA) in the Respirable Dust (PM10) Emitted from Unpaved Rural Roads of Argentina". *Science of The Total Environment* 773 (2021), 145055.

[16] Yuzhao Hu, Shuai Wu, Wei Lyu, Jun Ning, and Dongmei She. "Risk assessment of human exposure to airborne pesticides in rural greenhouses". *Scientific Reports* 13.1 (2023), 5138.

[17] Amelie Foucault, Nicolas Vallet, Noemie Ravalet, Frederic Picou, Marie C Bene, Emmanuel Gyan, and Olivier Herault. "Occupational pesticide exposure increases risk of acute myeloid leukemia: a meta-analysis of case–control studies including 3,955 cases and 9,948 controls". *Scientific Reports* 11.1 (2021), 2007.

[18] Chris Pappas and Brenda Foos. "Pesticide data program: 30 years of food residue data and trends". *Journal of Exposure Science & Environmental Epidemiology* 33.5 (2023), 805–812.

[19] W.A. Battaglin, M.T. Meyer, K.M. Kuivila, and J.E. Dietze. “Glyphosate and Its Degradation Product AMPA Occur Frequently and Widely in U.S. Soils, Surface Water, Groundwater, and Precipitation”. *JAWRA Journal of the American Water Resources Association* 50.2 (2014), 275–290.

[20] Stephen O Duke and Stephen B Powles. “Glyphosate: A Once-in-a-Century Herbicide”. *Pest Management Science* 64.4 (2008), 319–325.

[21] Andreas Hartmann, Scott Jasechko, Tom Gleeson, Yoshihide Wada, Bartolomé Andreo, Juan Antonio Barberá, Heike Brielmann, Lhoussaine Bouchaou, Jean-Baptiste Charlier, W. George Darling, Maria Filippini, Jakob Garvelmann, Nico Goldscheider, Martin Kralik, Harald Kunstmann, Bernard Ladouche, Jens Lange, Giorgia Lucianetti, José Francisco Martín, Matías Mudarra, Damián Sánchez, Christine Stumpp, Eleni Zagana, and Thorsten Wagener. “Risk of Groundwater Contamination Widely Underestimated Because of Fast Flow into Aquifers”. *Proceedings of the National Academy of Sciences* 118.20 (2021), e2024492118.

[22] Philip J. Landrigan. “Pesticides and Human Reproduction”. *JAMA internal medicine* 178.1 (2018), 26–27.

[23] Maria Ospina, Andre Schütze, Pilar Morales-Agudelo, Meghan Vidal, Lee-Yang Wong, and Antonia M. Calafat. “Exposure to Glyphosate in the United States: Data from the 2013–2014 National Health and Nutrition Examination Survey”. *Environment International* 170 (2022), 107620.

[24] S. Parvez, R. R. Gerona, C. Proctor, M. Friesen, J. L. Ashby, J. L. Reiter, Z. Lui, and P. D. Winchester. “Glyphosate Exposure in Pregnancy and Shortened Gestational Length: A Prospective Indiana Birth Cohort Study”. *Environmental Health* 17.1 (2018), 23.

[25] Roy R. Gerona, Jill L. Reiter, Igor Zakharevich, Cathy Proctor, Jun Ying, Robin Mesnage, Michael Antoniou, and Paul D. Winchester. “Glyphosate Exposure in Early Pregnancy and Reduced Fetal Growth: A Prospective Observational Study of High-Risk Pregnancies”. *Environmental Health* 21.1 (2022), 95.

[26] T. E. Arbuckle, Z. Lin, and L. S. Mery. "An Exploratory Analysis of the Effect of Pesticide Exposure on the Risk of Spontaneous Abortion in an Ontario Farm Population". *Environmental Health Perspectives* 109.8 (2001), 851–857.

[27] Adriana Camacho and Daniel Mejia. "The Health Consequences of Aerial Spraying Illicit Crops: The Case of Colombia". *Journal of Health Economics* 54 (2017), 147–160.

[28] Eliane Dallegrave, Fabiana DiGiorgio Mantese, Ricardo Soares Coelho, Janaína Drawans Pereira, Paulo Roberto Dalsenter, and Augusto Langeloh. "The Teratogenic Potential of the Herbicide Glyphosate-Roundup® in Wistar Rats". *Toxicology Letters* 142.1 (2003), 45–52.

[29] Alejandra Paganelli, Victoria Gnazzo, Helena Acosta, Silvia L. López, and Andrés E. Carrasco. "Glyphosate-Based Herbicides Produce Teratogenic Effects on Vertebrates by Impairing Retinoic Acid Signaling". *Chemical Research in Toxicology* 23.10 (2010), 1586–1595.

[30] Thu Ha Pham, Lohann Derian, Christine Kervarrec, Pierre-Yves Kernanec, Bernard Jégou, Fatima Smagulova, and Aurore Gely-Pernot. "Perinatal Exposure to Glyphosate and a Glyphosate-Based Herbicide Affect Spermatogenesis in Mice". *Toxicological Sciences* 169.1 (2019), 260–271.

[31] Juan P. Munoz, Tammy C. Bleak, and Gloria M. Calaf. "Glyphosate and the Key Characteristics of an Endocrine Disruptor: A Review". *Chemosphere* (2020).

[32] N. Benachour, H. Sipahutar, S. Moslemi, C. Gasnier, C. Travert, and G. E. Séralini. "Time- and Dose-Dependent Effects of Roundup on Human Embryonic and Placental Cells". *Archives of Environmental Contamination and Toxicology* 53.1 (2007), 126–133.

[33] Ashley E. Larsen, Steven D. Gaines, and Olivier Deschenes. "Agricultural Pesticide Use and Adverse Birth Outcomes in the San Joaquin Valley of California". *Nature Communications* 8.1 (2017), 302.

[34] Benjamin A. Jones. “Invasive Species Control, Agricultural Pesticide Use, and Infant Health Outcomes”. *Land Economics* 96.2 (2020), 149–170.

[35] Charles A Taylor. “Cicadian Rhythm: Insecticides, Infant Health and Long-term Outcomes” (2022).

[36] Eyal Frank. “The Economic Impacts of Ecosystem Disruptions: Costs From Substituting Biological Pest Control” (2021).

[37] Eyal Frank and Charles A Taylor. “The “Golden Age” of Pesticides? Trade-offs of DDT and Health in the US” (2022).

[38] Joan Calzada, Meritxell Gisbert, and Bernard Moscoso. “The Hidden Cost of Bananas: The Effects of Pesticides on Newborns’ Health”. *Journal of the Association of Environmental and Resource Economists* 10.6 (2023), 1623–1663.

[39] Jason Fletcher and Hamid Noghanibehambari. “The Siren Song of Cicadas: Early-life Pesticide Exposure and Later-Life Male Mortality”. *Journal of Environmental Economics and Management* 123 (2024), 102903.

[40] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist’s companion*. Princeton University Press, 2009.

[41] Dennis Engist, Laura Melissa Guzman, Ashley Larsen, and Frederik Noack. “The impact of genetically modified crops on bird diversity”. Working Paper. 2023.

[42] Bastiaan T Heijmans, Elmar W Tobi, Aryeh D Stein, Hein Putter, Gerard J Blauw, Ezra S Susser, P Eline Slagboom, and LH Lumey. “Persistent epigenetic differences associated with prenatal exposure to famine in humans”. *Proceedings of the National Academy of Sciences* 105.44 (2008), 17046–17049.

[43] Prashant Bharadwaj, Petter Lundborg, and Dan-Olof Rooth. “Birth Weight in the Long Run”. *Journal of Human Resources* 53.1 (2017), 189–231.

[44] National Center for Health Statistics. *Restricted-Use Vital Statistics Data*. As compiled from data provided by the 57 statistics jurisdications through the Vital Statistics Cooperative Program. 1990–2013.

[45] FAO and IIASA. *Global Agro Ecological Zones Version 4 (GAEZ V4)*. http://www.fao.org/gaez/en.

[46] Spencer Banzhaf, Lala Ma, and Christopher Timmins. "Environmental Justice: The Economics of Race, Place, and Pollution". *Journal of Economic Perspectives* 33.1 (2019), 185–208.

[47] Felix Elwert and Christopher Winship. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable". *Annual Review of Sociology* 40.1 (2014), 31–53.

[48] Kenneth Y. Chay and Michael Greenstone. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession". *The Quarterly Journal of Economics* 118.3 (2003), 1121–1167.

[49] Janet Currie, Michael Greenstone, and Katherine Meckel. "Hydraulic Fracturing and Infant Health: New Evidence from Pennsylvania". *Science Advances* 3.12 (2017), e1603021.

[50] Andrew L. Beam, Inbar Fried, Nathan Palmer, Denis Agniel, Gabriel Brat, Kathe Fox, Isaac Kohane, Anna Sinaiko, John A. F. Zupancic, and Joanne Armstrong. "Estimates of healthcare spending for preterm and low-birthweight infants in a commercially insured population: 2008–2016". *Journal of Perinatology* 40.7 (2020), 1091–1099.

[51] Norman J. Waitzman, Ali Jalali, and Scott D. Grosse. "Preterm Birth Lifetime Costs in the United States in 2016: An Update". *Seminars in Perinatology*. Economic Considerations in Neonatal Intensive Care 45.3 (2021), 151390.

[52] Gale Thelin and Wesley Stone. *Estimation of Annual Agricultural Pesticide Use for Counties of the Conterminous United States, 1992-2009*. Scientific Investigations Report 2013-5009. U.S. Geological Survey, 2013.

[53] *USDA ERS - Documentation*. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/.

[54] Wolfram Schlenker and Michael J. Roberts. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change". *Proceedings of the*

National Academy of Sciences of the United States of America 106.37 (2009), 15594–15598.

[55] Bernhard Lehner and Günther Grill. “Global River Hydrography and Network Routing: Baseline Data and New Approaches to Study the World’s Large River Systems”. *Hydrological Processes* 27.15 (2013), 2171–2186.

[56] Soil Survey Staff. *Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States*. 2021.

[57] PRISM Climate Group. "*Prism Gridded Climate Data*". 2021.

[58] CIESIN. *U.S. Census Grids (Summary File 1), 2010*. 2017.

[59] Jorge Fernandez-Cornejo, Seth Wechsler, Mike Livingston, and Lorraine Mitchell. *Genetically Engineered Crops in the United States*. SSRN Scholarly Paper ID 2503388. Rochester, NY: Social Science Research Network, 2014.

[60] KM Baer and BJ Marcel. *Glyphosate*. 2014.

[61] Birgit Muller. “Glyphosate-A Love Story. Ordinary Thoughtlessness and Responseability in Industrial Farming”. *Journal of Agrarian Change* 21.1 (2020), 160–179.

[62] Geoffrey Barrows, Steven Sexton, and David Zilberman. “Agricultural Biotechnology: The Promise and Prospects of Genetically Modified Crops”. *Journal of Economic Perspectives* 28.1 (2014), 99–120.

[63] Arthur Grube, David Donaldson, Timothy Kiely, and La Wu. “Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates” (2011).

[64] EPA. “Glyphosate: Response to Comments, Usage, and Benefits” (2019).

[65] Tyler J. Lark, Nathan P. Hendricks, Aaron Smith, Nicholas Pates, Seth A. Spawn-Lee, Matthew Bougie, Eric G. Booth, Christopher J. Kucharik, and Holly K. Gibbs. “Environmental Outcomes of the US Renewable Fuel Standard”. *Proceedings of the National Academy of Sciences* 119.9 (2022), e2101084119.

[66] United Nations. *UN-Water SDG 6 Data Portal*. https://sdg6data.org/. Online; accessed 11 June 2024. 2024.

Appendix A Methods

A.1 Data

Infant health data We have the universe of births in the United States between 1990 and 2019 from the National Vital Statistics System (NVSS). The birth data contain information recorded on the birth certificates, including month and year of birth, sex, birthweight, APGAR score, live birth order, total birth order, whether the birth was a C-section, gestation length, and dummies for a battery of birth defects. The birth certificates also contain certain demographics of the mother and father—namely race, ethnicity, age, education, marital status, and residence status. We have access to the restricted versions of these files, which identify the county of birth and county of mother's residence for all births, compared to the publicly available data, which hides geographic identifiers for counties with less than 100,000 residents.

Pesticide Use Estimates Our measure of GLY use comes from the United States Geological Survey's Pesticide National Synthesis Project [7, 52]. The USGS surveyed farmers to calculate pesticide use rates per acre of different crops planted at the crop reporting district level. They then multiply these usage rates by the total acreage of each crop within the crop reporting district to estimate the total amount of each pesticide used, measured in kilograms. Each pesticide has two estimates, high and low, where the high value assumes crop reporting districts with a missing usage rate for a pesticide applied the pesticide at the same rate as their neighbors on each crop. The low value assumes a missing usage rate for a pesticide means that farmers did not apply that pesticide. We use the high value throughout our analysis. Additionally, we normalize by the total area of each county, thus our measure of GLY and other pesticides are in kilograms of active ingredient per square kilometer (kg/km^2).

Attainable Yield We use the FAO-GAEZ attainable yield for soybeans, corn, and cotton to measure the suitability of a county for genetically modified crops [45]. These data assign potential yield values to one square kilometer pixels based on environmental factors such as soil type, slope, and climate. We aggregate the pixels to the county by taking the average of all pixels within a county. We take the difference between the high-input attainable yield and the low-input attainable yield to focus on the counties with the largest incentive to adopt GM crops. The underlying model calculates the high-input scenario assuming that farmers have access to modern technology for crop management, including GM seeds. Whereas the model calculates the low-input scenario assuming more traditional farming methods. We aggregate the three GM crops by standardizing the attainable yield difference so that each crop has a mean of zero and a standard deviation of one. We then take the simple average across the standardized yield differences, then re-scale this average into national percentiles.

Other Data We supplement our analysis with several additional data sources. First, we use the 2003 Rural-Urban Continuum codes from the US Department of Agriculture (USDA) to classify counties as rural. A rural county is any non-metro county, where the USDA defines a metro county as, "broad labor-market areas that include central counties with one or more urban areas with populations of 50,000 or more people. They also include outlying counties that are economically tied to the core counties as measured by labor-force commuting" [53].

We get workforce data from the Bureau of Labor Statistics (BLS), specifically their Local Area Unemployment Statistics. These data report the annual average number of employed and unemployed workers going back to 1990. We supplement these data with annual, county-level income and employment data split by farm and non-farm from the Bureau of Economic Analysis (BEA). They report these data going back to 1969 and source the data from the US Bureau of Labor Statistics and Internal Revenue Service.

Next, we get annual county-level acreage and yield of various crops—including the main GM crops, corn, soy, and cotton—from the USDA National Agricultural Statistics Service (NASS). One issue with these data is that if there are few farms in a particular county, then the USDA will mask the data for privacy reasons. Therefore, we aggregate the acreage and yield data up to the Agricultural-Statistics-District-level when using these data in section C.8.

A.2 Empirical Strategy

We aim to estimate the causal effect of GLY exposure—induced by the rollout of GM seeds—on infant health. To isolate exogenous variation in GLY use, we leverage temporal variation due to the commercial release of GM crops in 1996 and spatial variation in GM adoption due to differences in the suitability of the environment for growing those crops.

A linear model for the effect of local GLY intensity $\left(\beta^l\right)$ and the effect of upstream GLY intensity $\left(\beta^u_d\right)$ on perinatal health is

$$\text{Health}_{ijt} = \beta^l \text{GLY}^l_{jt} + \sum_d \beta^u_d \text{GLY}^u_{jtd} + \Gamma X_{ijt} + \alpha_j + \lambda_t + \varepsilon_{ijt}. \tag{1}$$

for individual i, in county j, in year t. GLY^l_{jt} represents local GLY exposure, measured as the total mass (kg) of GLY sprayed per square kilometer in county j in year t. GLY^u_{jtd} denotes estimated exposure to GLY from GLY applied in upstream distance bin d of county j in year t. X_{ijt} provides a vector of controls. α_j and λ_t are county and month-of-sample (e.g., January 2012) fixed effects.

Estimating Equation (1) with OLS is unlikely to identify the true effect of GLY on health due to measurement error and endogeneity.

While we have micro-data (birth-level) on birth outcomes, we do not have precisely measured GLY exposure. We expect some mothers within a county are highly exposed to GLY while others are not. However, data limitations force us to assign the same level of exposure to all births within a county. This measurement concern relates to the ecological fallacy and would likely lead us to underestimate the magnitude of GLY's damages to an individual's health. Banzhaf, Ma, and Timmins [46] write "When measuring the correlation between pollution and demographics, the 'ecological fallacy' can arise when inferring relationships between individual units (like households) from larger, more aggregated units (like counties) that contain those units." The main endogeneity concern is that the adoption of GM technology and GLY may correlate with unobservable factors that affect perinatal health.

To rectify the measurement and endogeneity issues, we use instruments—in a 2SLS estimator—that isolate exogenous variation in both local and upstream GLY. Our instruments are the percentile of attainable yield for GM crops for county j, denoted GM^l_j, interacted with dummy variables for each year. GM^u_{jd} is a measure of the attainable yield for watersheds upstream of county j in distance bin d. We describe our construction of GM^l_c in section A.1 and GM^u_{jd} in section A.4.

Identification Our model includes county and month-of-sample fixed effects. Stated as a parallel-trends assumption, the identifying assumption is: If GM crops had not been released, then the *difference* between high and low attainable yield counties would have remained constant in terms of GLY use and infant health (conditional on fixed effects and controls). Below we discuss threats to our identification strategy.

Differential trends in high and low GM attainable yield counties, prior to the release of GM crops, would violate our estimation approach. Additionally, a single outlier year for higher or lower yield counties that drives the average result could affect our results despite the smooth increase in GLY over time. In order to assess these concerns, we estimate an event study model,

$$\text{Health}_{ijt} = \sum_{\tau \neq 1995} \left(\gamma_\tau^l \text{GM}_j^l \times \mathbb{1}(t = \tau) + \sum_d \gamma_\tau^d \text{GM}_{jd}^u \times \mathbb{1}(t = \tau) \right) + \Gamma X_{ijt} + \alpha_j + \lambda_t + \varepsilon_{ijt}, \quad (2)$$

The measured effects (γ_τ) represent the difference in perinatal health between the highest and lowest attainable yield counties relative to their 1995 difference (the year before the GM rollout and the ensuing GLY intensification). The county-level fixed effects absorb average differences between higher and lower yield counties. Consequently, an event study with no trend in the γ_τ^l prior to treatment (pre-1996) supports our identifying assumption. Figure 2 suggests this identifying assumption is plausible in our context: the estimated γ_τ^l are near zero, do not reject zero, and do not suggest a pre-treatment trend.

As described in the Empirical Approach Section, we estimate Equation (1) with two-stage least squares (2SLS). Interpreting the 2SLS estimates as causal requires that our instrument satisfy an exclusion restriction [40]. The exclusion restriction in our context: Our instruments, GM attainable yield interacted with year, only affect infant health through changes in GLY use, conditional on our controls. Non-GLY pesticides that farmers replaced with GLY potentially violate this exclusion restriction. Consequently, our 2SLS approach would underestimate the effect of GLY on health because it measures the effect of GLY relative to the profile of herbicides farmers used before adopting genetically modified crops and GLY. Accordingly, we present estimates with controls for non-GLY pesticides (the *GLY effect*) and without non-GLY controls (the *policy effect*). Table A1 compares summary statistics for high-yield counties, low-yield counties, and urban counties before the release of GM crops. Births in high- and low-yield rural counties were quite similar during this period—as were any other outcomes.

Finally, settings with potentially heterogeneous treatment effects require an additional assumption: monotonicity. In our context, monotonicity requires that increasing attainable yield in a county would result in that county using weakly more GLY—i.e., increasing suitability for GM crops would not reduce a counties' GLY intensity. This assumption appears quite reasonable in our setting, and we have no reason to believe that *decreasing* a county's suitability for corn, soy, or cotton would increase GLY, all else equal.

A.3 Predicting birthweight

Models like Equation (1) identify a single, *average* effect (β^{local})—as in Table A2. If individual infants respond differently to GLY exposure, an average effect obscures this heterogeneity. Further, this heterogeneity is potentially critical to understanding the health impacts and policy remedies of the documented losses of birthweight—for example, if the lost birthweight came from lower-weight infants, as opposed to higher-weight infants. Thus, *who* lost weight could be key.

We estimate heterogeneity in the effect of GM seeds and GLY on infant birthweight and gestation as a function of the infant's birthweight percentile—with two nuances. The first nuance centers on an issue of causal identification; the second is operational.

First, rather than using an infant's *actual* birthweight percentile, we use the percentile of the infant's *predicted* birthweight. We train a random forest to predict birthweight using features from the NVSS that describe the infant, mother, father, and birth location.[1] The training sample for this prediction is the universe births in the contiguous US before the mass introduction of GM crops and the ramp-up of GLY application (before 1996)—i.e., before our *treatment* began. We tune[2] the model on 80% of the pre-1996 births using five-fold cross-validation. Finally, we train the selected (minimum RMSE) model in a five-fold pattern—ensuring each prediction comes from a model that has not seen the predicted individual. This hold-out approach in the prediction step and our predicted birthweight approach help avoid bias in the heterogeneity regressions. This bias could arise because birthweight—our heterogeneity dimension—is the outcome variable; conditioning on the outcome can introduce endogeneity [47]. Instead, we condition on *predicted* birthweight, which is a function of (1) the infant's family's observable features and (2) *other*, pre-treatment infants' birthweights. Because our predictions are relatively accurate (predicted birthweight is, on average, quite close to actual birthweight as shown in Figure A5), we can estimate how GLY differentially affects lower- and higher-birthweight infants without introducing bias from conditioning on the outcome.

The second nuance relates to the structure of the heterogeneous treatment effect. Because the shape of the heterogeneity is unknown, we take a semi-parametric approach that allows us to remain relatively agnostic. We split the sample using infants' predicted birthweight percentiles (e.g., quintiles) and then separately estimate the 2SLS model for each group. For example, Figure 3 contains the estimated effect of GLY on birthweight for the first through tenth deciles. This bin-based approach is commonly applied to recover heterogeneous treatment effects, as it allows one to approximate arbitrary nonlinear functions without making stronger assumptions about functional form (for example, Schlenker and Roberts [54]). This approach still returns an *average* treatment effect estimate *within* each bin/group. Fig-

[1] These features are the infant's sex; the parents' races, ethnicities, ages, and marital status; the mother's education, residence status, plurality, and tobacco use; the birth location's state; whether the birth location or mother's home are in a rural county; whether the birth occurred in a facility; and month of year. When missing, we impute these features' values.

[2] We tune the number of random features selected and the minimum number of observations in a terminal node.

ure A6 shows robustness to different bins in predicted birthweight across all of our main outcomes.

A.4 Exposure to GLY sprayed upstream through water

We use a spatial water model to estimate GLY exposure based on the amount of GLY sprayed upstream of each county using a methodology similar to that of Dias, Rocha, and Soares [10]. The HydroBASINS watershed shapes form the building blocks of this model [55]. We summarize the process here but leave the details in the appendix section C.10.

The amount of GLY that runs off into surface water will be affected by the erodibility of the soil, the slope of the land, and precipitation. We collect soil erodibility and slope data from the USGS gridded soil survey in each watershed, which we aggregate to the watershed level by taking the average over all grid cells within a watershed [56]. We take the interaction of soil erodibility and slope and then convert that interaction into a percentile based on the distribution from all watersheds in the US. These data are static and do not change over time.

Next, we use gridded monthly precipitation from the PRISM climate group to capture whether there was potential for GLY to run off [57]. We aggregate rainfall during the growing season (April through September) for each watershed and again convert it into a percentile from the distribution of all watershed-month-years.

We then map our county-level attainable yield percentile to watersheds and take the interaction between high erodibility, high precipitation, and attainable yield to create an instrument for upstream GLY use. We expect there to be effects from upstream spraying only when there is both high soil erodibility and high precipitation. We also estimate the effect of high attainable yield upstream without the interaction with soil erodibility and precipitation. The HydroBASINS data allows for easy linking of upstream and downstream watersheds. In the linking process, we calculate the distance between two watersheds by tracing along centroids of all watersheds between those two watersheds. This measure allows us to aggregate variables over 50-kilometer distance bins upstream and downstream from each watershed. The downstream variables serve as a nice placebo test since we do not expect GLY sprayed downstream to affect infant health.

Once we estimate values upstream of each watershed, we must aggregate to the county level to analyze them with our health metrics. We take the weighted average of the upstream variables for all watersheds in a county, where the weights are the portion of the county's population that lives within that watershed. We use population estimates for one square kilometer pixels from SEDAC to calculate the population weights for each watershed

[58].

In addition to the binned approach described above, we use a machine-learning model to predict concentrations of GLY in surface water using a limited dataset of samples taken across the US. We utilize the geographic structure and physical characteristics of land, along with spatially disaggregated herbicide use, to predict downstream concentrations of GLY in surface water. We then regress perinatal health outcomes on these predictions og GLY and AMPA exposure from water. The details are in appendix C.10.

Appendix B Extended Data

	Rural				Non-rural	
	High GM Yield		*Low GM Yield*			
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Number of Counties	863	0	1160	0	1086	0
Birthweight (*g*)	3354	83.3	3390	89.5	3379	67
Percent LBW	7.7	2.17	6.18	2	7.05	1.66
Percent Male	51.1	1.85	51.3	2.98	51.3	1.15
Total Births	369	297	253	287	3076	8185
GLY (kg/km^2)	0.0026	0.0029	0.0011	0.0013	0.0020	0.0035
Total Population (1000s)	27.3	21.1	19.3	20.9	197	443
Percent Hispanic	1.49	3.2	7.03	15.3	4.97	9.99
Unemployment Rate	6.96	2.54	6.82	3.85	6.15	2.61
Pct. Some HS Degree	35.5	9.11	30.2	10.6	26.8	9.35
Pct. HS Degree	35.7	6.09	34.5	5.82	33.1	6.29
Pct. Some College	18.5	4.34	22.6	6.18	23.5	5.61
Pct. College Degree	10.4	3.54	12.8	5.27	16.7	8.11
Income per Capita	16.5	2.16	17.1	3.35	20	4.38

Table A1: Summary statistics for high- and low-attainable yield counties between 1992 and 1995
Means and standard deviations are calculated on county-year level averages between 1992 and 1995. High GM yield are rural counties with above median attainable yield for GM crops, and low GM yield are rural counties with below median attainable yield for GM crops. Rural vs non-rural defined using USDA rural-urban continuum codes from 2003.

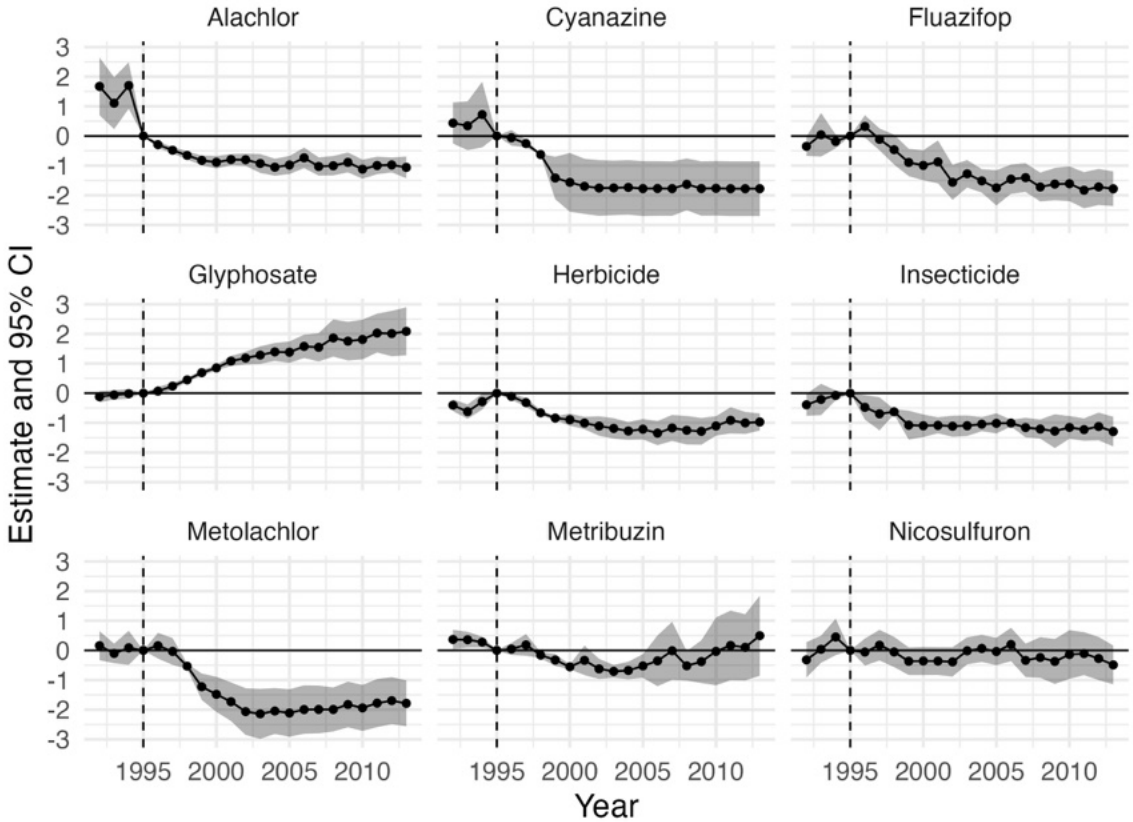


Figure A1: Counties with high suitability for GM crops increased GLY intensity and reduced non-GLY pesticides with the introduction of GLY-resistant seeds. Each event study come from separate regressions where the given pesticide is regressed on local GM attainable yield percentile interacted with year dummies with year and county fixed effects. All coefficients are scaled by the standard deviation of their respective variables. *Herbicide* and *Insecticide* each aggregate all other herbicides and insecticides not individually analyzed. Results from rural US counties. Standard errors are cluster by state and year. A unit of observation is county by year; regressions weight by total number of births.

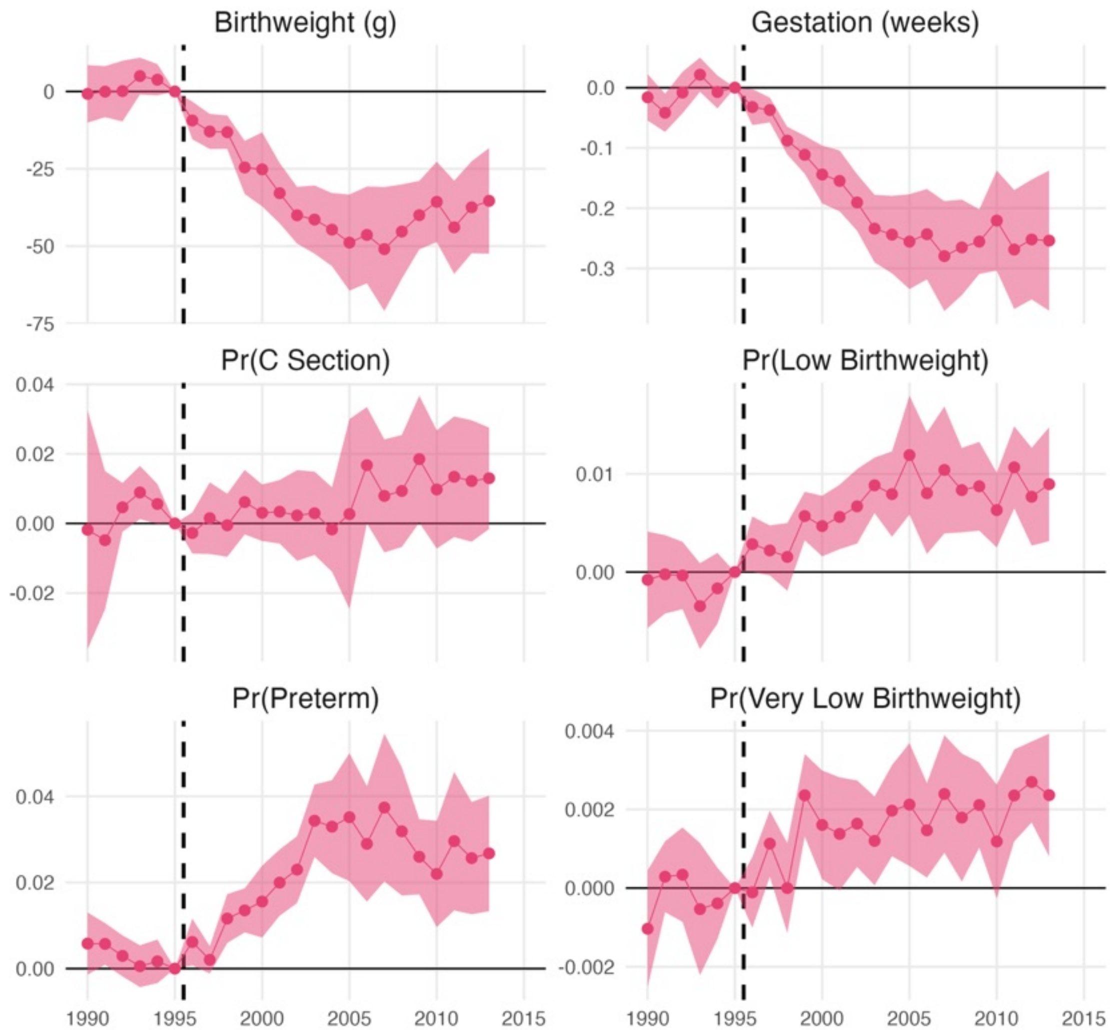


Figure A2: Perinatal health declined in GM-crop suitable counties after the introduction of GLY-resistant seeds The subfigures extend Figure 2b to additional health outcomes—i.e., the estimated effect of local GM attainable yield percentile on perinatal health outcomes relative to 1995. All regressions include county and year by month fixed effects and cluster errors by state and year. All regressions also control for family demographics, including mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

	BW	LBW	VLBW	Gestation	Preterm	C-section
Panel A: Policy effect						
GLY/km^2	–1,002.2	0.220	0.052	–6.60	0.685	0.326
	(362.3)	(0.090)	(0.016)	(1.83)	(0.248)	(0.217)
Controls						
Pesticides	N	N	N	N	N	N
Unemployment	N	N	N	N	N	N
Panel B: GLY effect						
GLY/km^2	–1,376.2	0.330	0.075	–9.46	0.963	0.474
	(640.7)	(0.148)	(0.033)	(3.42)	(0.442)	(0.344)
Controls						
Pesticides	Y	Y	Y	Y	Y	Y
Unemployment	Y	Y	Y	Y	Y	Y
Fixed-effects *(Both panels)*						
Family Demog	Y	Y	Y	Y	Y	Y
County	Y	Y	Y	Y	Y	Y
Yr $\times$ Mo	Y	Y	Y	Y	Y	Y
Summaries *(Both panels)*						
N (millions)	10.73	10.73	10.73	10.71	10.71	9.510
2012 mean	3,271.1	0.081	0.014	38.6	0.207	0.278

Table A2: 2SLS estimates of the policy and direct GLY effects on perinatal health. Each coefficient estimate (column-panel combination) provides results from a separate 2SLS regression. The six outcomes are birthweight (BW), the probabilities of low birthweight (LBW; BW $<$ 2500g) and very low birthweight (VLBW; BW $<$ 2500g), gestation length, and the probability of a preterm birth (gestation $<$ 37 weeks). Both panels include family demographic, county, and year by month fixed effects. GLY effect (Panel B) additionally controls for other pesticides and unemployment. Sample restricted to births occurring in rural counties or to mothers residing in rural counties. Instruments are the attainable yield percentile for GM crops in each county interacted with year. Family demographic controls include mother's age, mother's race, mother's origin, mother's education, sex of child, total birth order, mother's residence status, and birth facility. Pesticide controls include alachlor, atrazine, cyanizine, fluazifop, metolachlor, metribuzin, and nicosulfuron. GLY/km^2 is kg/km^2. Standard errors in parentheses. We two-way cluster errors by year and state.

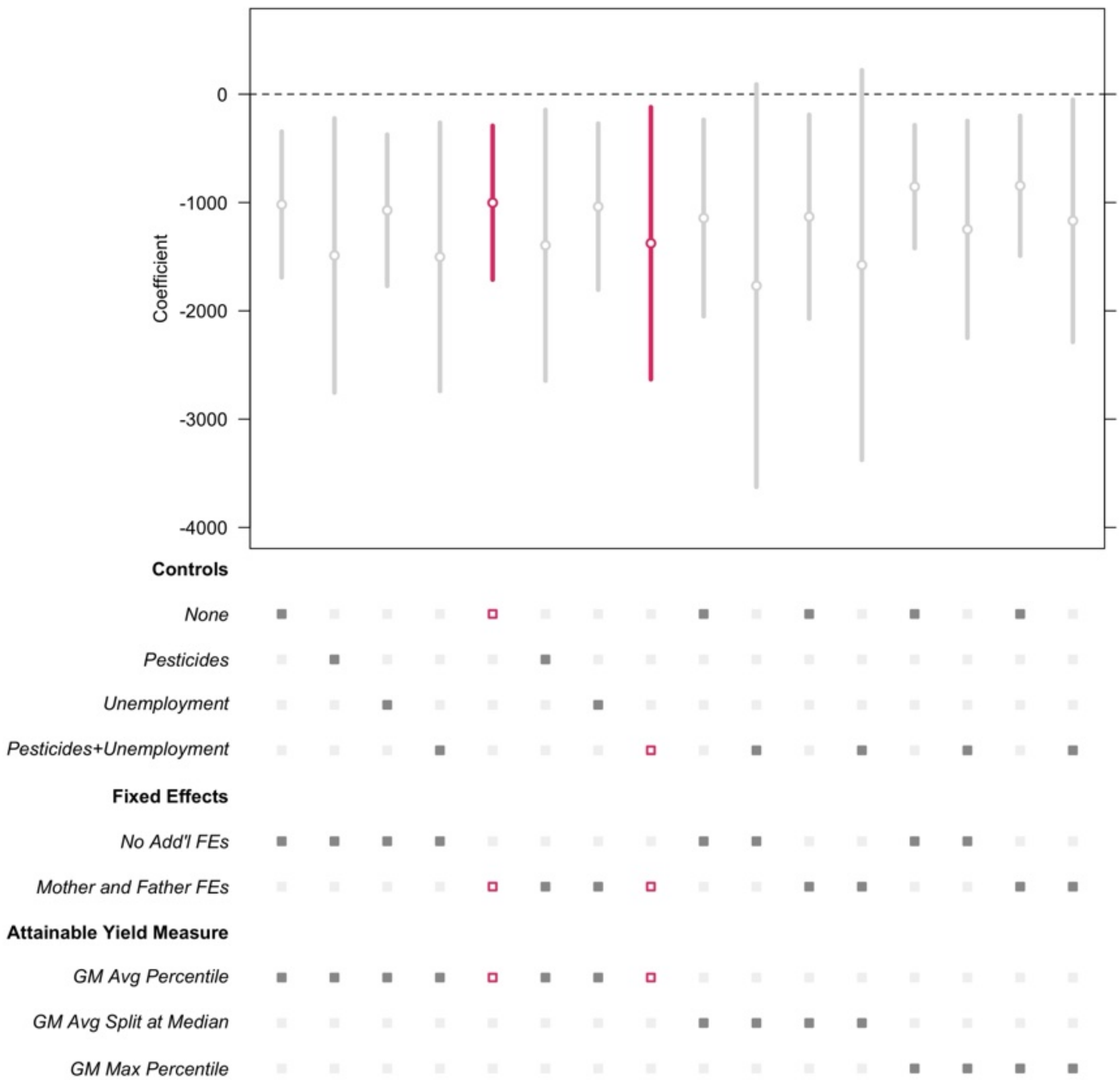


Figure A3: The estimated effect of GLY on birthweight is robust to alternative specifications. Coefficients are the estimated marginal effect of GLY (kg/km^2) on birthweight. Our main specifications are highlighted. All regressions include county and year by month fixed effects, standard errors are clustered by state and year. Pesticide controls include Mother and Father FE's include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. We vary the construction of GM attainable yield: "GM Avg Percentile" is our main specification, "GM Average, Split at Median" uses a binary high vs low attainable yield, where a county is high attainable yield if they are above the median attainable yield, and "GM Max Percentile" takes the maximum standardized attainable yield among corn, soy, and cotton (rather than the average) before re-scaling into a percentile. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

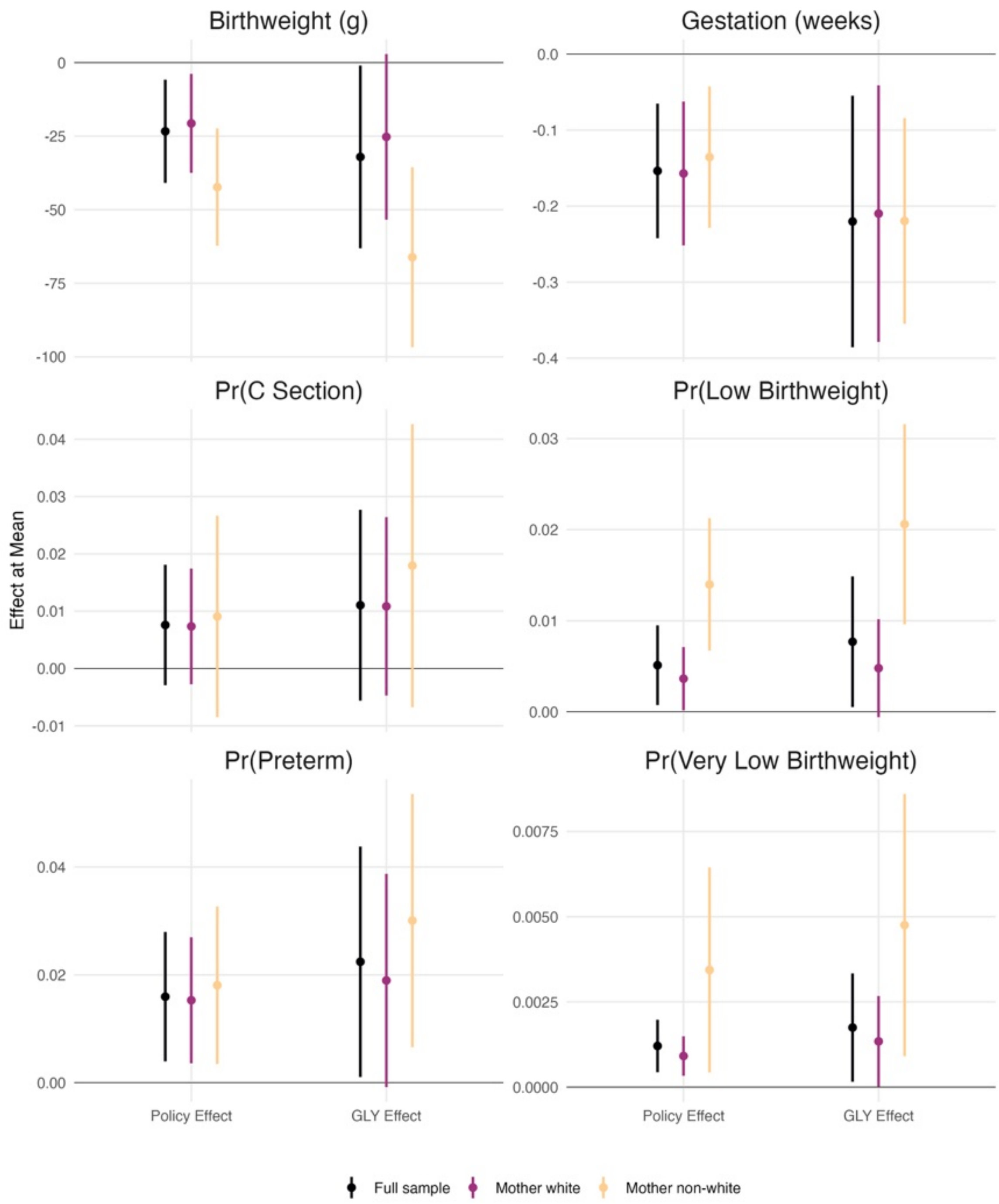


Figure A4: GLY effects for infants born to non-white mothers are larger for birthweight and for the probabilities of preterm birth, LBW, and VLBW. Policy and GLY effects for all outcomes at the mean level of GLY in 2012, estimated separately by mother's race. All regressions include county and year by month fixed effects, and control for family demographics. Standard errors are clustered by state and year. The GLY Effect adds controls for other pesticides and unemployment. The sample is restricted to births occurring in rural counties or to mothers residing in rural counties.

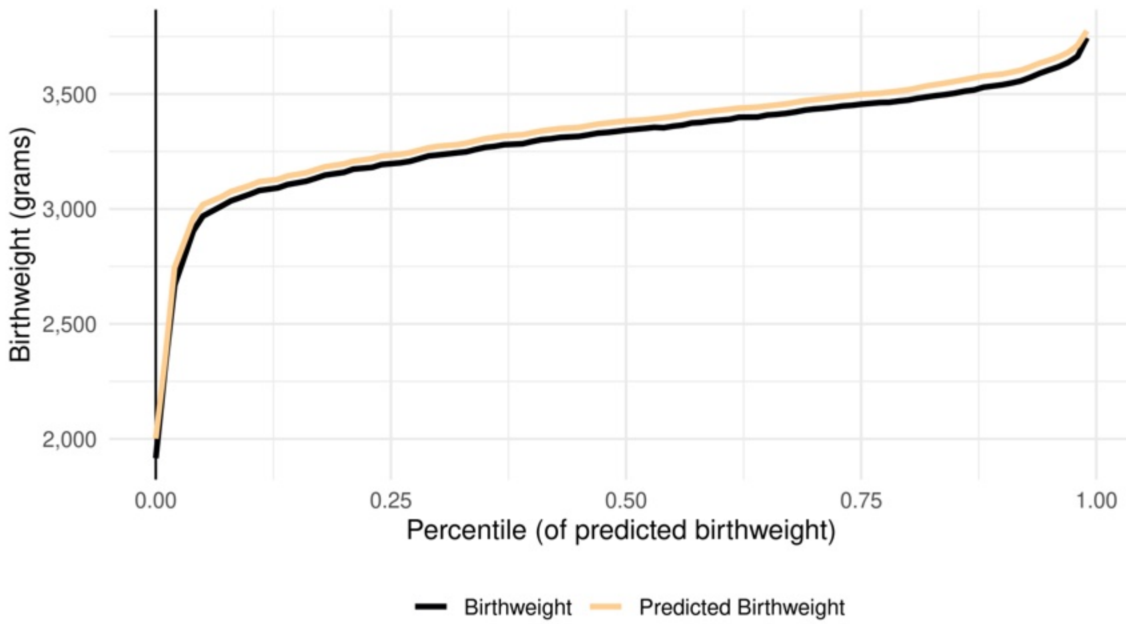


Figure A5: Predicted birthweights closely match actual birthweights across the predicted birthweight distribution. At each predicted birthweight percentile (x-axis), we take the average actual birthwight and average predicted birthweight, which are both plotted in the y-axis. Sample includes births to mothers with rural residences from 1990 to 2013.

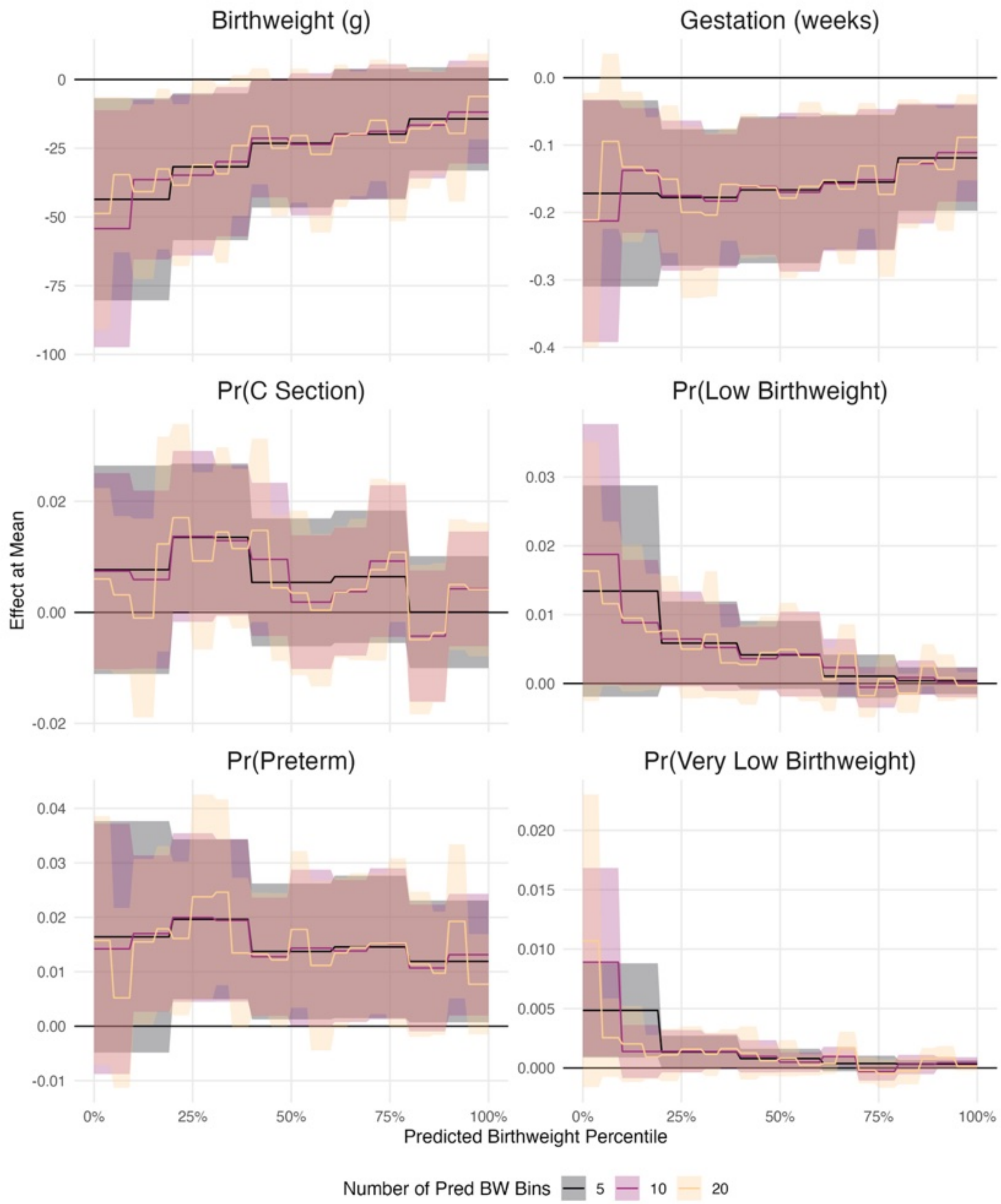
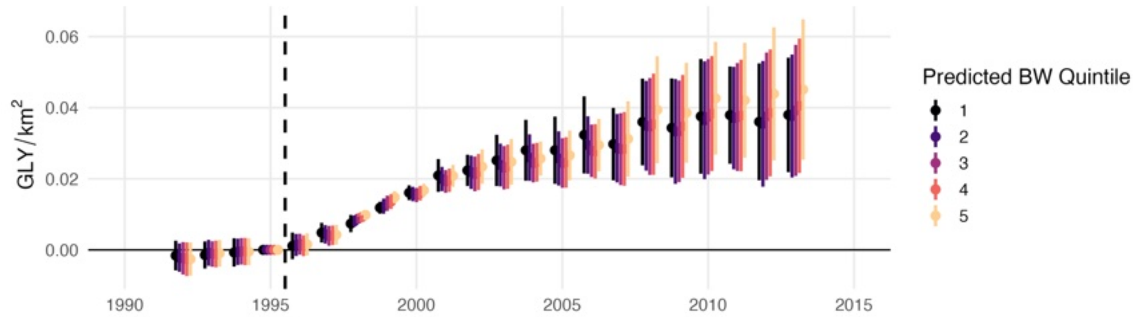
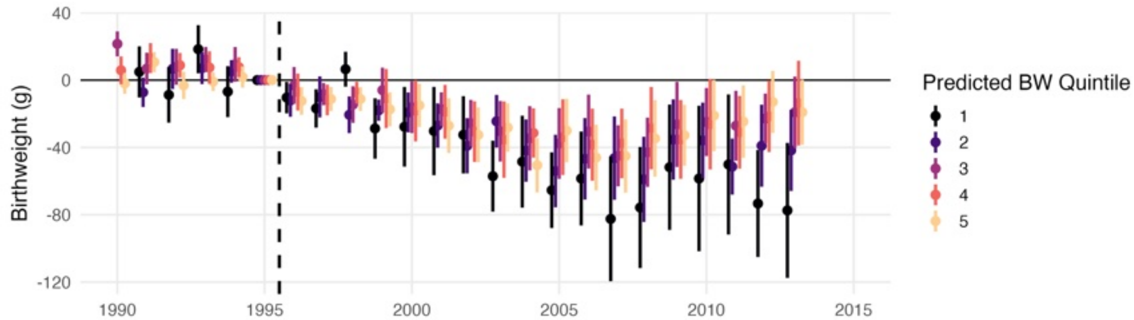


Figure A6: Heterogeneity in policy effect is consistent across various predicted birthweight bin sizes, greater disparities among birthweight outcomes. Estimated policy effect at mean of GLY/km^2 on various perinatal health outcomes instrumented with GM attainable yield interacted with year. All regressions include county, year by month, and family demographic fixed effects. Standard errors are clustered by state and year. Family demographics include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.



(a) *First-stage* effect of GM suitability on GLY intensity by predicted birthweight quintile



(b) *Reduced-form* effect of GM suitability on birthweight by predicted birthweight quintile

Figure A7: First-stage event study coefficients are similar across predicted BW quintiles, reduced form shows larger effects in lower quintiles. **(a)** Estimated event-study coefficients for the effect of local GM attainable yield percentile on GLY by year relative to 1995 by predicted birthweight quintile. Pesticide data only go back to 1992—there are no coefficients in 1990–1991. **(b)** Similar event study but with birthweight as outcome. Estimates from each predicted birthweight quintile come from separate regressions. All regressions include county, year by month, and family demographic fixed effects. Standard errors are clustered by state and year. Family demographics include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

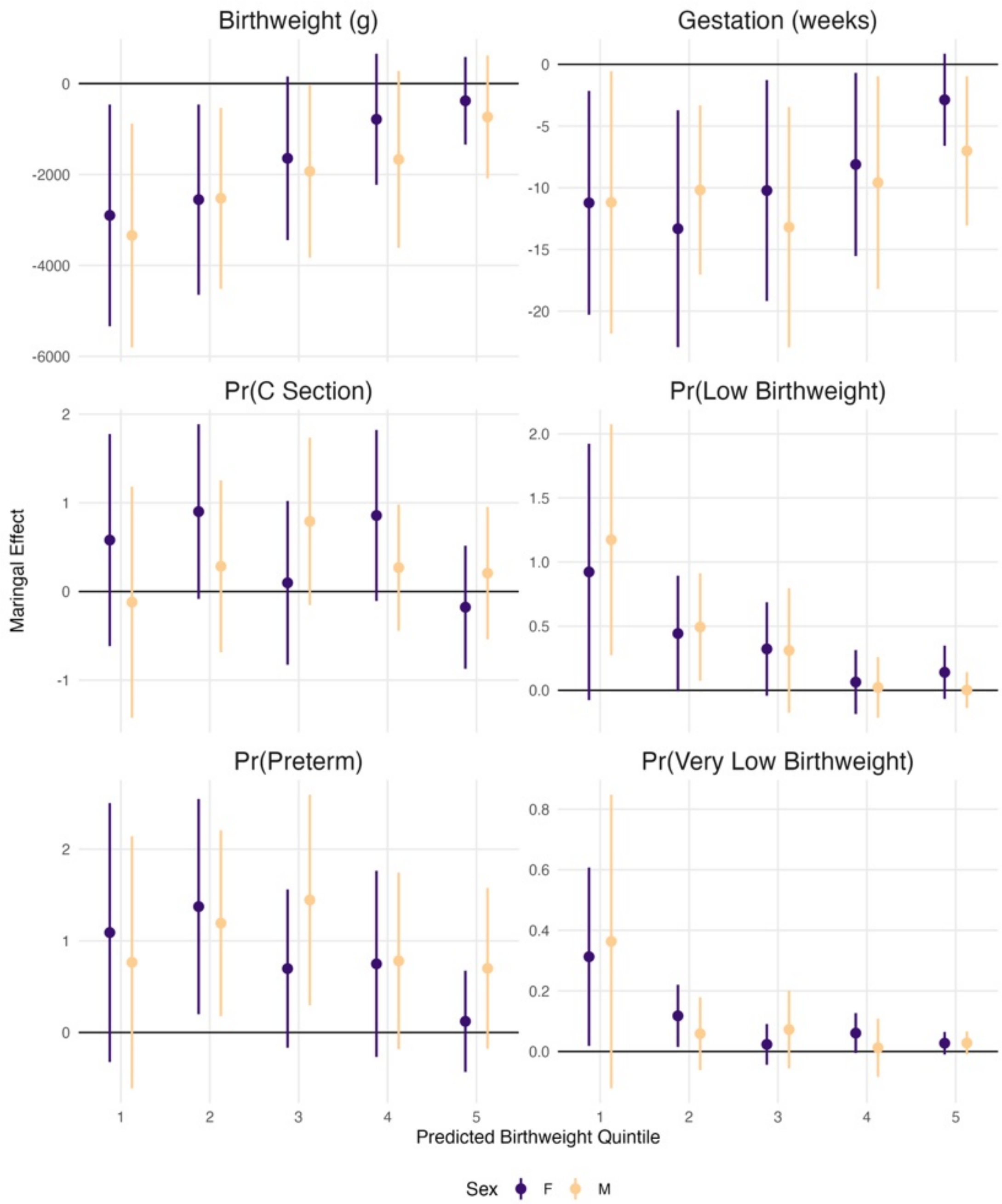


Figure A8: Limited evidence of heterogeneous marginal effects by sex within predicted BW quintile. Estimated marginal effect of GLY/km^2 on various perinatal health outcomes instrumented with GM attainable yield interacted with year. All regressions include county, year by month, and family demographic fixed effects and control for other pesticides and unemployment. Standard errors are clustered by state and year. Family demographics include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

Appendix C Supplementary Information

C.1 Background

Genetically modified crops Monsanto developed the first genetically modified crops, releasing GM soy, corn, and cotton in 1996 in the United States. These plants are resistant to GLY, allowing farmers to spray their fields with GLY to kill weeds but not harm their crops. The pairing of GM seeds with GLY provides a simple and effective method for controlling weeds—previously, farmers had to use different pesticides, each effective on a smaller subset of weeds at different points in the cultivation process. This herbicide portfolio was supplemented by mechanical tilling. GLY previously had to be used sparingly since it would also kill the crops themselves. Farmers adopted GM seeds rapidly in the United States. In 2000, just four years after their release, GM seeds constituted 54 percent of soy acres, 61 percent of cotton acres, and 25 percent of corn acres—however GM corn, soy, and cotton now make up over 90 percent of US acreage [59].

GLY and health GLY is a broad-spectrum herbicide discovered and commercialized by Monsanto in the 1970s. Its popularity grew over the next twenty years because of its relatively favorable properties. GLY has a low toxicity relative to other chemicals used on farms. It breaks down fairly quickly and binds to the soil, decreasing runoff [20]. However, it is water-soluble, which means that the part that does not bind to soil enters the water supply [60]. It is an effective weed killer, working on a broad spectrum of plants. However, GLY does not just kill weeds, it also kills fungi and microorganisms in the soil, which can lead to the crops being susceptible to disease [61]. It also breaks the nutrient cycle, forcing farmers to increase their dependence on fertilizer to feed their crops [62]. Farmers in the US spend nearly $8 billion on pesticides each year [63], applying GLY to 298 million acres of crops annually [64].

C.2 OLS Results

Panels A and B of Table A3 contain results for our main specifications, but estimated with OLS rather than 2SLS. We find precise null effects across all outcomes, demonstrating the importance of isolating exogenous varaiation in GLY using our instruments.

C.3 Shift-share specification

We can recast our identifying variation to be used similarly to that of a traditional "shift-share" specification, where the "shift" is national GLY use and the "share" is attainable yield in each county. Thus, the identifying variation is very similar to our main results—we

	BW	LBW	VLBW	Gestation	Preterm	C-section
Panel A: Policy effect						
GLY/km^2	22.4	−0.022	−0.003	−0.511	0.030	0.022
	(76.1)	(0.022)	(0.006)	(0.373)	(0.050)	(0.068)
Controls						
Pesticides	N	N	N	N	N	N
Unemployment	N	N	N	N	N	N
Panel B: GLY effect						
GLY/km^2	54.2	−0.018	−0.003	−0.355	0.024	−0.001
	(61.1)	(0.021)	(0.006)	(0.305)	(0.043)	(0.065)
Controls						
Pesticides	Y	Y	Y	Y	Y	Y
Unemployment	Y	Y	Y	Y	Y	Y
Fixed-effects *(Both panels)*						
Family Demog	Y	Y	Y	Y	Y	Y
County	Y	Y	Y	Y	Y	Y
Yr $\times$ Mo	Y	Y	Y	Y	Y	Y
Summaries *(Both panels)*						
N obs. (millions)	10.73	10.73	10.73	10.71	10.71	9.51

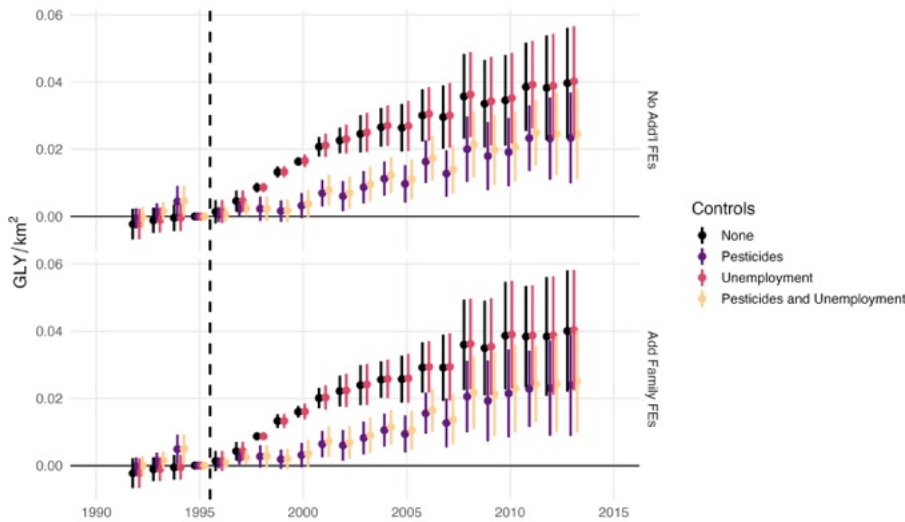
Table A3: OLS estimates of the policy and direct GLY effects on perinatal health. Each coefficient estimate (column-panel combination) provides results from a separate OLS regression. Both panels include family demographic, county, and year by month fixed effects. GLY effect (Panel B) additionally controls for other pesticides and unemployment. Sample restricted to births occurring in rural countries or from mothers residing in rural counties. Family demographic controls include mother's age, mother's race, mother's origin, mother's education, sex of child, total birth order, mother's residence status, and birth facility. Pesticide controls include alachlor, atrazine, cyanizine, fluazifop, metolachlor, metribuzin, and nicosulfuron. GLY/km^2 is kg/km^2. Standard errors in parentheses. We two-way cluster errors by year and state.

get temporal variation driven by the nation-wide increase in GLY use after the release of GM crops, and we get spatial variation from the suitability of the land in each county for corn, soy, and cotton. In the shift-share specification, our instruments are national GLY and national GLY interacted with the attainable yield percentile for corn, soy, and cotton. Thus, the difference between this specification and our main specification is that the first stage uses the national GLY trend directly, rather than interacting attainable yield with year dummies. When calculating the national GLY for each county, we exclude GLY sprayed within 100km of the county and any GLY sprayed upstream of the county to ensure that the national GLY instrument satisfies the exclusion restriction—that national GLY only affects perinatal health through it's affect on local GLY. Table A4 shows the results, which are generally similar, but smaller in magnitude than our main results.

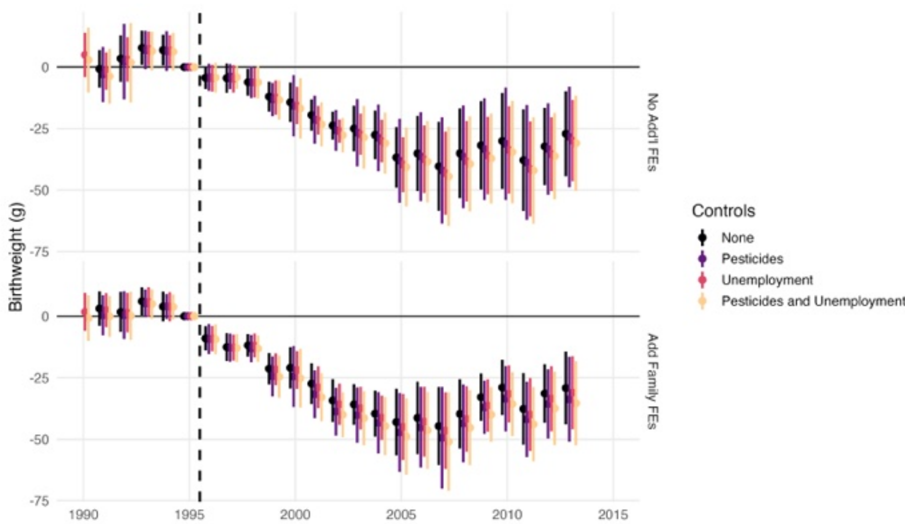
	BW	LBW	VLBW	Gestation	Preterm	C-section
Panel A: Policy effect						
GLY/km^2	-599.3	0.132	0.033	-4.11	0.440	0.166
	(213.0)	(0.054)	(0.011)	(1.01)	(0.153)	(0.149)
Controls						
Pesticides	N	N	N	N	N	N
Unemployment	N	N	N	N	N	N
Panel B: GLY effect						
GLY/km^2	-870.5	0.205	0.049	-5.49	0.622	0.130
	(305.7)	(0.072)	(0.017)	(1.47)	(0.224)	(0.187)
Controls						
Pesticides	Y	Y	Y	Y	Y	Y
Unemployment	Y	Y	Y	Y	Y	Y
Fixed-effects *(Both panels)*						
Family Demog	Y	Y	Y	Y	Y	Y
County	Y	Y	Y	Y	Y	Y
Yr $\times$ Mo	Y	Y	Y	Y	Y	Y
Summaries *(Both panels)*						
N obs. (millions)	10.73	10.73	10.73	10.71	10.71	9.51
Effects at mean						
Policy effect at mean	-14.0	0.003	0.0008	-0.096	0.010	0.004
GLY effect at mean	-20.3	0.005	0.001	-0.128	0.015	0.003

Table A4: Shift-share estimates of the policy and direct GLY effects on perinatal health. Each coefficient estimate (column-panel combination) provides results from a separate OLS regression. Both panels include family demographic, county, and year by month fixed effects. GLY effect (Panel B) additionally controls for other pesticides and unemployment. Sample restricted to births occurring in rural countries or from mothers residing in rural counties. Family demographic controls include mother's age, mother's race, mother's origin, mother's education, sex of child, total birth order, mother's residence status, and birth facility. Pesticide controls include alachlor, atrazine, cyanizine, fluazifop, metolachlor, metribuzin, and nicosulfuron. GLY/km^2 is kg/km^2. Standard errors in parentheses. We two-way cluster errors by year and state.

C.4 Robustness of first-stage and reduced form results

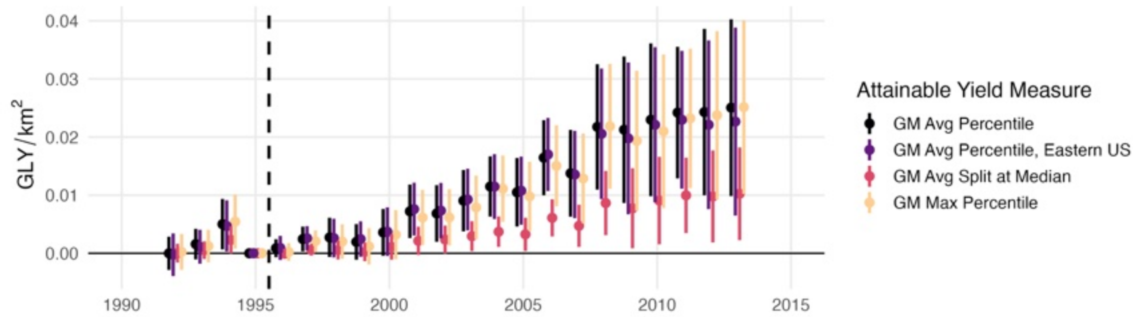


(a) First-stage effect of local GM attainable yield on GLY

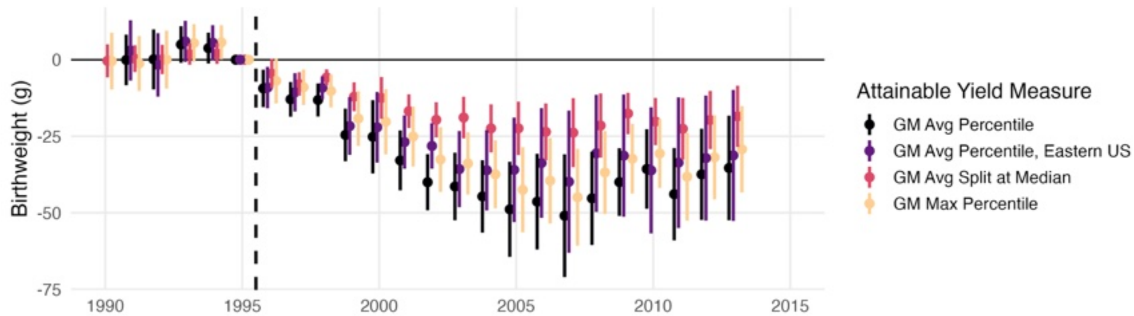


(b) Reduced-form effect of local GM attainable yield on birthweight

Figure A9: Robustness of birthweight effect to alternative controls and fixed effects. Estimated effect of local GM attainable yield percentile on birthweights relative to 1995. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. Family demographics include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

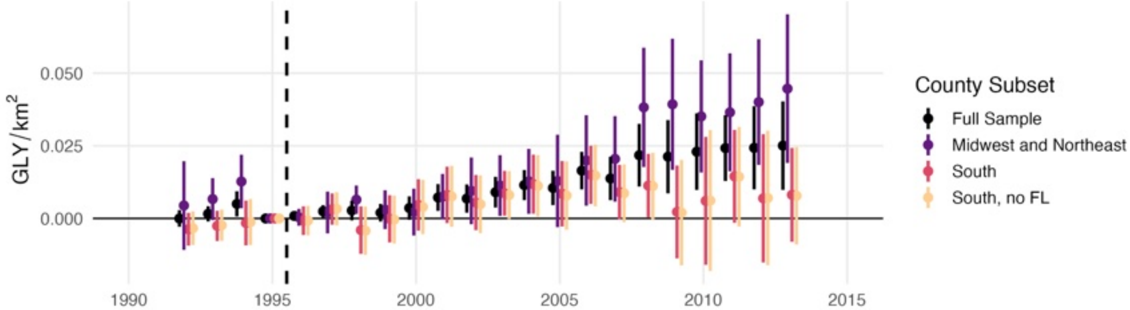


(a) First-stage effect of various instruments on GLY/km^2

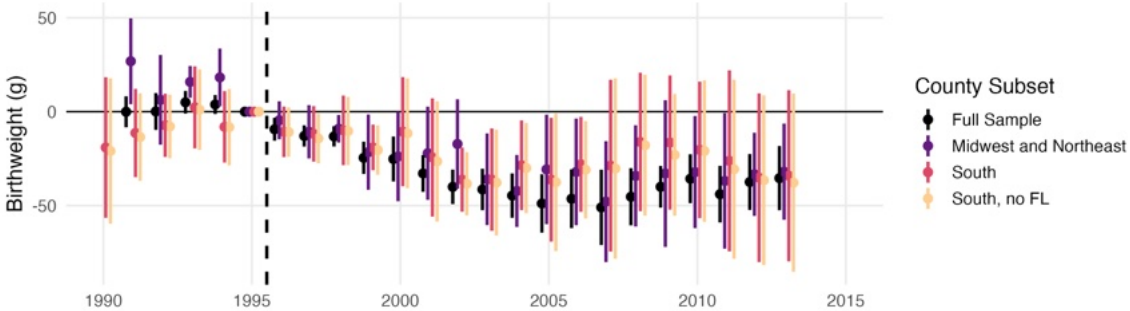


(b) Reduced-form effect of various instruments on birthweight

Figure A10: Robustness of birthweight effect to alternative instruments. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. We vary the construction of GM attainable yield: "GM Avg Percentile" is our main specification, "GM Average Percentile, Eastern US" limits the sample to just counties east of the 100th meridian, "GM Average, Split at Median" uses a binary high vs low attainable yield, where a county is high attainable yield if they are above the median attainable yield, and "GM Max Percentile" takes the maximum standardized attainable yield among corn, soy, and cotton (rather than the average) before re-scaling into a percentile. The regressions control for unemployment and family demographics, including mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.



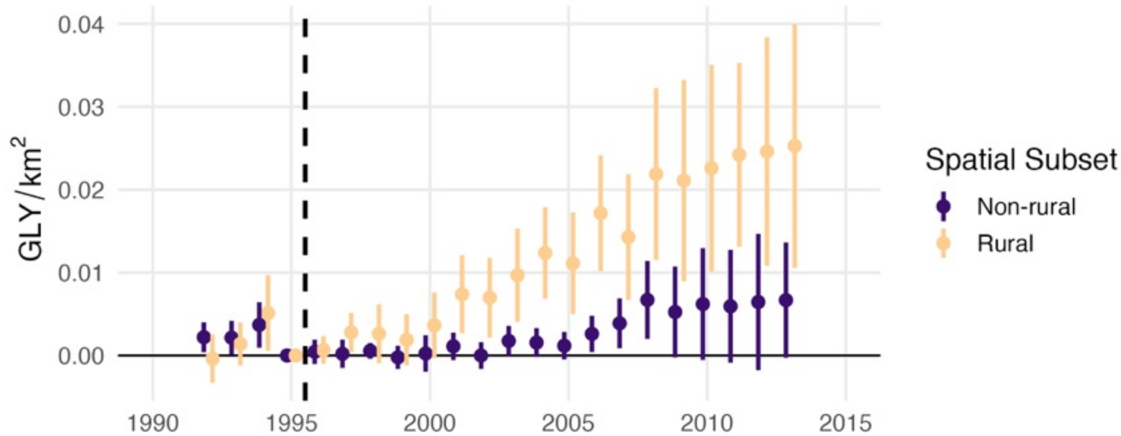
(a) First-stage effect on GLY/km^2



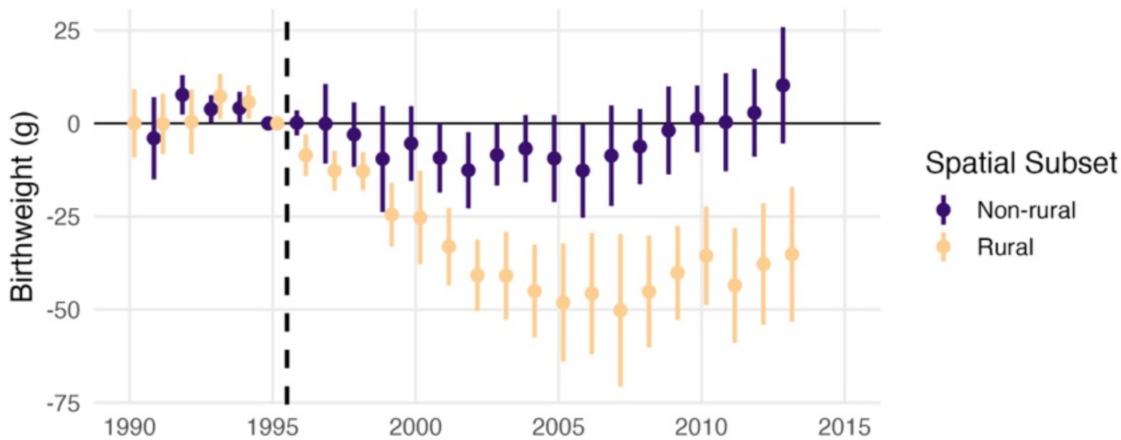
(b) Reduced-form effect on birthweight

Figure A11: Heterogeneity in Birthweight Effect by Geographic Subsets. Estimated effect of local GM attainable yield percentile on birthweight relative to 1995. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. The geographic subsets are primarily defined using census regions (Midwest, Northeast, or South). Fig A10 shows results with just the eastern US. All regressions also control for unemployment and family demographics, including mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

Figure A12 estimates our model for births to mothers with rural and non-rural residences separately. There is a small but largely insignificant decrease in birthweight after the release of GM seeds in 1996 in high GM attainable yield counties relative to low GM attainable yield counties. However, this effect is gone by 2010. We attribute this difference largely to measurement error in exposure—we do not think that exposure is very high for urban mothers, who are unlikely to be in contact with drift, dust, or water contaminated with GLY sprayed within that county. The lack of direct measurement of exposure to GLY is a weakness of our study, as all we know is the amount of GLY used in a county each year.



(a) First-stage effect on GLY/km^2, rural and non-rural counties.



(b) Reduced-form effect on birthweight, rural and non-rural counties.

Figure A12: Birthweight event studies by rural and non-rural counties. Estimated effect of local GM attainable yield percentile on birthweight relative to 1995 for births to mothers residing and occurring in rural and non-rural counties. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. All regressions also control for other pesticides, unemployment, and family demographics, including mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race.

C.5 Robustness of 2SLS results for other outcomes

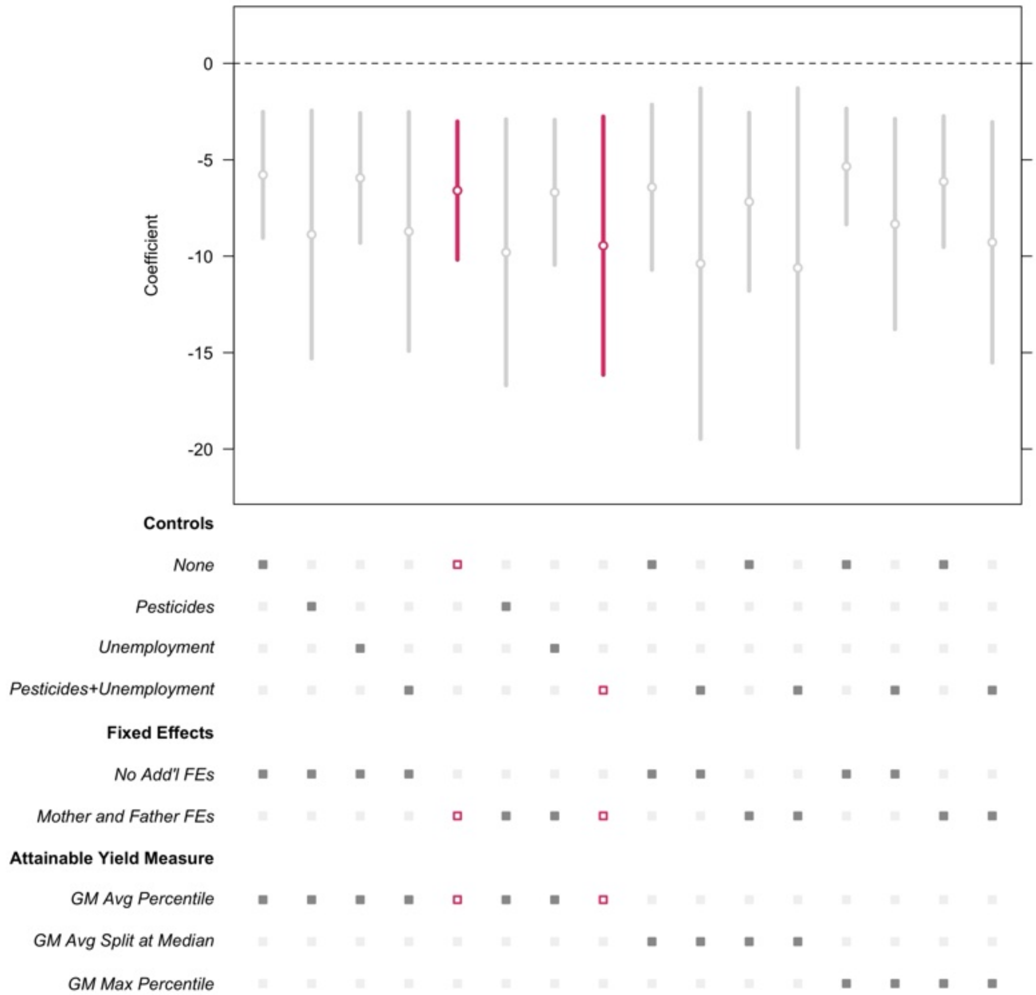


Figure A13: Robustness of gestation length effect to alternative specifications. Coefficients are the estimated effect of GLY (kg/km^2) on gestation length. Our main specifications are highlighted. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. Pesticide controls include Mother and Father FE's include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. We vary the construction of GM attainable yield: "GM Avg Percentile" is our main specification, "GM Average, Split at Median" uses a binary high vs low attainable yield, where a county is high attainable yield if they are above the median attainable yield, and "GM Max Percentile" takes the maximum standardized attainable yield among corn, soy, and cotton (rather than the average) before re-scaling into a percentile. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

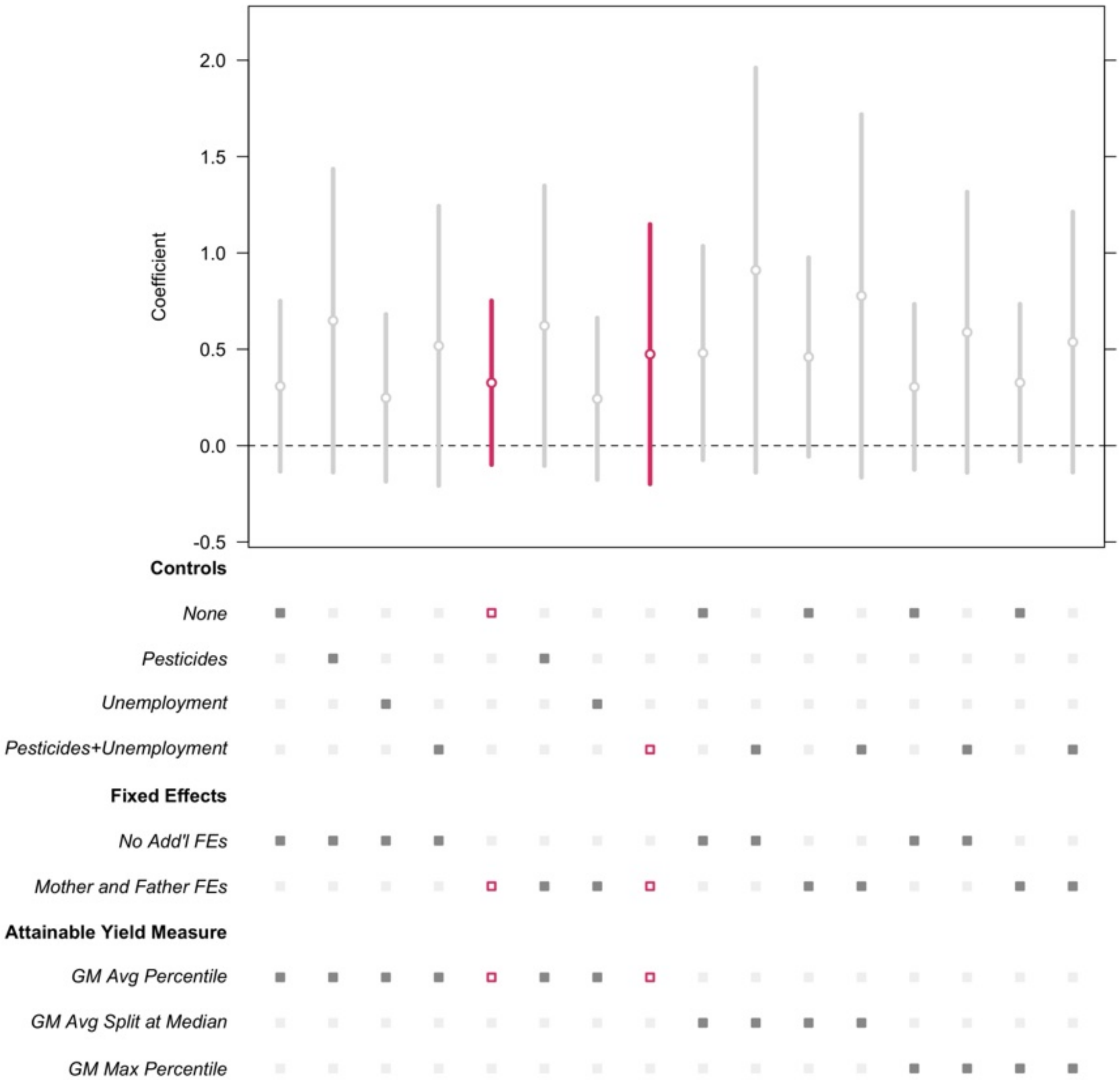


Figure A14: Robustness of C-section effect to alternative specifications. Coefficients are the estimated effect of GLY (kg/km^2) on the probability of having a C-section. Our main specifications are highlighted. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. Pesticide controls include Mother and Father FE's include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. We vary the construction of GM attainable yield: "GM Avg Percentile" is our main specification, "GM Average, Split at Median" uses a binary high vs low attainable yield, where a county is high attainable yield if they are above the median attainable yield, and "GM Max Percentile" takes the maximum standardized attainable yield among corn, soy, and cotton (rather than the average) before re-scaling into a percentile. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

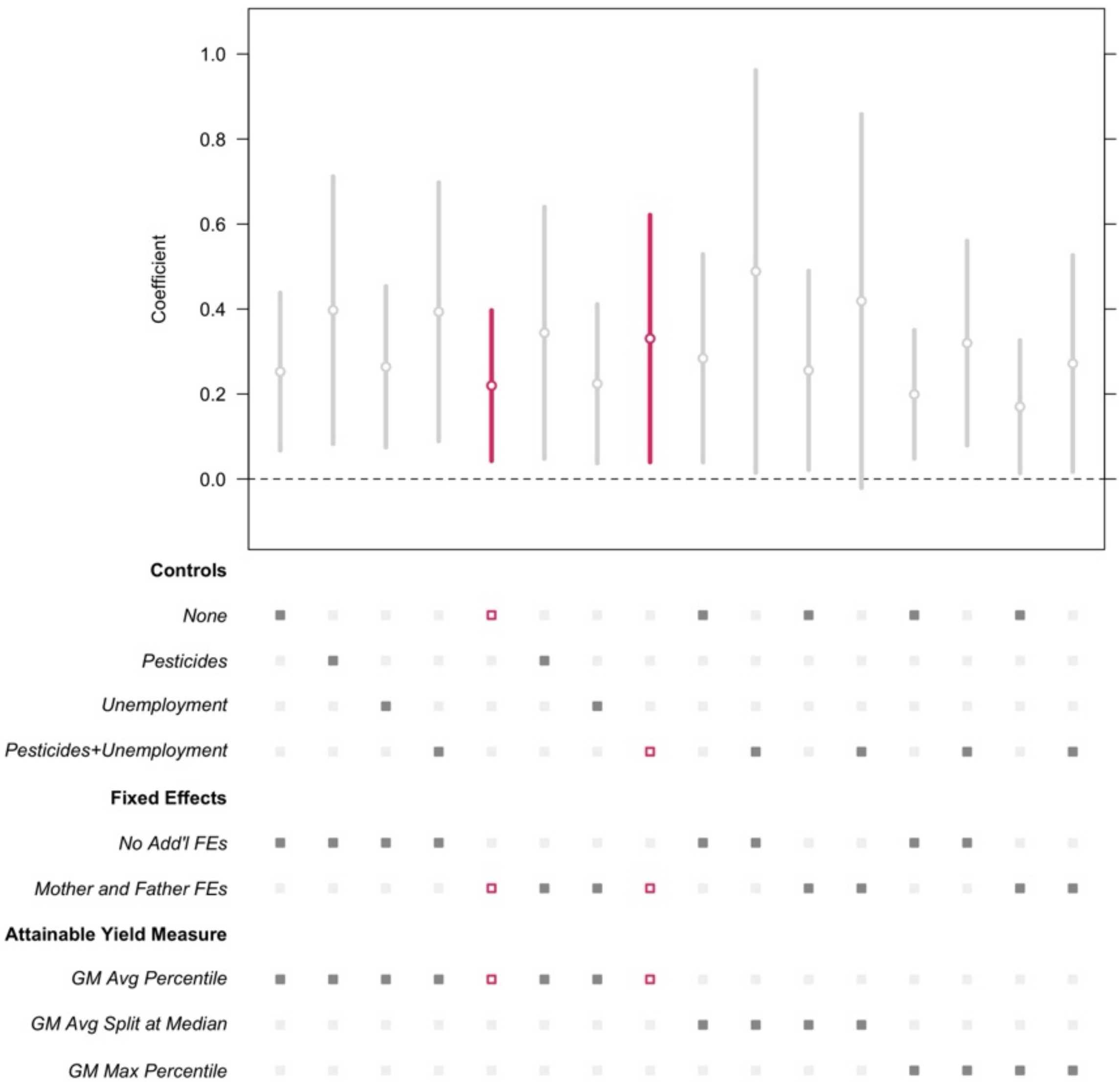


Figure A15: Robustness of LBW effect to alternative specifications. Coefficients are the estimated effect of GLY (kg/km^2) on the probability of low birthweight. Our main specifications are highlighted. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. Pesticide controls include Mother and Father FE's include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. We vary the construction of GM attainable yield: "GM Avg Percentile" is our main specification, "GM Average, Split at Median" uses a binary high vs low attainable yield, where a county is high attainable yield if they are above the median attainable yield, and "GM Max Percentile" takes the maximum standardized attainable yield among corn, soy, and cotton (rather than the average) before re-scaling into a percentile. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

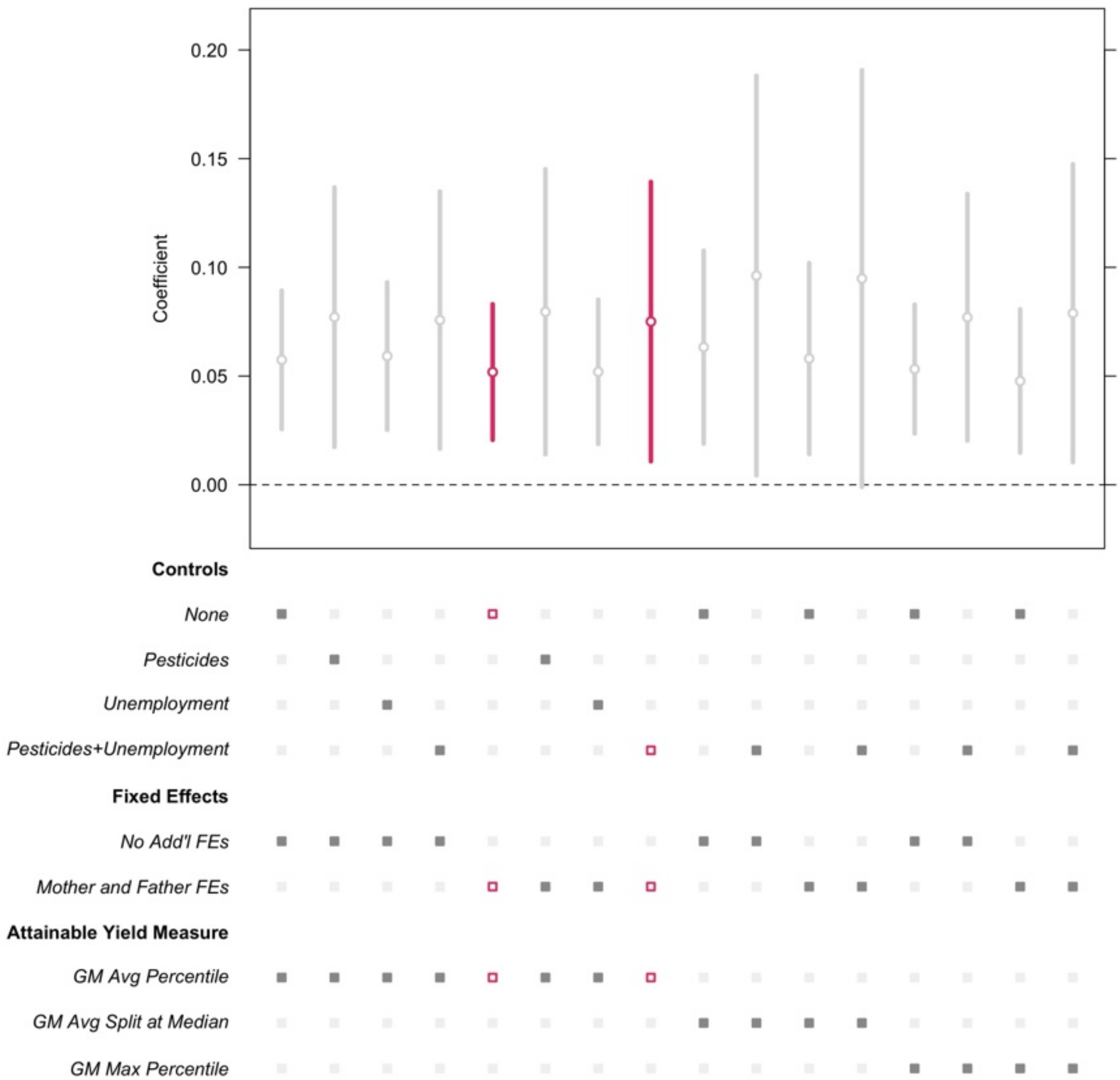


Figure A16: Robustness of VLBW effect to alternative specifications. Coefficients are the estimated effect of GLY (kg/km^2) on the probability of very low birthweight. Our main specifications are highlighted. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. Pesticide controls include Mother and Father FE's include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. We vary the construction of GM attainable yield: "GM Avg Percentile" is our main specification, "GM Average, Split at Median" uses a binary high vs low attainable yield, where a county is high attainable yield if they are above the median attainable yield, and "GM Max Percentile" takes the maximum standardized attainable yield among corn, soy, and cotton (rather than the average) before re-scaling into a percentile. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

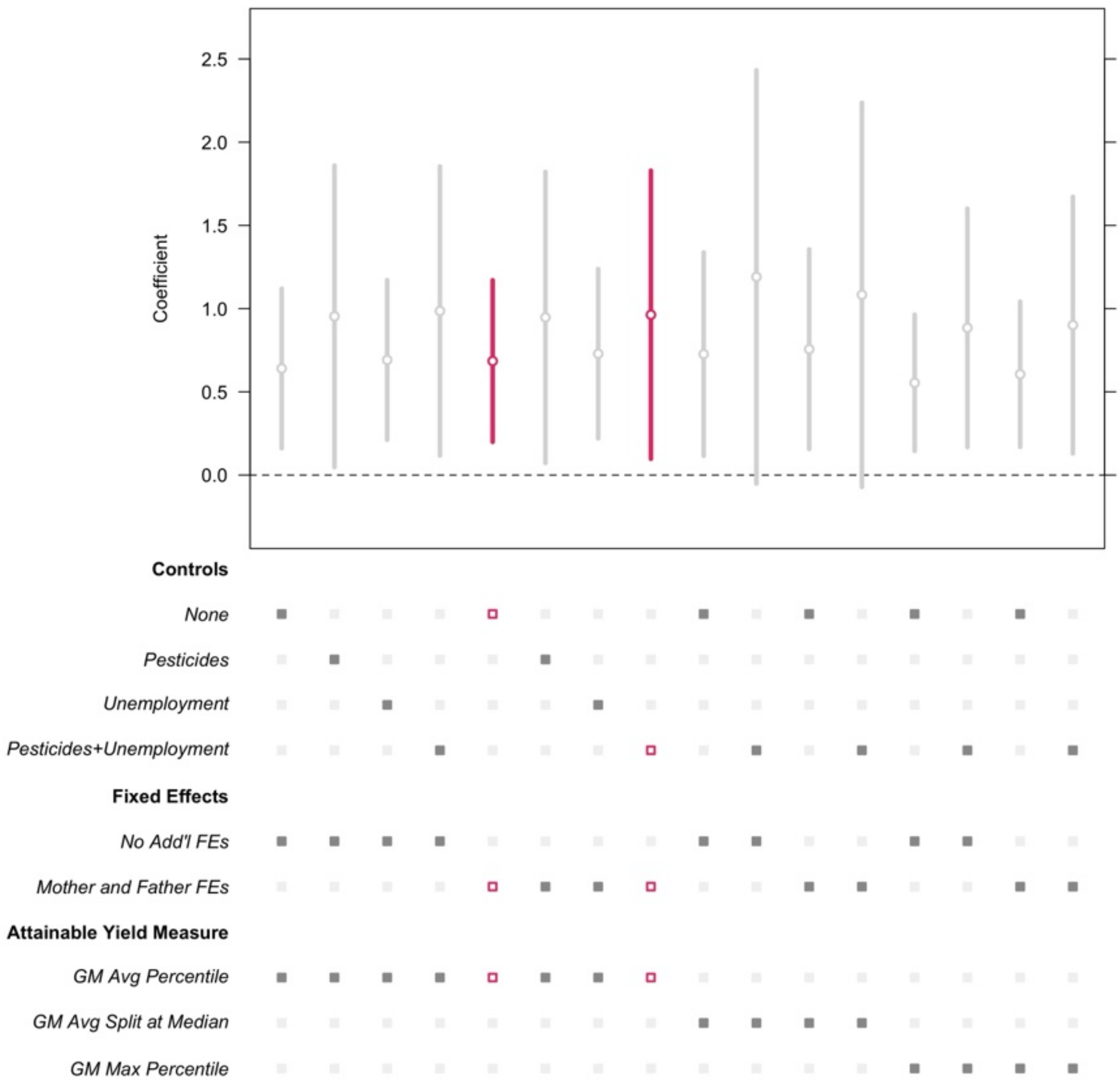


Figure A17: Robustness of preterm effect to alternative specifications. Coefficients are the estimated effect of GLY (kg/km^2) on the probability of preterm birth. Our main specifications are highlighted. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. Pesticide controls include Mother and Father FE's include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. We vary the construction of GM attainable yield: "GM Avg Percentile" is our main specification, "GM Average, Split at Median" uses a binary high vs low attainable yield, where a county is high attainable yield if they are above the median attainable yield, and "GM Max Percentile" takes the maximum standardized attainable yield among corn, soy, and cotton (rather than the average) before re-scaling into a percentile. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

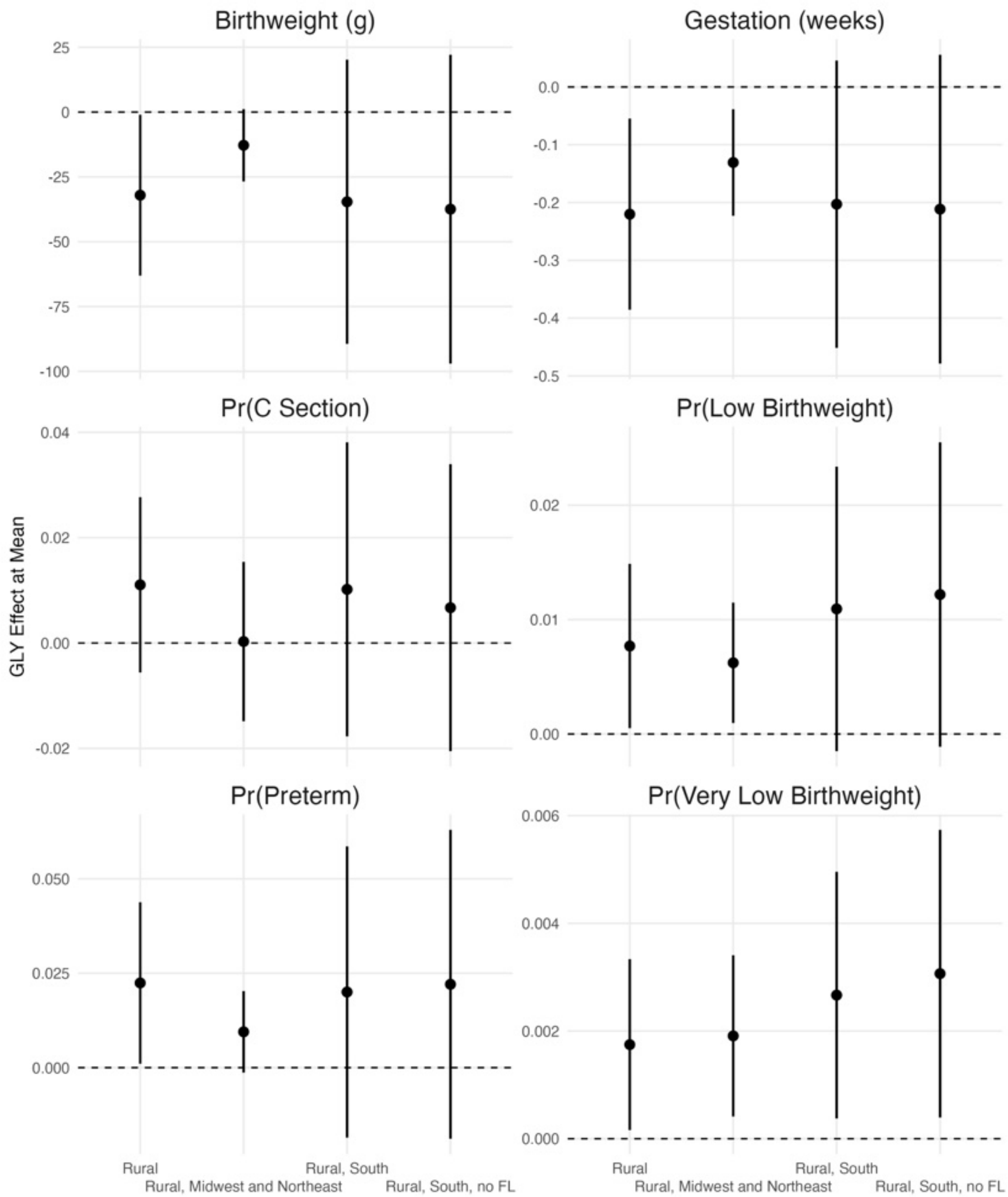


Figure A18: Robustness to spatial subsets, all outcomes. Estimated effect of GLY/km^2 on various perinatal health outcomes instrumented with GM attainable yield interacted with year. All regressions include county, year by month, and family demographic fixed effects and standard errors are clustered by state and year. Family demographics include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race.

C.6 Demographic Trends

One concern for identification in our model is that the underlying composition of the population is changing in high vs. low GM attainable yield counties during the period of our study. Figure A19 shows event studies where we use demographics of the mother as outcomes with county and year-by-month fixed effects and no other controls to test whether demographics are changing over time. We find that births in high-yield counties are less likely to come from black mothers after the release of GM crops—this would otherwise be concerning for our main estimates, however, we (1) control for race and other demographics in our main estimation and they do not meaningfully impact the results, (2) *predicted* birthweight does not change over the time period of the study, and (3) we find significant effects of GLY on birthweight for babies with both white and non-white mothers.

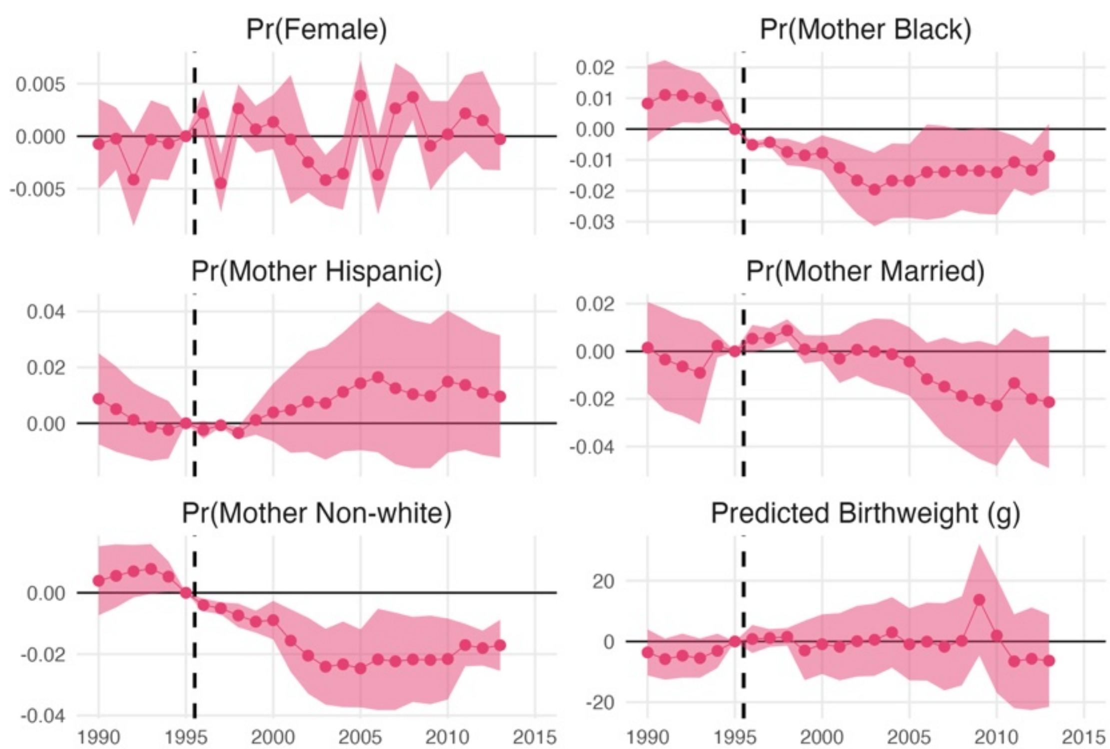


Figure A19: Demographic event studies. Estimated effect of local GM attainable yield percentile on various demographics outcomes relative to 1995. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

C.7 Other forms of heterogeneity

Mother's Race Based on heterogeneity in predicted birthweight, we expect there to be differences in effect by mother's race. Fig A20 shows reduced from event studies for different outcomes by mother's race. Births to non-white mothers have a noisy, but generally larger effect than briths to white mothers.

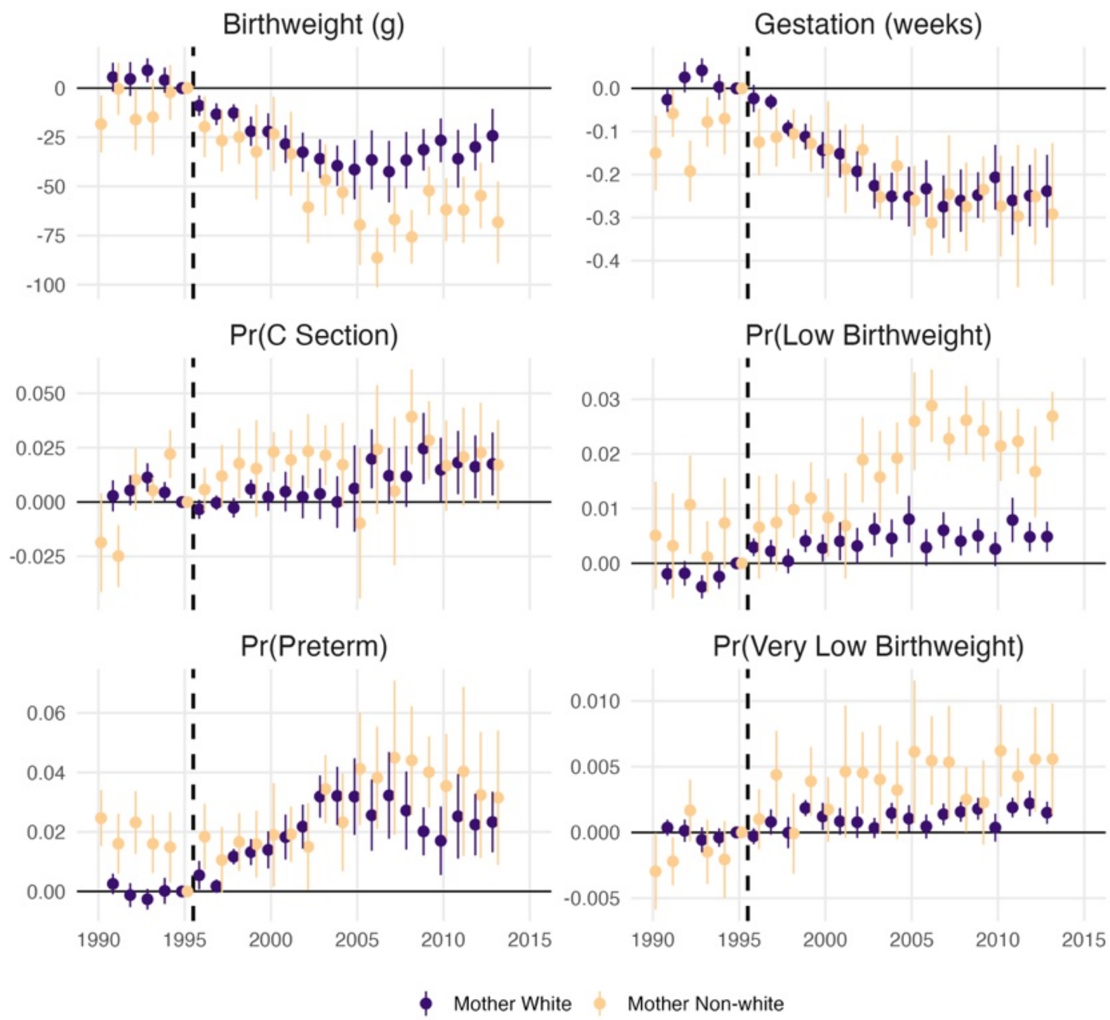


Figure A20: Reduced form heterogeneity by mother's race. Estimated effect of local GM attainable yield percentile on various perinatal health outcomes relative to 1995. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. All regressions also control for family demographics, including mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births occurring in rural counties or to mothers residing in rural counties.

Heterogeneity by month of birth These results do not exhibit consistent heterogeneity by month of birth, as seen in Figure A21. There are slightly higher effects during the first months of the year—which means that their gestational period began in the spring and early summer the time when the most GLY is applied.

C.8 Effect of GM on Acreage and Yield

Changes in agricultural activity unrelated to GLY that result from GM seed adoption could also affect infant health, threatening our identified effect of GLY on birth weight. For example, GM technology could lead farmers to bring marginal, not previously farmed land into agricultural production. This additional production could be associated with increased runoff into water or air pollution from dust or drift. The previously unfarmed land may have provided ecosystem services protecting infant health. Additionally, if yield increased with GM seeds, the local economy could see a boost from higher revenues. In order to rule out these as mechanisms for the observed effect of GM attainable yield on birth weight, we explore the effect of GM attainable yield on crop acreage and actual yield.

We use USDA NASS data on crop acreage and yield. Since the USDA masks counties with few farms in the raw data, we aggregate up to the Agricultural Statistics District level. We estimate the same event study models as in the main analysis but use crop acreage or yield as the outcome and area-weighted attainable yield as the treatment variable. We continue to weight by total infant births. However, these plots reflect all counties, not just rural counties, as in the main analysis, since we cannot distinguish between rural and non-rural counties included in the masked district-level data.

Figure A22 shows the event study estimation for the effect of GM attainable yield on corn, soy, cotton, and total acres planted. Soy acreage increased slightly in high-suitability districts relative to low-suitability districts in the first few years after 1995 but quickly returned to pre-1995 levels. Meanwhile, corn acreage remained relatively constant between high- and low-suitability districts until 2007, when the renewable fuel standard increased incentives for farmers to plant corn [65]. Total crop acreage did not change in high relative to low attainable yield districts, but cotton saw a small, consistent, but not statistically significant increase. We emphasize that the magnitudes of these changes are very small—the peak soy acreage increase in 1998 represents just a 0.14 standard deviation increase in soy acreage. Figure A23 shows yield for corn, soy, and cotton—revealing no clear trend in the difference between high and low attainable yield counties after the introduction of GM seeds.

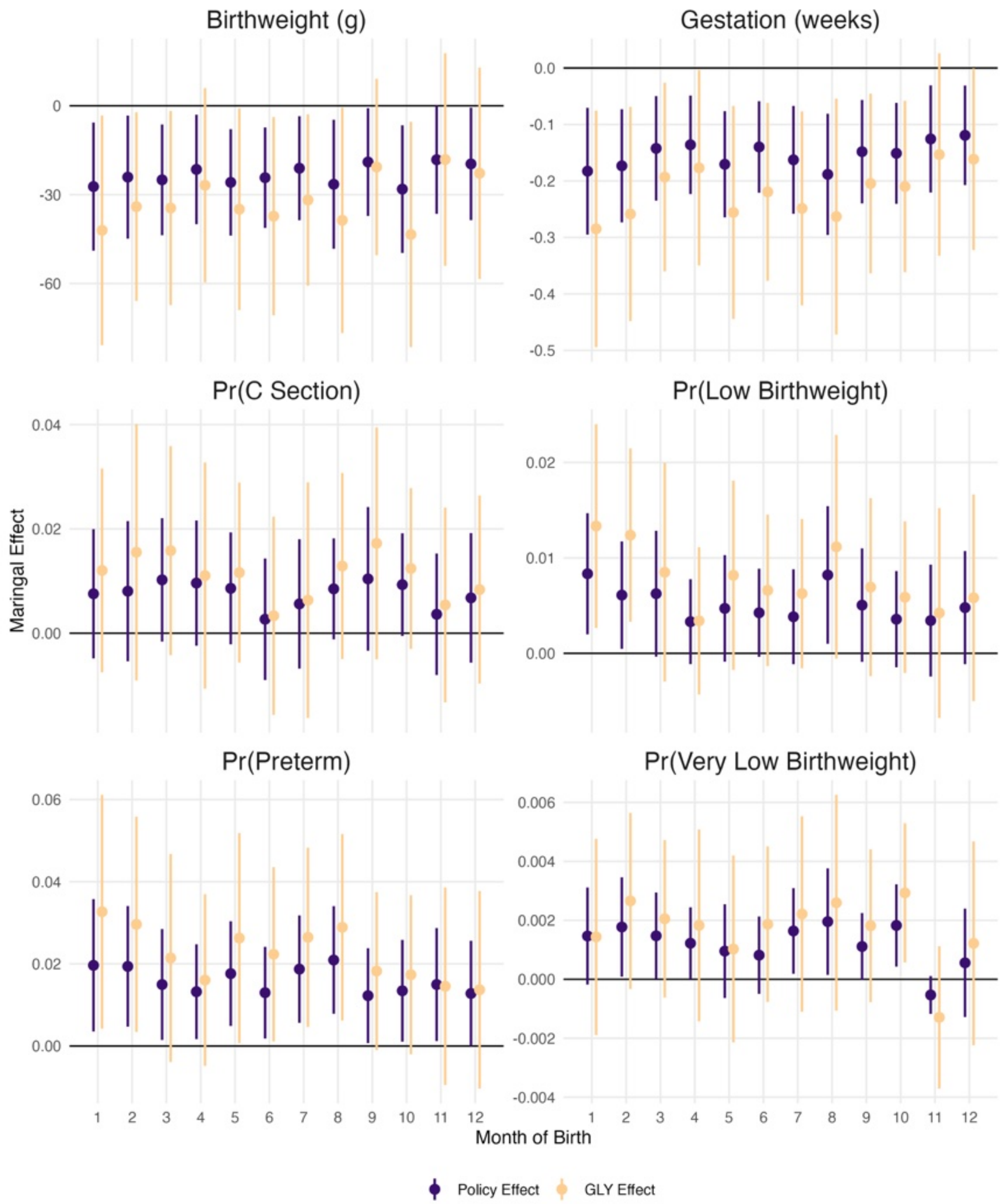
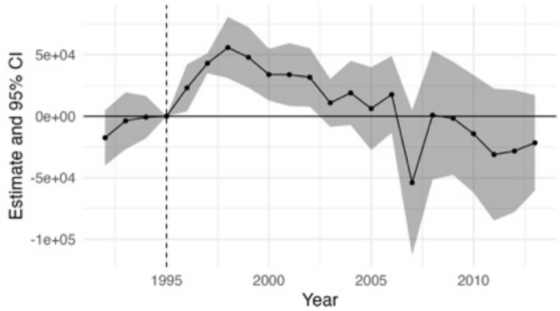
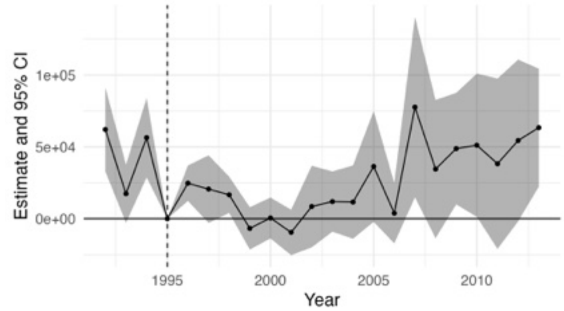


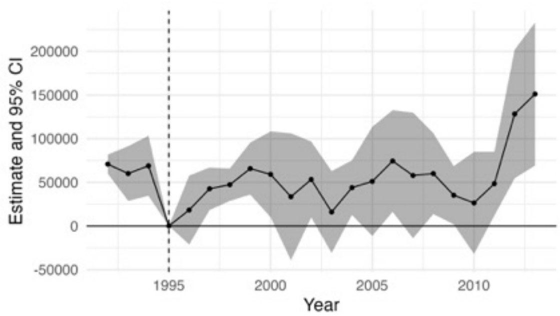
Figure A21: Heterogeneity in effect by different month of birth, all outcomes. Estimated effect of GLY/km^2 on various perinatal health outcomes instrumented with GM attainable yield interacted with year. All regressions include county, year, and month fixed effects and standard errors are clustered by state and year. Family demographics include mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race.



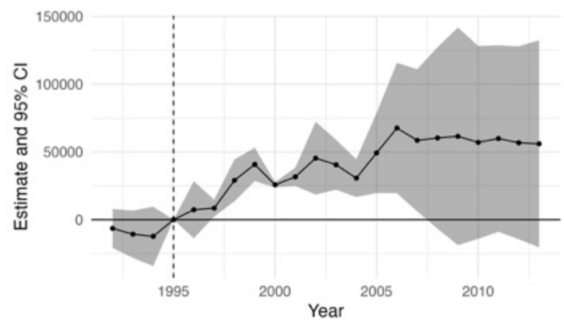
(a) Soy acreage



(b) Corn acreage

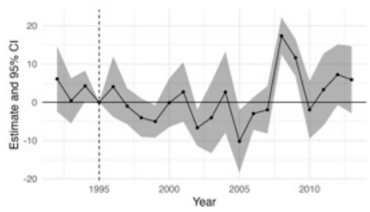


(c) Total crop acreage

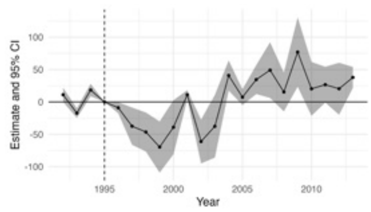


(d) Cotton acreage

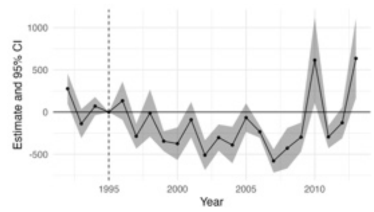
Figure A22: Effect of local GM attainable yield on crop acreage. Standard errors are clustered by state and year.



(a) Soy yield

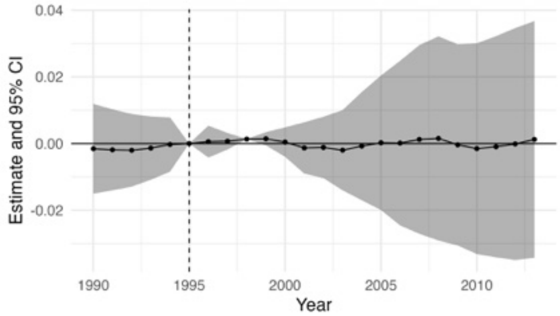


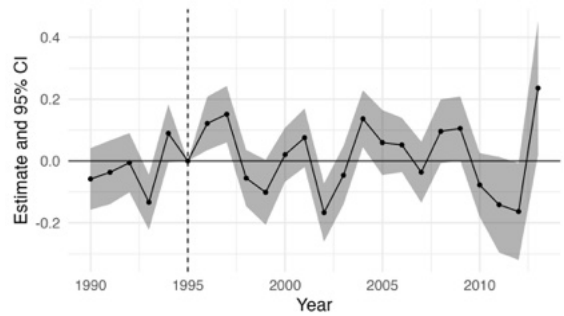
(b) Corn yield



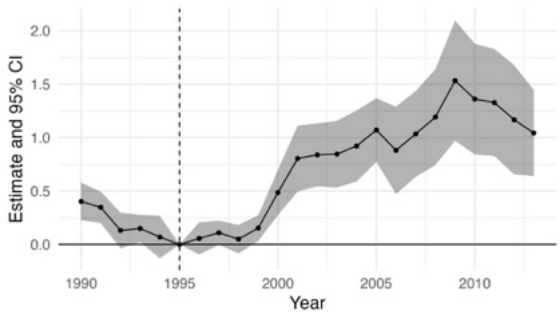
(c) Cotton yield

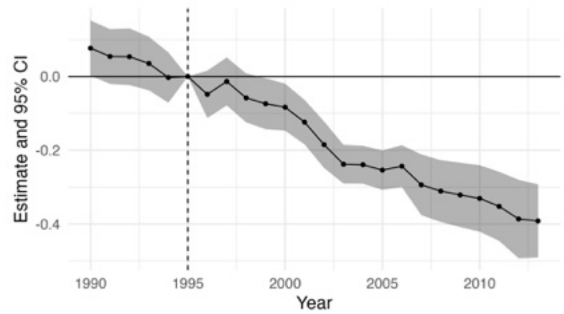
Figure A23: Effect of local GM attainable yield on crop yield. Soy and corn yield is measured in bushels/acre, while cotton is measured in lbs/acre. Standard errors are clustered by state and year





(a) Effect of GM suitability on non-farm income. **(b)** Effect of GM suitability on farm income





(c) Effect of GM suitability on unemployment rate **(d)** Effect of GM suitability on farm employment

Figure A24: Coefficients from an event study regression of various socioeconomic variables on GM suitability for rural counties. Standard errors are clustered by state and year.

C.9 Other socioeconomic outcomes

Here, we explore the relationship between our attainable yield instrument and some socioeconomic outcomes in order to rule them out as mechanisms for the measured birth weight effect. We regress the socioeconomic variables on GM attainable yield interacted with year dummies with county and year fixed effects. The sample is a county-year panel of rural counties in the US between 1990 and 2013. Figure A24 shows the results. There is no change in farm or non-farm income, however there do appear to be changes in employment. The unemployment rate jumps after 2000—thus, we control for unemployment in our main regression, but note that this is four years after the release of GM seeds, thus the timing does not align to have been caused by GM. Meanwhile, farm employment is also declining, however there is a clear pre-trend. The release of GM seeds does not appear to affect this trend.

C.10 Effects of Upstream GLY in Water

C.10.1 Predicting GLY in Water with Machine Learning

To measure spillover effects from GLY sprayed upstream, we must have some measure of GLY exposure in water. Ideally, this would come from extensive monitoring, which consistently reports pesticide concentrations in water for a comprehensive set of water sources. Unfortunately, such a monitoring network does not exist, so we must create an alternative methodology to estimate GLY exposure from upstream spraying. We train a machine learning model to predict GLY concentrations using the limited GLY monitoring in water, along with water flow and other environmental characteristics.

Data preparation Our training data come from Medalie et al. [14], who took 3204 samples of GLY and its main degradate AMPA from 70 sites in the National Water Quality Network (NWQN), a nationally representative set of water bodies, between 2015 and 2017. Both chemicals are nearly omnipresent, with GLY detected in 75 percent of samples and AMPA detected in 90 percent. We link these measurements to data on GLY use, soil type, slope, and rainfall upstream from the sampling location.

We use a spatial water model to aggregate the amount of GLY sprayed upstream and downstream of each sampling location. Specifically, we use the level 8 HydroBASINS product from HydroSHEDS [55]. These data are watershed polygons that delineate water basins across the globe in a standardized way. Importantly, they are assigned codes in a way that makes it possible to find all watersheds upstream and downstream from any given watershed.

We begin with the pesticide data. As in our local analysis, one may be concerned with the endogeneity of GLY use. Our estimates will be biased if spraying upstream of a sampling location correlates with other factors affecting health outcomes. We deal with this issue by using only exogenous variation in GLY use driven by the same instruments from our local analysis, namely that driven by the timing of the release of GM seeds and the suitability of a county for corn, soy, and cotton. We regress GLY on the GM attainable yield percentile interacted with year dummies, with year and county fixed effects to generate county-year level predictions of GLY. To disaggregate these county-level predictions into watersheds, we assume that spraying is uniform across the county and multiply the GLY prediction for each county by the portion of the county's total area covered by the watershed. Figure A25 shows the spatial distribution of predicted GLY by watershed across the United States in 2004.

Additionally, we collect several other variables that affect the runoff of GLY in a method loosely following the commonly used universal soil loss equation (USLE). This soil loss equa-

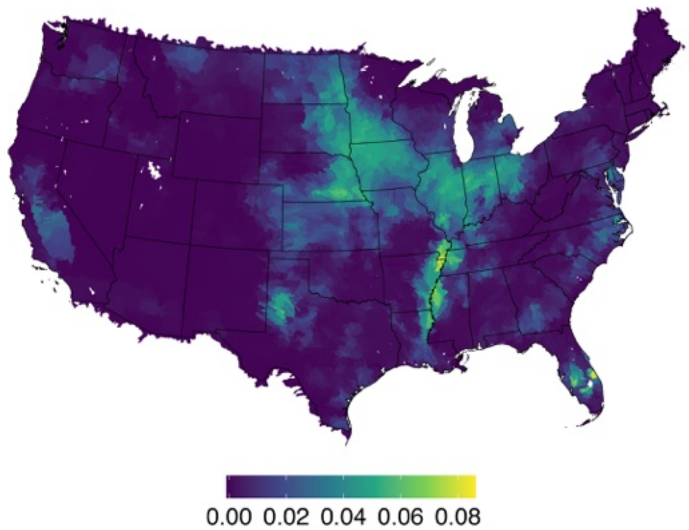
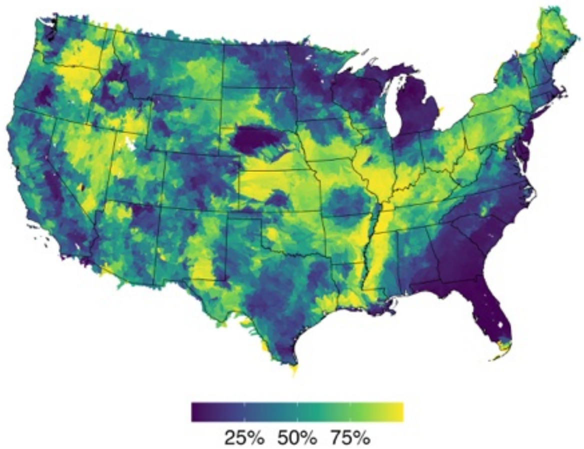


Figure A25: Predicted GLY disaggregated into watersheds in 2004. These predictions come from our first stage model regressing GLY on local GM attainable yield percentile with county and year fixed effects. We disaggregate from county into watersheds using the portion of the county's area that is covered by each watershed. We generate predictions for each year, but only show 2004 to accompany the exposition.

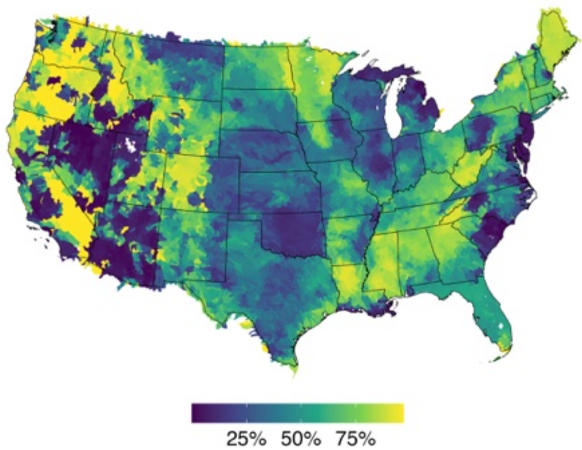
tion multiplies the erodibility of the soil, the slope of the land, rainfall, and two measures associated with land use. We aggregate soil erodibility and slope from the gridded soil survey to the watershed level by taking the average over all 30-meter cells in each watershed [56]. Similarly, we use gridded, monthly precipitation from PRISM to help inform the potential for GLY to run into water [57]. We aggregate the 4-kilometer cells to the watershed level by taking the simple average of cells within a watershed. Additionally, we aggregate to the annual level by taking the sum over the growing season, April through September, when most GLY is applied. Figure A26 shows national percentiles of soil erodibility, slope, and precipitation by watershed.

We then utilize the "Pfafstetter" watershed coding system used by the HydroBASINS data to find all watersheds upstream from each watershed. We have selected an example watershed in Washington County, Illinois, just east of St. Louis, for demonstration purposes. Figure A27 shows the example watershed in red and then highlights all of the watersheds upstream, which reach further north into Illinois, and all of the watersheds downstream, which follow the Mississippi River to the Gulf of Mexico.

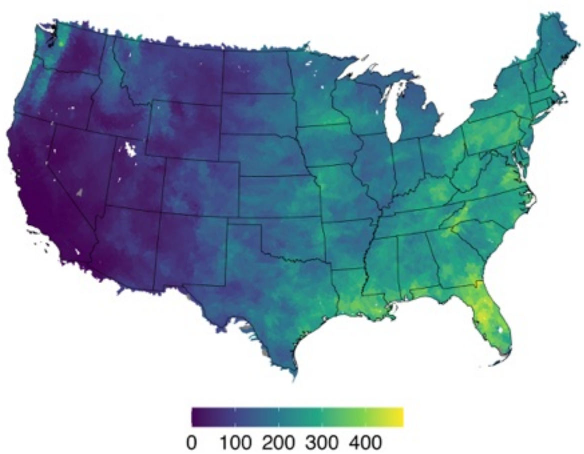
When linking upstream and downstream watersheds, we calculate the distance between any two watersheds by summing the distance between centroids of each watershed that lies along the water flow between the two watersheds. We then aggregate the variables described above into 50-kilometer distance bins from -100 to 350, where negative values



(a) Soil erodibility (K factor) percentile.



(b) Slope percentile.



(c) Growing season (Apr to Sep) precipitation in 2004.

Figure A26: Spatial variation in water ML predictors. Each map depicts a watershed-level average of the given variable. See text for details.

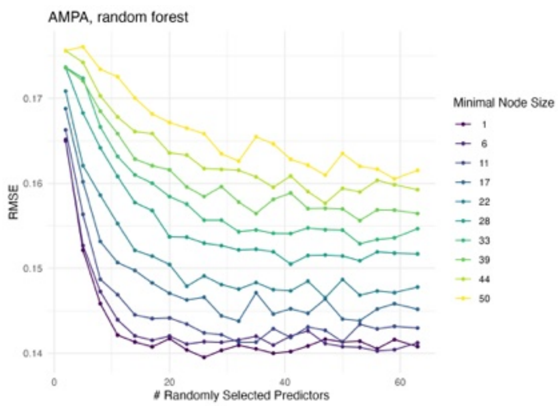
(a) Upstream and downstream watersheds.

(b) Distance bins.

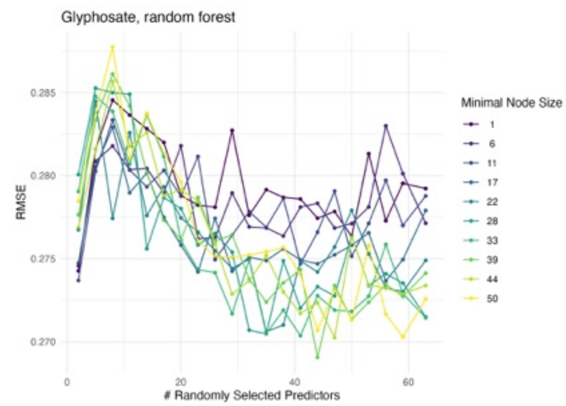
Figure A27: Capturing upstream and downstream watersheds. For an example watershed in Illinois (highlighted in red), we show all of the watersheds upstream and all of the watersheds downstream. We calculate distance upstream and downstream using the distance between the centroids of watersheds along the path, then categorize these into 50-kilometer distance bins.

denote values for downstream watersheds. Figure A27 demonstrates the distance bins for our example watershed. The final dataset contains 2,142 water samples, where we removed 1064 samples from sites with no upstream watersheds entirely outside the site's county. We remove these to ensure that our measure of upstream spraying does not capture non-water mechanisms of GLY exposure, such as dust, drift, or direct contact.

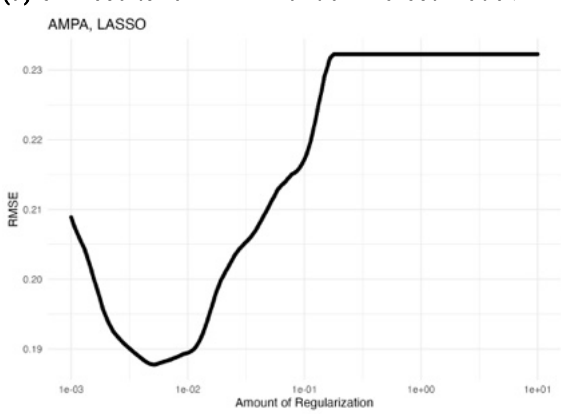
Training the water concentration ML model We train LASSO and Random Forest (RF) models using the above mentioned dataset. We generate a fully saturated set of interaction terms between GLY, soil erodibility, slope, and rainfall as predictors in the LASSO model. The month of the sample is the only other predictor variable. Since the model's primary goal is to predict GLY concentrations back in time, we train the model on 1,385 observations from after October 2015 and validate performance with 757 observations from before October 2015. Within the training set, we tune parameters using 4-fold cross-validation, where each fold trains on 15 months of data and then tests performance on the preceding six months of data. Then, we select the parameter with the lowest average RMSE across folds to estimate



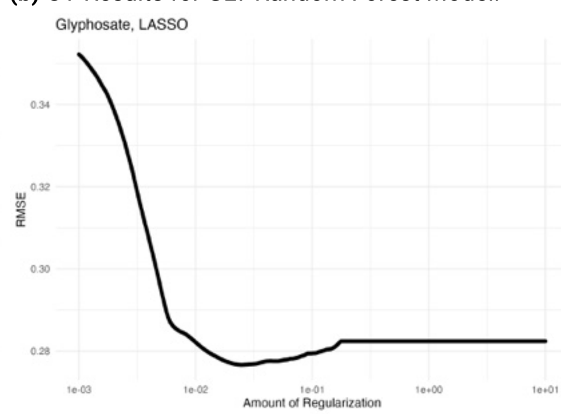
(a) CV Results for AMPA Random Forest Model.



(b) CV Results for GLY Random Forest Model.



(c) CV Results for AMPA LASSO Model.



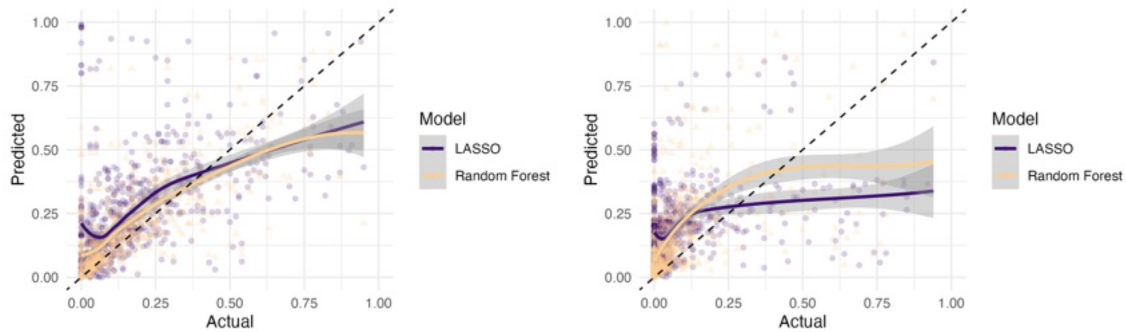
(d) CV Results for GLY LASSO Model.

Figure A28: Cross Validation Results.

the model on the entire training set. Figure A28 shows the cross-validation results.

We then assess performance of the tuned models using the 757 held out observations. Figure A29 shows the out-of-sample predictions versus their actual values. Both models predict AMPA concentrations much better than GLY concentrations, with an R-squared of 0.59 and 0.31 for the random forest and LASSO models respectively. Figure A30 shows the density of the out-of-sample predictions for each model, as well as actual values. Generally, the models slightly over-predict at low values, moreso for GLY than AMPA.

Generating predictions We use the model to predict county-month-level GLY and AMPA concentrations. We do this by making predictions for every watershed for each month between January of 1992 and December of 2017. We then take the weighted average of the predictions, where the weights are the proportion of the county's population that lives in the watershed. Our population estimates come from SEDAC's 2010 population grid [58].



(a) Predicted vs Actual AMPA Concentrations.

(b) Predicted vs Actual GLY Concentrations.

Figure A29: Out-of-sample prediction performance for LASSO and Random Forest models. Predictions are made on the 757 held-out observations in order to assess the model fit for LASSO and random forest models. Smooth lines are that of a generalized additive model.

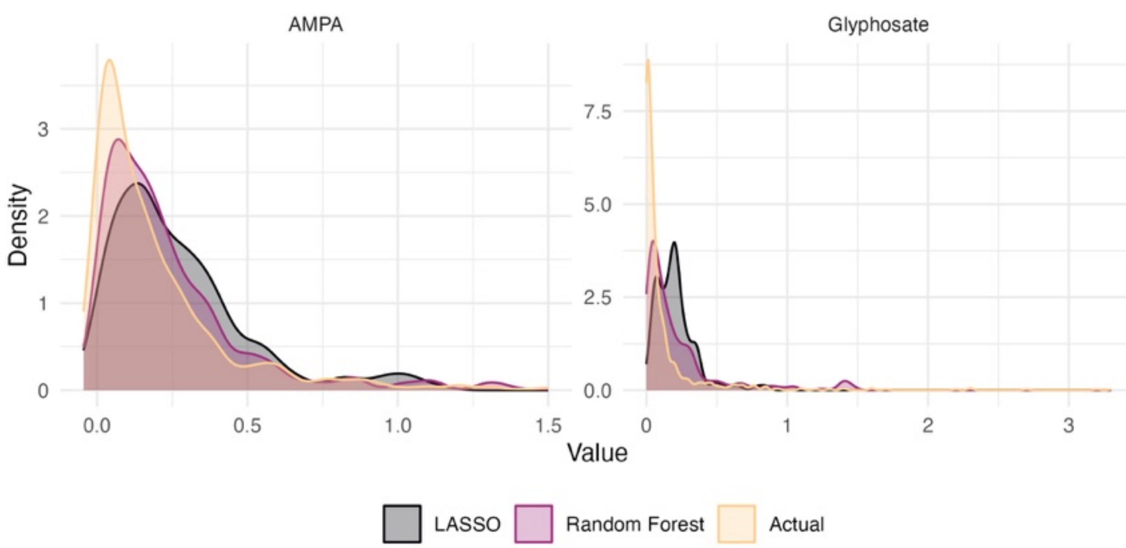


Figure A30: Density of out-of-sample predictions relative to the actual values.

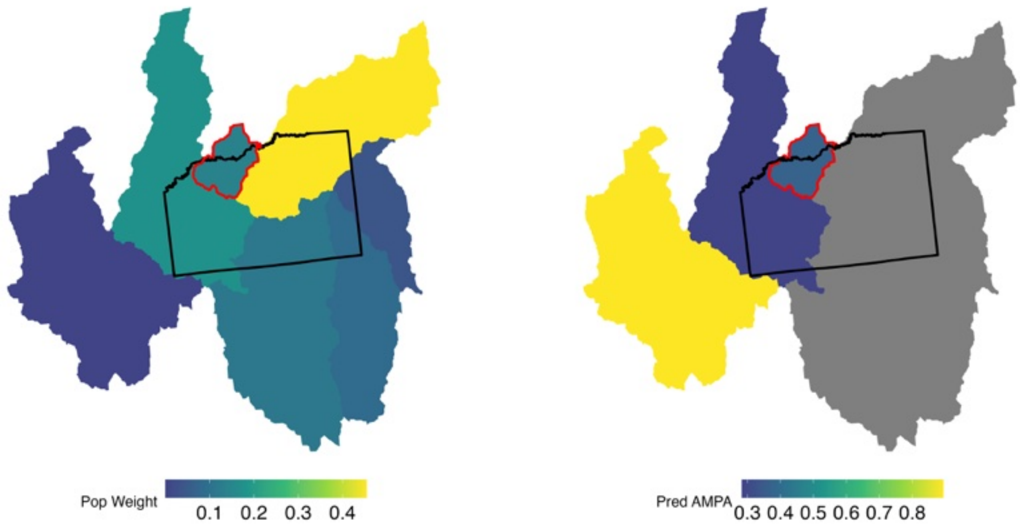


Figure A31: Aggregating watersheds to counties. For the same watershed as in Fig A27 with the red border, on the left, we have population weights for Washington County (black outline). On the right, we have our predicted AMPA in July, 2004 using the LASSO model in each watershed touching Washington County. Thus, to generate county-month level predicted AMPA, we take the weighted average of predictions (left), where the weights come from the population in each watershed (right).

This grid estimates the population for one square kilometer pixels across the United States. We add the population counts for pixels within each watershed and then divide by the total population count for cells within the county to obtain the population weights. Figure A31 shows predicted AMPA in July of 2004 from the LASSO model from each watershed touching Washington County on the right and the population weight for those watersheds on the left. Figure A32 shows predicted AMPA in water for each county in July of 2004. We can then link the county-month-level predictions of GLY and AMPA to the birth certificate data.

C.10.2 Results: Effect from Upstream GLY in Water

We find no significant perinatal health effects associated with exposure to GLY sprayed upstream of a mother's county of residence. Figure A33 displays event study plots illustrating the effect of average attainable yield in upstream watersheds on birthweight, categorized into 50-kilometer distance bins. These results suggest that having land more suitable for GM crops *upstream* of a county does not lead to a change in birthweight after the release of GM seeds in 1996.

As emphasized in Dias, Rocha, and Soares [10], the potential effects of upstream GLY spraying would be strongest in places where there is more runoff from farms. We estimate the event study allowing for heterogeneity by high-soil-erodibility and high-precipitation,

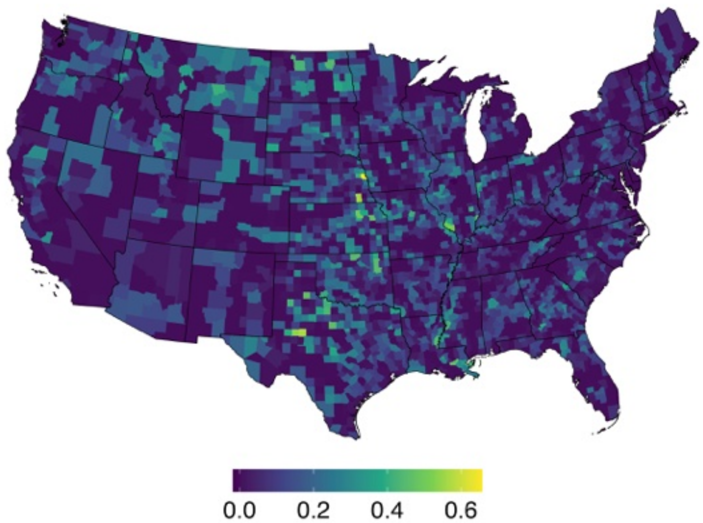


Figure A32: Predicted county-level AMPA in July of 2004. This is a map for one month (July), in one year (2004), using one of four predictive models (LASSO predicting AMPA). We generate county level predictions like this for all months and years between 1992 and 2017 with LASSO and random forest models predicting GLY and AMPA.

two factors that could increase runoff of GLY into surface water. Figure A34 shows the results for both high- and low-erodibility and precipitation. Neither demonstrate a consistent effect on birthweight.

Finally, we estimate the effect of predicted GLY and AMPA in water from the machine-learning models described above. These predictions are plausibly exogenous, as the predictions are trained only on exogenous data. Table A5 shows the results of regressing these predictions of GLY or AMPA in water on birthweight. All four estimates, coming from either a LASSO or random forest model predicting either AMPA or GLY concentrations demonstrate a null effect of GLY or AMPA on birthweight.

We approach these findings cautiously; however, they suggest the absence of substantial downstream health spillovers resulting from GLY runoff. The lack of effect may be reasonably expected in the US relative to Brazil, as drinking water treatment in the US is more robust than that in Brazil [66]. However, we cannot definitively exclude water exposure as a potential mechanism driving the local results. GLY runoff into the water could be causing issues within a county but not downstream of a county if the chemicals degrade quickly enough. Additionally, given the inherent measurement error in this process and the absence of a more refined chemical transport model, we refrain from making definitive claims about the existence of downstream spillovers from GLY use.

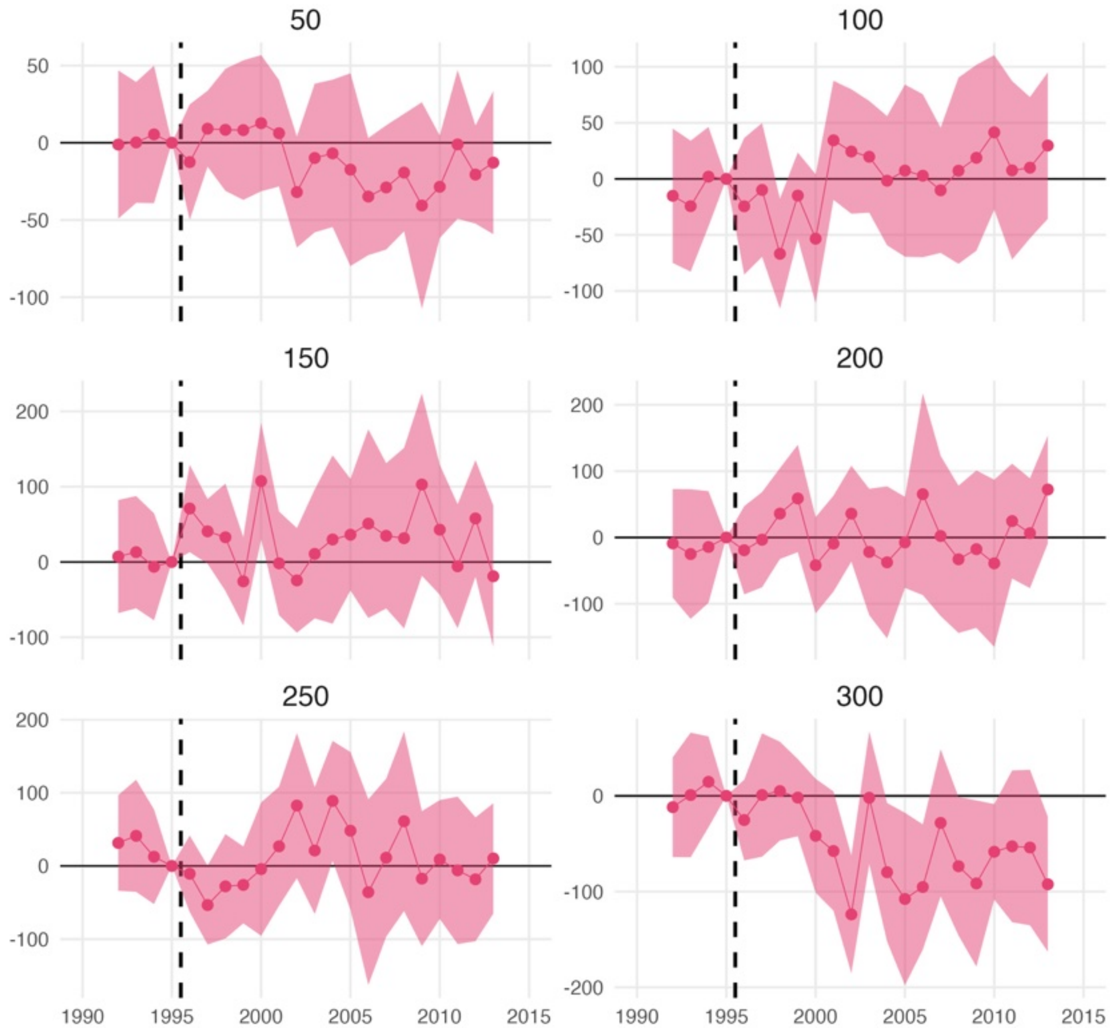


Figure A33: Effect of upstream GLY by distance bin. Estimated effect of upstream GM attainable yield percentile on various perinatal health outcomes relative to 1995. Bin labels represent the lower bound distance between the county and the upstream watershed, thus "50" is an aggregate of watersheds 50 to 100km upstream of a county. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. All regressions also control for local attainable yield interacted with year, unemployment and family demographics, including mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births from mothers with rural residence.

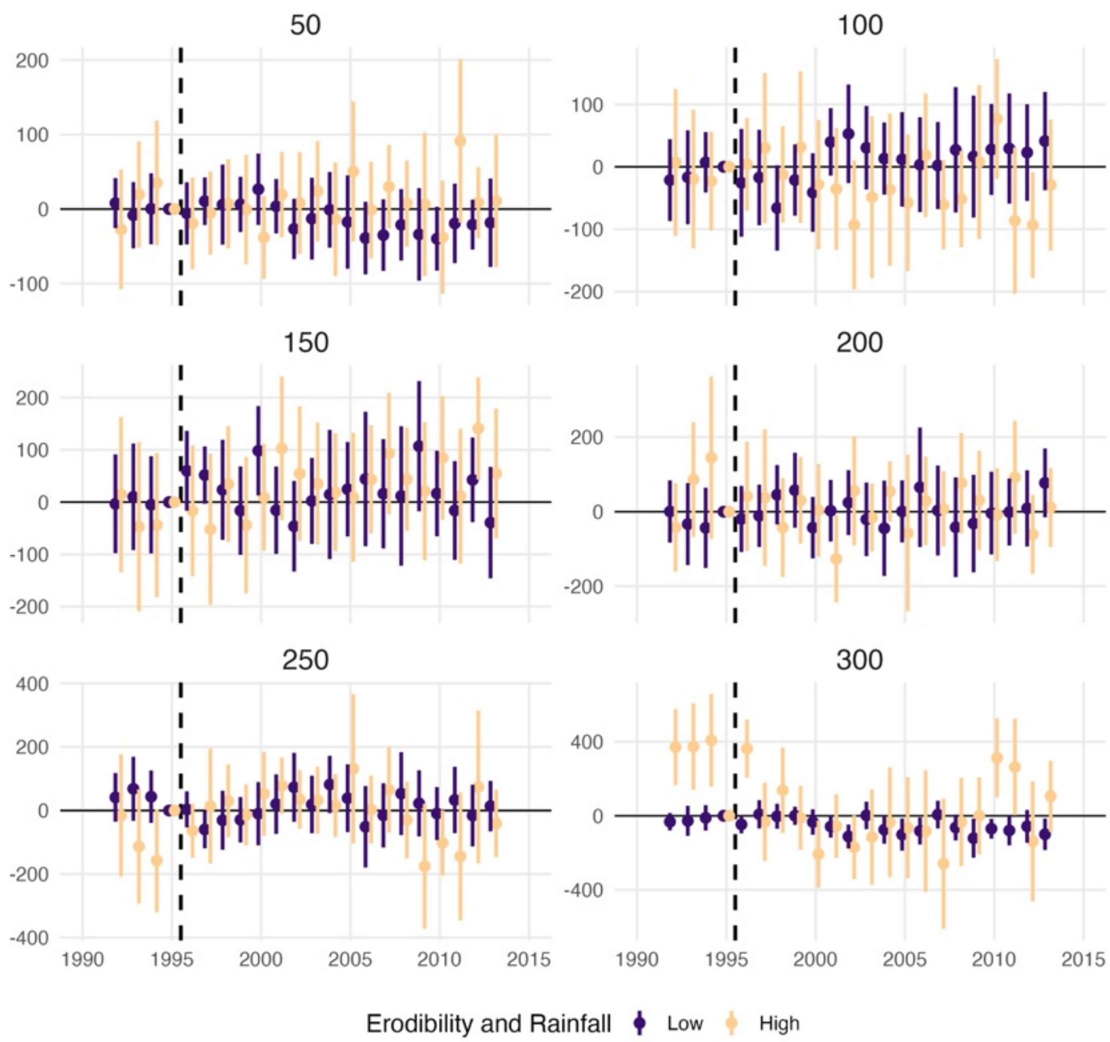


Figure A34: Effect of upstream GLY by distance bin by high and low soil erodibility and precipitation. Estimated effect of upstream GM attainable yield percentile on various perinatal health outcomes relative to 1995. Bin labels represent the lower bound distance between the county and the upstream watershed, thus "50" is an aggregate of watersheds 50 to 100km upstream of a county. All regressions include county and year by month fixed effects and standard errors are clustered by state and year. All regressions also control for local attainable yield interacted with year, unemployment and family demographics, including mother's age, race, education, marital status, birth facility, resident status, previous births, sex of infant, and father's age and race. Sample restricted to births from mothers with rural residence.

Table A5: Effect of predicted GLY or AMPA in water on birthweight.

Dep Var	BW			
Model:	(1)	(2)	(3)	(4)
Predicted AMPA (LASSO)	16.1			
	(11.2)			
Predicted AMPA (RF)		-3.14		
		(8.83)		
Predicted GLY (LASSO)			-6.01	
			(16.7)	
Predicted GLY (RF)				-5.94
				(6.69)
Local attainable yield	Yes	Yes	Yes	Yes
Local pesticides	Yes	Yes	Yes	Yes
Unemployment	Yes	Yes	Yes	Yes
Fixed-effects				
Family Demog	Yes	Yes	Yes	Yes
Yr x Mo + Cnty	Yes	Yes	Yes	Yes
Fit statistics				
N (millions)	7.910	7.910	7.910	7.910

Clustered (Year & State) standard-errors in parentheses