Marginal causal effect of fungicide seed treatments on soybean yield and uncertain profitability in the US Midwest

Supplementary Figures, Tables, Methods, Appendix

27 February, 2026

Figures

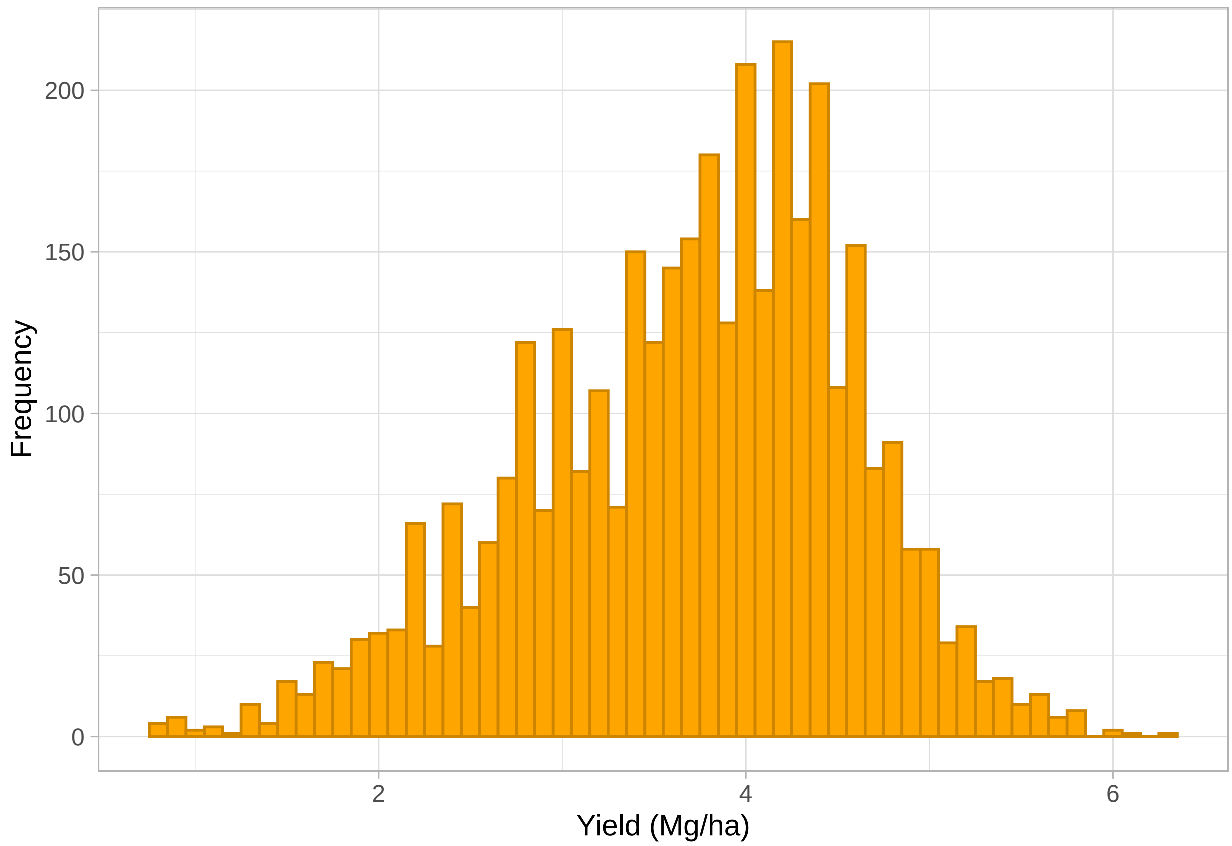

Figure S1: Distribution of grower-reported soybean yield in 3,614 fields. Data were collected from 2014 to 2016 across ten Midwest states.

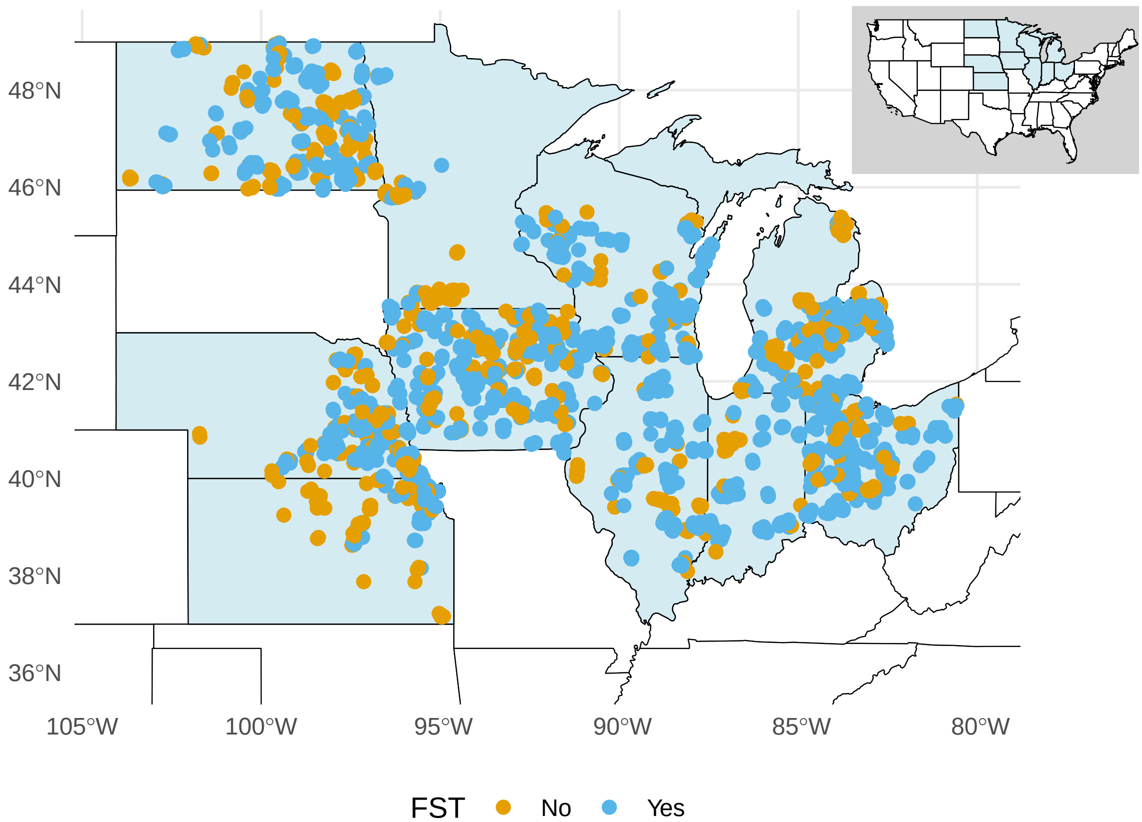

Figure S2: Locations of 3,614 Midwest soybean fields assessed for yield and management practices. Fields were sown in 2014, 2015 or 2016. Locations are colored by whether the field was sown to fungicide-treated seed (FST = Yes) or not (FST = No). Inset: the 10 states in which the fields were located, relative to the contiguous United States.

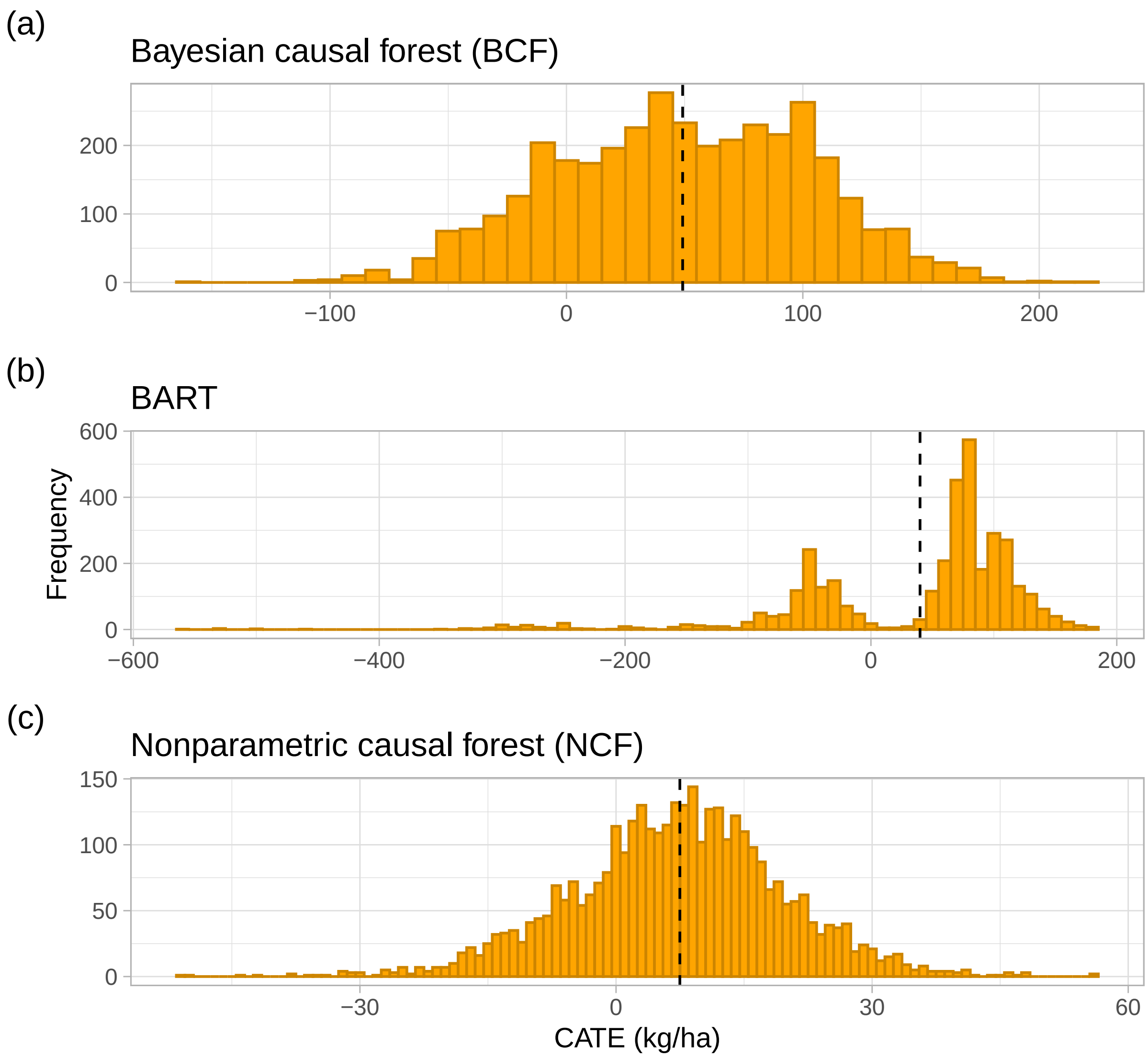

Figure S3: Histograms of the 3,614 individual conditional average treatment effect (CATE) yield estimates due to sowing fungicide-treated soybean seed in the Midwest United States. Estimates were produced by three methods: **(a)** Bayesian causal forest (BCF), **(b)** Bayesian additive regression trees (BART) and **(c)** Nonparametric causal forest (NCF) via generalized random forests. The x-axis scales are different for each panel, but the binwidth is the same (10) for panels **a** and **b**. The dashed vertical black line is at the respective mean for each distribution.

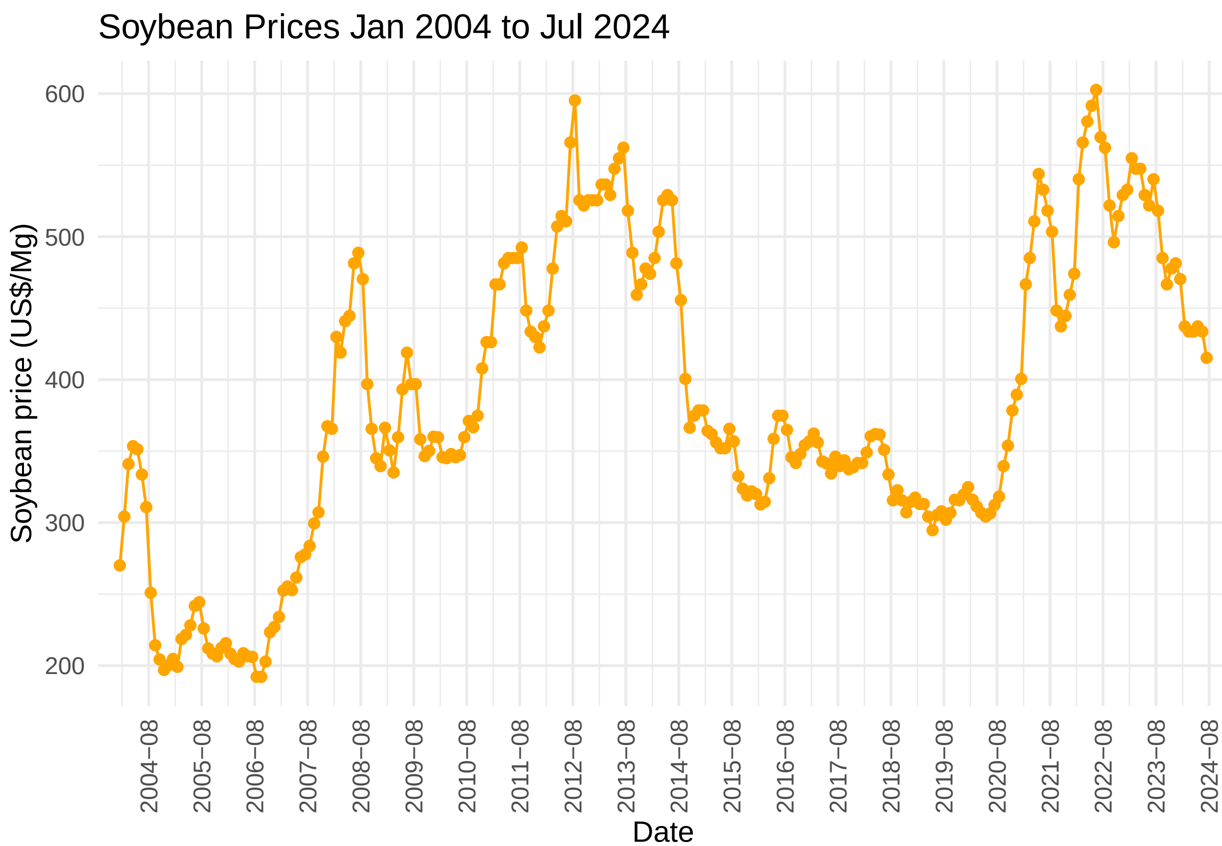

Figure S4: Soybean prices (US$/Mg) from 2004 to 2024. Soybean prices were obtained from the United States Department of Agriculture's National Agricultural Statistics Service (https://www.nass.usda.gov/).

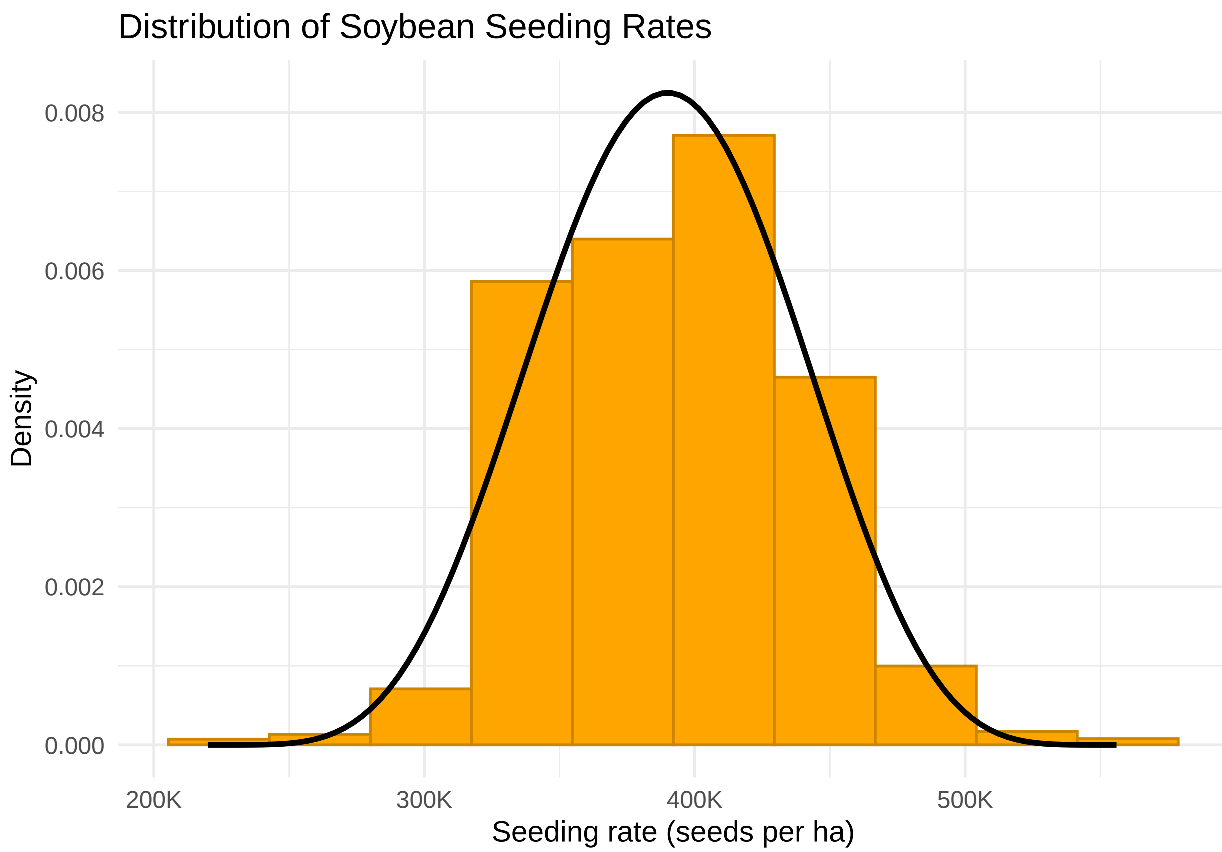

Figure S5: Histogram of soybean seeding rate in 5,184 US Midwest fields with a fitted beta distribution density overlay.

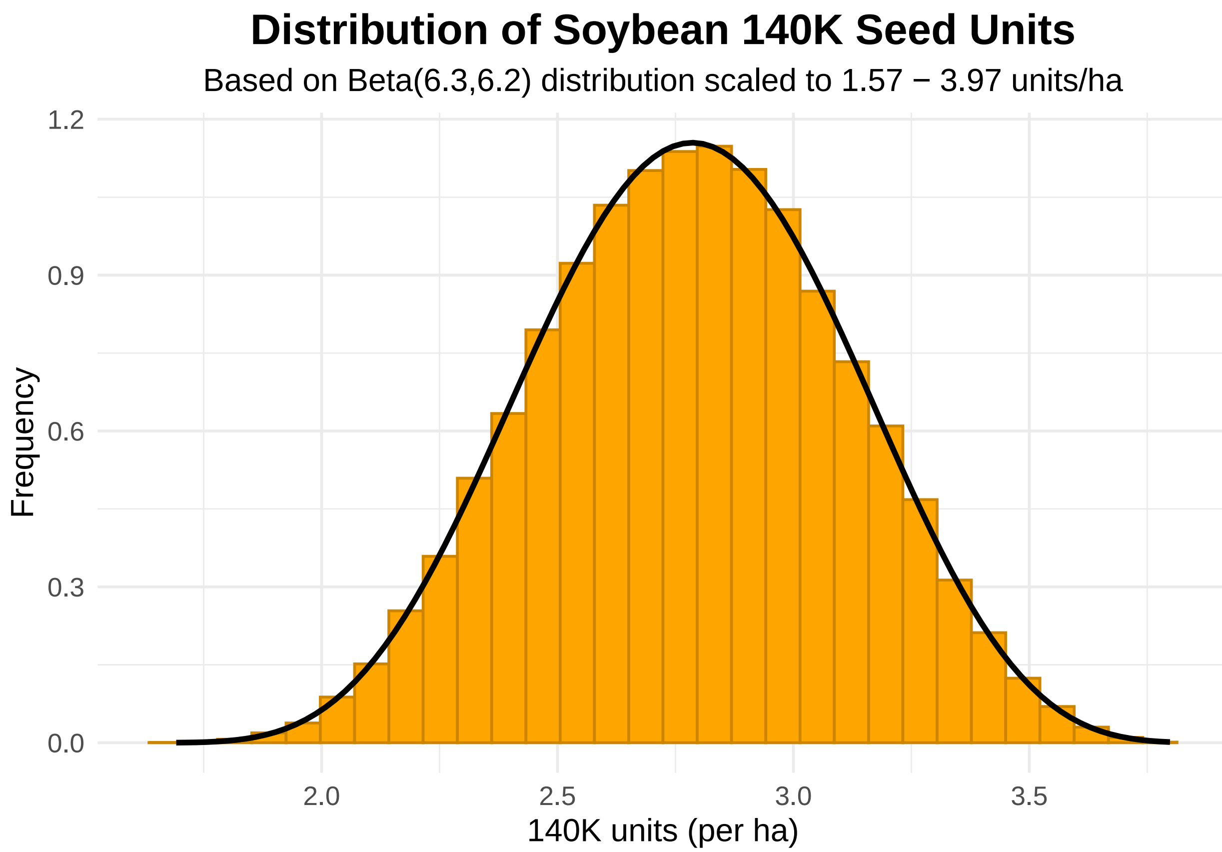

Figure S6: Histogram of 50,000 simulated seeding rates from a fitted beta distribution, expressed as 140K units sown per ha.

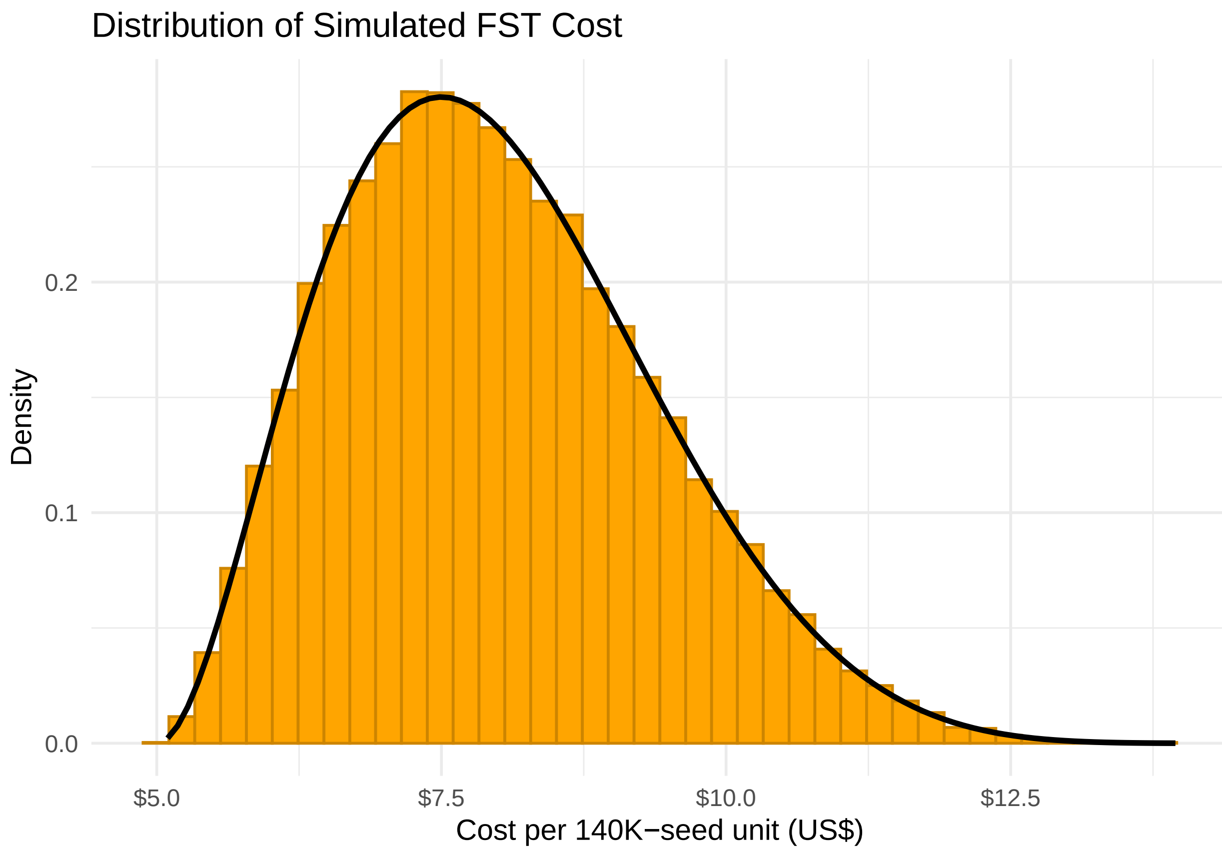

Figure S7: Histogram of 50,000 simulated values of fungicide seed treatment (FST) cost per 140K-seed unit, generated from an assumed beta distribution. The histogram is overlaid with the density curve of the beta distribution.

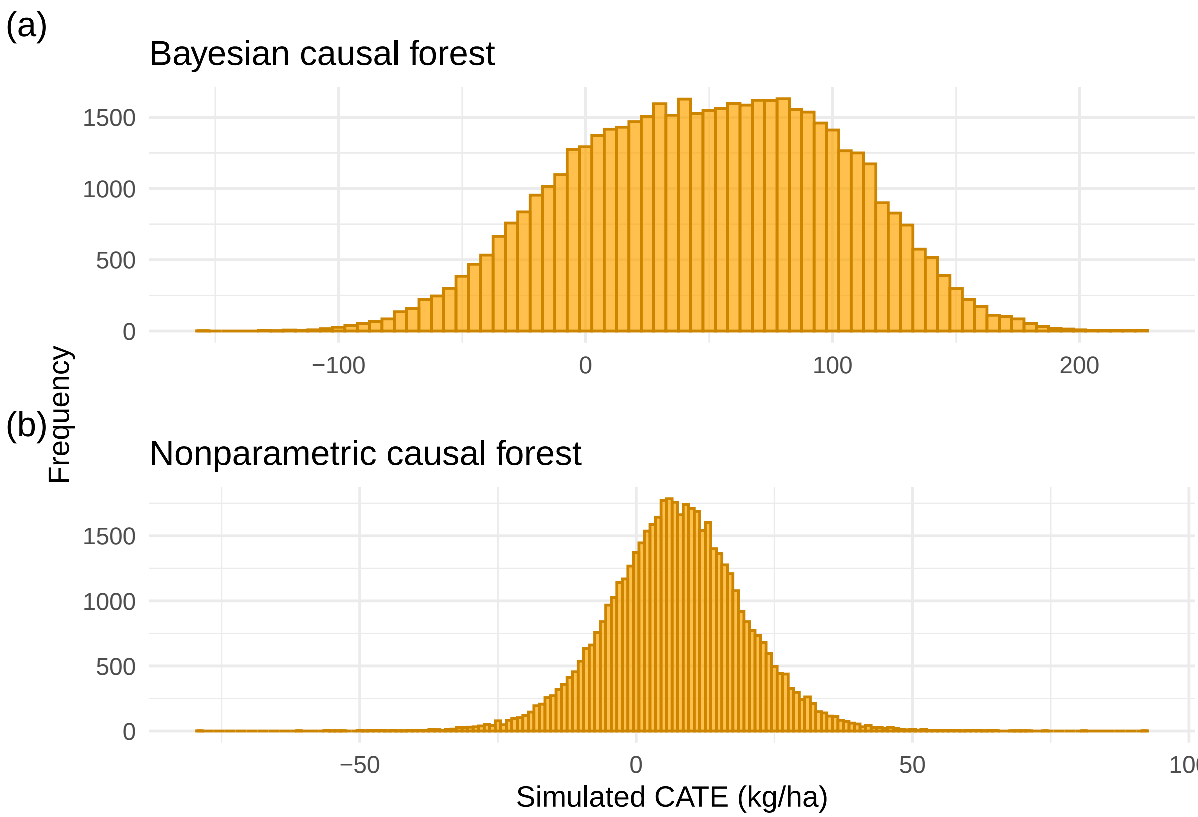

Figure S8: Histograms of simulated individual conditional average treatment effect (CATE) yield effects due to sowing fungicide-treated soybean seed. Values were simulated from **(a)** a two-component Gaussian mixture distribution fitted to the CATE values from a Bayesian causal forest; and **(b)** a non-standardized t-distribution fitted to the CATE values from a nonparametric causal forest via generalized random forests. 50,000 values were drawn from each distribution. Note the x-axis scales are different between the two panels.

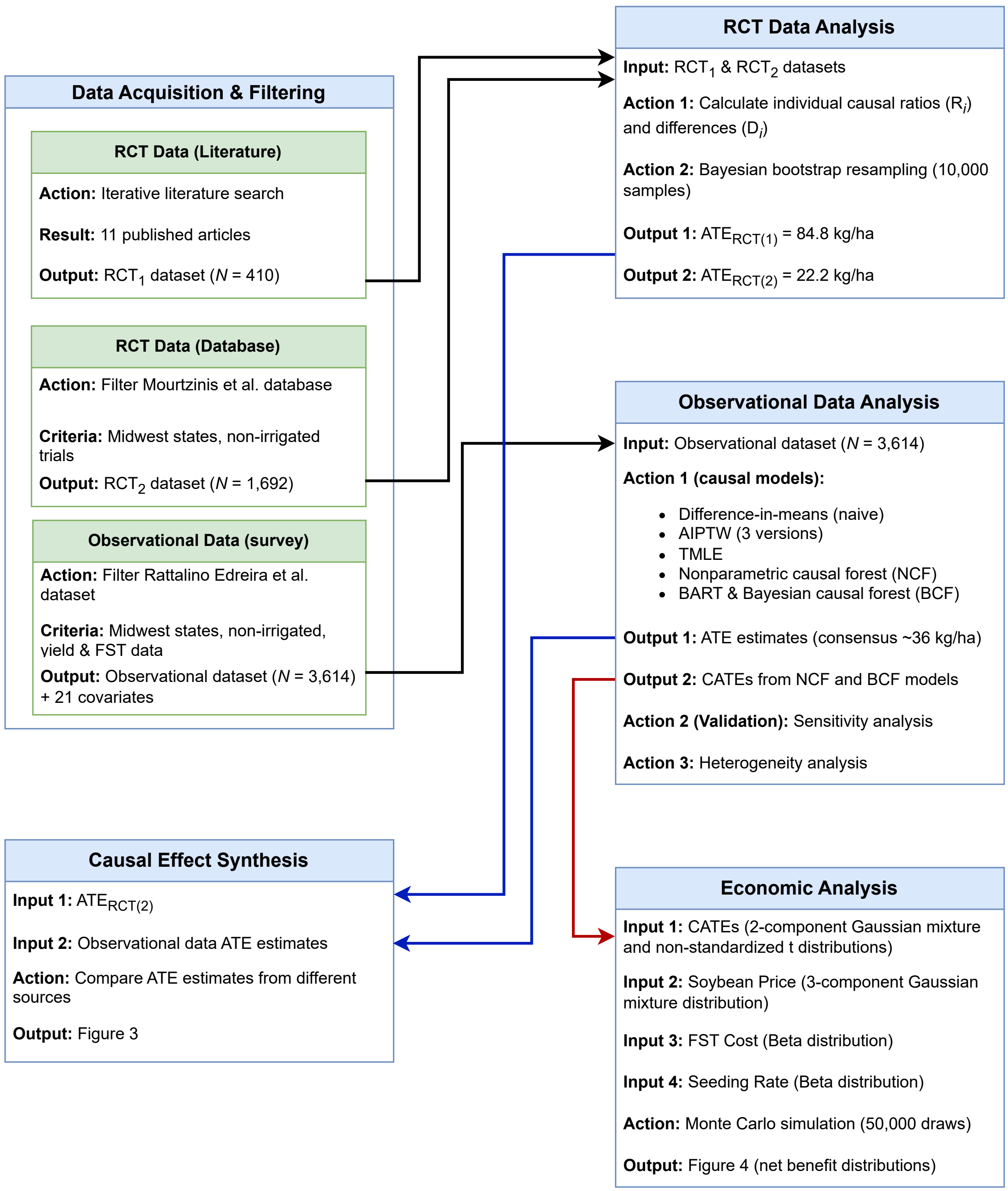

Figure S9: Flow diagram of the data acquisition, filtering, and analysis pipeline.

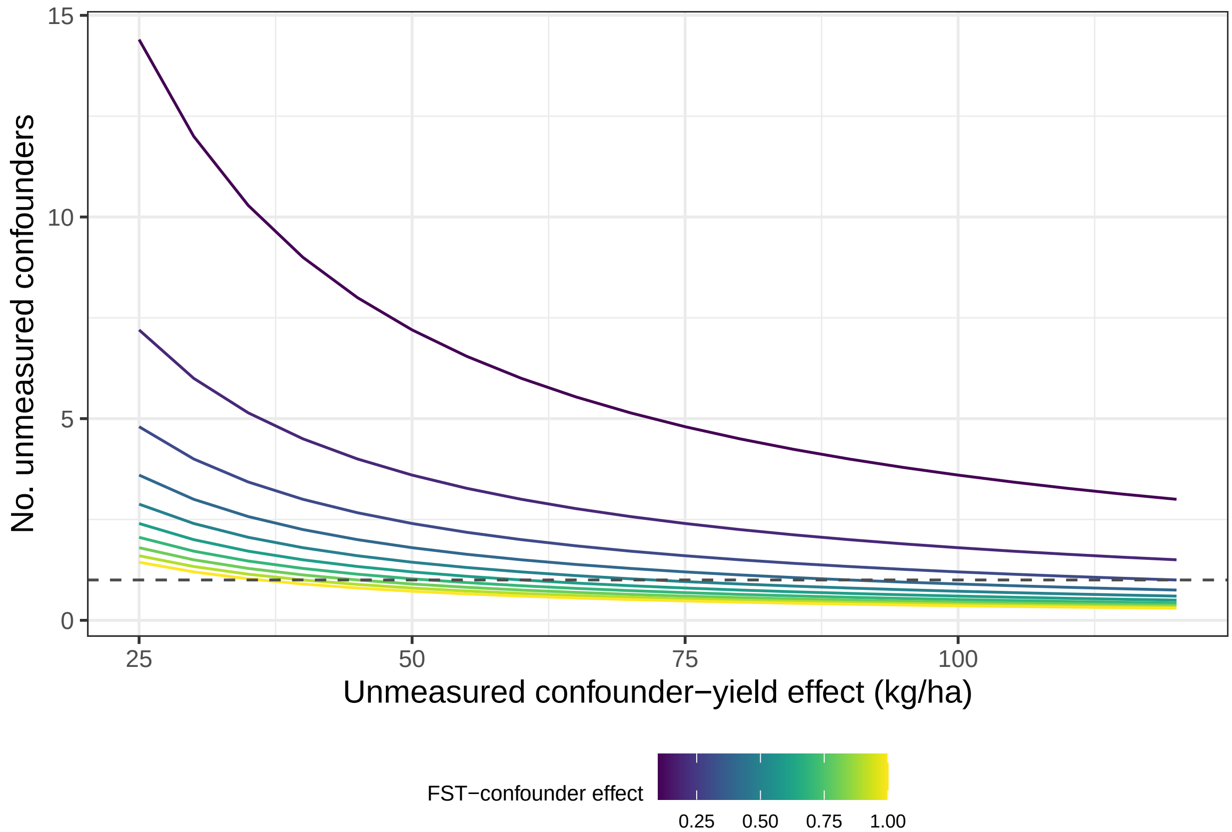

Figure S10: The number of unmeasured continuous confounders (variables) required to nullify the average treatment effect of fungicide seed treatment (FST) on yield in 3,614 US Midwest soybean fields, assuming a FST effect of 36 kg/ha, the consensus among various doubly-robust causal analysis methods applied to the observational dataset. The yield effect that an unmeasured continuous confounder is posited to have is shown on the x-axis. The color scheme indicates the difference in the means of the normalized unmeasured confounder between soybean fields sown with fungicide-treated seed and those sown with untreated seed. The y-axis is the number of unmeasured continuous confounders that would be required to nullify the FST causal effect on yield as a function of the effects the unmeasured confounders would need to have on both yield and the probability of FST use. For example, it would take five additional unmeasured continuous variables, each with an average yield effect of 25 kg/ha and each with a normalized difference of 0.3 between the two seed treatment groups (FST and no FST) to bring the causal effect of FST down to zero. For reference, the average continuous-confounder-yield effect estimated from the observed data was 78.7 kg/ha, and the estimated mean FST-continuous-confounder effect from the data was 0.13.

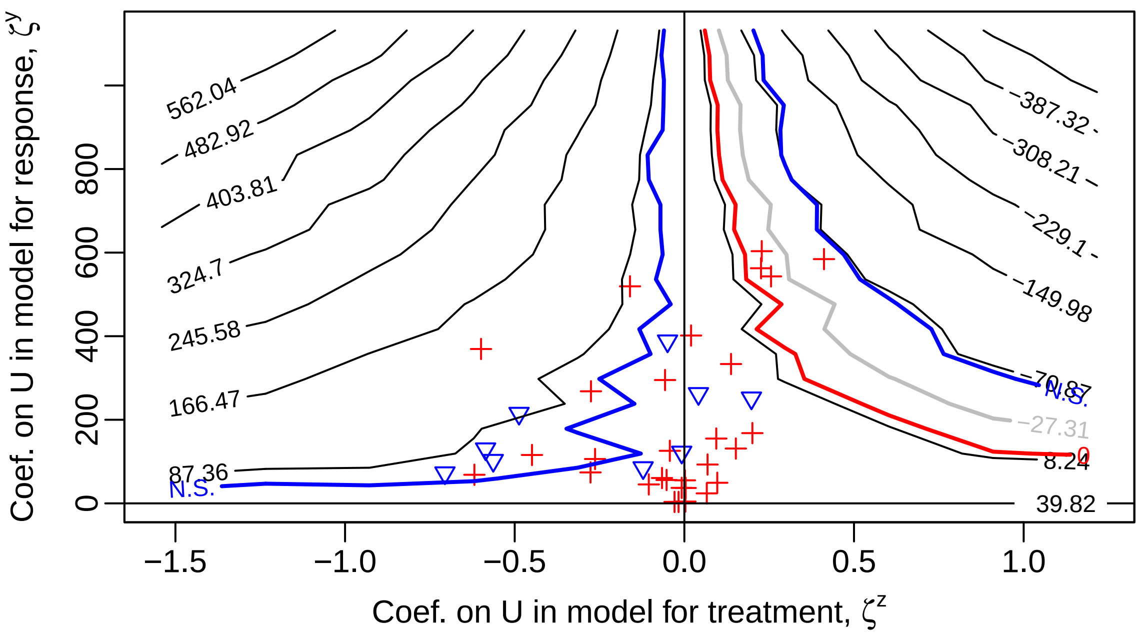

Figure S11: Bayesian sensitivity analysis for the effects of an unmeasured confounder on the causal effect of sowing fungicide-treated seed (FST) on soybean yield in 3,614 US Midwest fields. All variables are shown on their original measurement scale. The parameter values of the unmeasured confounder (when added to the model for yield and for the treatment) are shown on the x-axis (ζ^z = the coefficient of the unmeasured confounder in the model for the treatment, namely FST in our case) and y-axis (ζ^y = the coefficient of the unmeasured confounder in the model for the response, namely yield in our case). The thin black lines show the basic effect of the unmeasured confounder. The blue lines marked 'N.S.' show where the estimated FST effect would no longer be considered statistically significant (visualize moving from the outside inwards into the region bordered by the blue lines). The red line in the middle of the blue lines shows how strong an unmeasured confounder would need to be to completely cancel out the FST effect. The thick gray line represents what the FST effect would be if there was an unmeasured confounder as strong as the covariate measured with the strongest marginal effect. The original FST estimate, which doesn't account for any potential unmeasured confounders, is at the bottom right of the graph. Predictors with negative coefficients for the response surface have been transformed through multiplication by -1 and are displayed as inverted triangles.

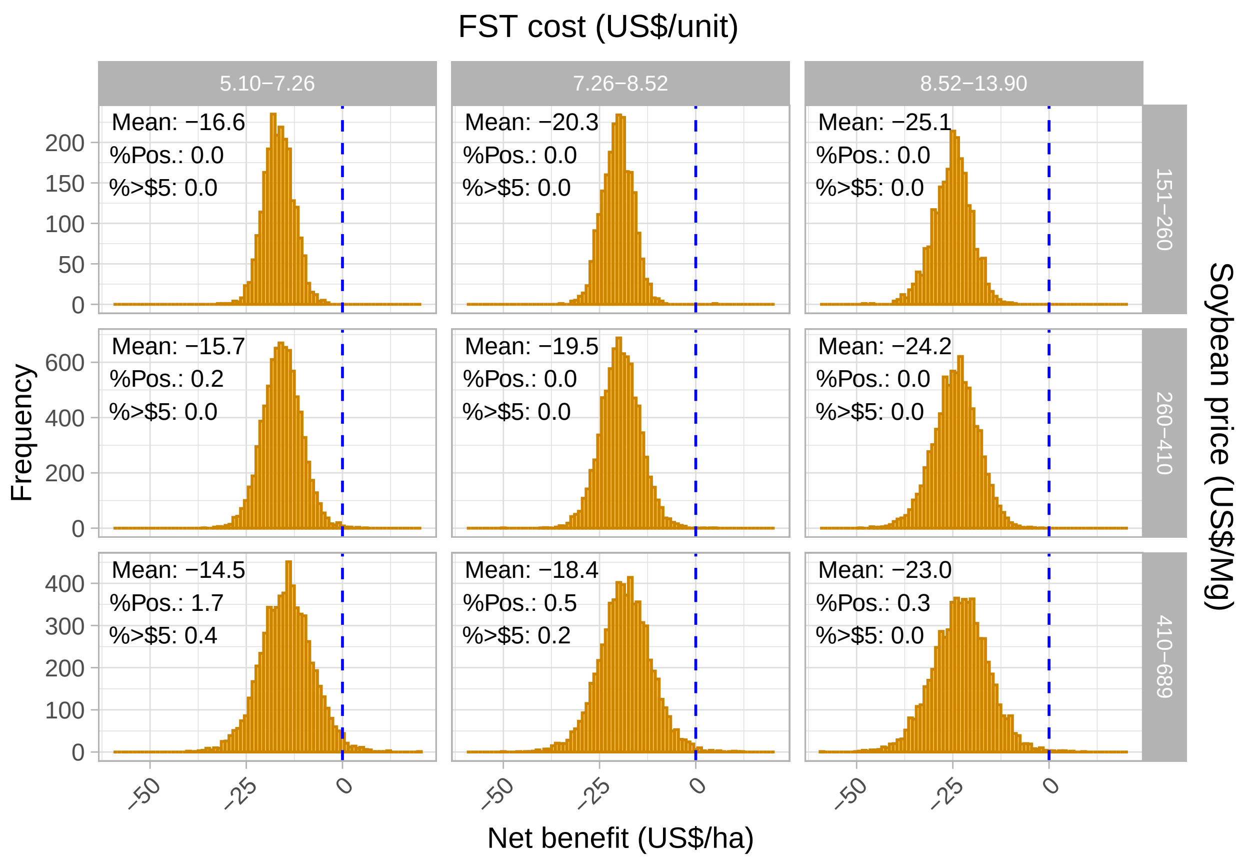

Figure S12: Distributions of the net financial benefit (return on investment) of using fungicide-treated seed (FST) in Midwest US soybean production estimated from 50,000 Monte Carlo simulations. Inputs to the Monte Carlo simulations were values sampled from probability distributions fit to (i) the individual conditional average treatment effects (CATEs) based on a nonparametric causal forest fit to data from the N = 3,614 observational fields, (ii) soybean price distribution over the 2004 to 2024 period, (iii) soybean seeding rate distribution derived from N = 5,184 observations, and (iv) a per-unit FST cost bounded between US$5 and US$15 per unit (140,000 seeds per unit). Results are faceted by three ranges for FST cost (columns; top labels) and soybean price (rows; right labels). The vertical dashed, black line in each panel references zero benefit. Three statistics are given in each panel: the mean of the distribution, the percentage of net benefit values that were positive (%Pos), the percentage of net benefit values that were greater than 5 US$ (%>$5).

Table S1. Properties of variables used in the causal modeling of Midwest soybean yield and fungicide seed treatment (FST). FST = 0 indicates that no FST was used in sowing the field; FST = 1 indicates that seeds were treated with a fungicide prior to sowing.

Variable	**FST = 0** **N = 1,053**[a]	**FST = 1** **N = 2,561**
Artificial drainage		
No	482 (46%)	1,024 (40%)
Yes	571 (54%)	1,537 (60%)
Type of tillage		
No-Till	546 (52%)	1,204 (47%)
Other	484 (46%)	1,302 (51%)
Unknown	23 (2.2%)	55 (2.1%)
Lime application		
No	946 (90%)	2,266 (88%)
Yes	107 (10%)	295 (12%)
Manure application		
No	966 (92%)	2,335 (91%)
Yes	87 (8.3%)	226 (8.8%)
Soybean cultivar maturity group		
0	120 (11%)	236 (9.2%)
I	135 (13%)	244 (9.5%)
II	411 (39%)	1,073 (42%)
III	241 (23%)	796 (31%)
IV	40 (3.8%)	70 (2.7%)
Unknown	106 (10%)	142 (5.5%)
Use of a starter fertilizer		
No	915 (87%)	2,251 (88%)
Yes	138 (13%)	310 (12%)
Plant available water holding capacity in the rooting zone (mm)		
1	3 (0.3%)	20 (0.8%)

2	30 (2.8%)	108 (4.2%)
3	131 (12%)	383 (15%)
4	113 (11%)	318 (12%)
5	180 (17%)	402 (16%)
6	492 (47%)	1,069 (42%)
7	104 (9.9%)	261 (10%)
Annual total growing degree days (°Cd)		
01	10 (0.9%)	15 (0.6%)
02	189 (18%)	361 (14%)
03	478 (45%)	1,159 (45%)
04	355 (34%)	1,017 (40%)
05	21 (2.0%)	9 (0.4%)
Aridity index		
2	35 (3.3%)	59 (2.3%)
3	124 (12%)	220 (8.6%)
4	79 (7.5%)	78 (3.0%)
5	197 (19%)	281 (11%)
6	294 (28%)	813 (32%)
7	191 (18%)	617 (24%)
8	133 (13%)	493 (19%)
Use of P_2O_5 non-starter fertilizer		
Applied	260 (25%)	751 (29%)
None	793 (75%)	1,810 (71%)
Use of K_2O non-starter fertilizer		
Applied	318 (30%)	1,023 (40%)
None	735 (70%)	1,538 (60%)
Root zone depth (cm)		
< 150 cm	367 (35%)	850 (33%)
150 cm	686 (65%)	1,711 (67%)

Sowing date (day of the year)	0.15 (-0.77, 0.74)	0.07 (-0.85, 0.57)
Row spacing (cm)	-0.47 (-0.47, 1.41)	-0.47 (-0.47, 1.41)
Seeding rate (seeds/ha)	0.16 (-0.96, 0.73)	-0.12 (-0.73, 0.45)
Percent organic matter in the topsoil (0-30 cm depth)	-0.19 (-0.58, 0.35)	-0.19 (-0.52, 0.32)
Topsoil pH	-0.11 (-0.64, 0.85)	-0.29 (-0.66, 0.67)
Topographic wetness index	0.02 (-0.65, 0.74)	0.03 (-0.66, 0.72)
Logratio of topsoil sand:silt	-0.05 (-0.80, 0.78)	-0.08 (-0.80, 0.76)
Logratio of topsoil clay:silt	-0.01 (-0.61, 0.80)	-0.04 (-0.59, 0.71)
Snow water equivalent from the beginning of the year to sowing (kg/m^2)	-0.38 (-0.80, 0.37)	-0.24 (-0.71, 0.53)
Soybean yield (Mg/ha)	3.70 (2.92, 4.23)	3.88 (3.16, 4.37)

[a] N = number of fields in each FST category (FST = 0 and FST = 1). For the categorical variables, N is subset for each level of the variable, and the percentage of fields in each subcategory is given in parentheses. For the continuous variables, which have been centered and scaled, the median is given followed by the 1st and 3rd quantile values in brackets.

Table S2: Means of seeding rate, yield return and net benefit conditioned by fungicide seed treatment (FST) cost and soybean price, based on 50,000 Monte Carlo simulations. Yield return was sampled from distributions representing the estimated individual conditional average treatment (CATE) values from Bayesian causal forests (BCF) and nonparametric causal forests (NCF).

			BCF		NCF	
FST cost (US$/unit)	Soybean price (US$/Mg)	Seeding rate (units/ha)	Return (Mg/ha)	Net benefit (US$/ha)[a]	Return (Mg/ha)	Net benefit (US$/ha)
5.10-7.26	151-260	2.785	0.051	-7.133	0.008	-16.581
5.10-7.26	260-410	2.784	0.049	-1.347	0.007	-15.722
5.10-7.26	410-689	2.782	0.049	6.167	0.007	-14.537
7.26-8.52	151-260	2.787	0.048	-11.410	0.008	-20.286
7.26-8.52	260-410	2.787	0.048	-5.413	0.007	-19.523
7.26-8.52	410-689	2.785	0.049	2.393	0.007	-18.356
8.52-13.90	151-260	2.785	0.050	-15.900	0.007	-25.146
8.52-13.90	260-410	2.783	0.049	-10.050	0.007	-24.230
8.52-13.90	410-689	2.781	0.048	-2.637	0.007	-22.950

[a] Net benefit was calculated under both assumed CATE distributions (BCF and NCF).

Methods

RCT-based data

Bayesian bootstrap resampling was done using the 'bayesboot'package (version 0.2.2) [1].

Observational data

Causal analysis

Propensity scores were estimated using the 'WeightIt' package (version 1.3.2) [2]. The AIPTW models were fitted using functions from the 'xgboost' (version 1.7.8.1) [3], 'SuperLearner' (version 2.0-29) [4] and 'AIPW' (version 0.6.9) [5] packages. TMLE estimation was done with the 'sl3' (version 1.4.5) [6] and 'tmle3' (version 0.2.0) [7] packages. Nonparametric causal forests were fitted with the 'grf' package (version 2.4.0) [8] and Bayesian causal forests with the 'bcf' package (version 2.0.2) [9]. The BART model was fitted using the 'bartCause' (version 1.0-9) [10] and 'stan4bart' (version 0.0-10) [11] packages.

Sensitivity analysis

The parametric sensitivity analysis was done with the 'tipr' package (version 1.0.2) [12]. Bayesian SA was done with the 'treatSens' package (version 3.0) [13].

Heterogeneity in FST effects

The heterogeneity analysis was implemented with the 'grf', 'rpart' (version 4.1.23) [14], 'rpart.plot' (version 3.1.2) [15] and 'partykit' (version 1.2-23) [16] packages.

References

1. Bååth, R. bayesboot: An implementation of Rubin's (1981) Bayesian bootstrap. (2018).
2. Greifer, N. WeightIt: weighting for covariate balance in observational studies. (2024).
3. Chen, T. *et al.* xgboost: Extreme gradient boosting. (2024).
4. Polley, E., LeDell, E., Kennedy, C. & Laan, M. van der. SuperLearner: Super learner prediction. (2024).
5. Zhong, Y., Kennedy, E. H., Bodnar, L. M. & Naimi, A. I. AIPW: an R package for augmented inverse probability weighted estimation of average causal effects. *Am. J. Epidemiol.* **190**, 2690–2699 (2021).
6. Coyle, J., Hejazi, N., Malenica, I., Phillips, R. & Sofrygin, O. sl3: Pipelines for machine learning and super learning. https://doi.org/10.5281/zenodo.1342293 (2025).
7. Coyle, J. tmle3: The extensible tmle framework. https://doi.org/10.5281/zenodo.4603358 (2025).
8. Tibshirani, J., Athey, S., Sverdrup, E. & Wager, S. grf: Generalized random forests. (2024).
9. Hahn, P. R., Murray, J. S. & Carvalho, C. M. Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Anal.* **15**, 965–2020 (2020).
10. Hill, J. L. Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Stat.* **20**, 217–240 (2011).
11. Dorie, V. stan4bart: Bayesian additive regression trees with Stan-sampled parametric extensions. (2024).

12. DÁgostino McGowan, L. tipr: An R package for sensitivity analyses for unmeasured confounders. *J. Open Source Softw.* **7**, 4495 (2022).

13. Carnegie, N. B., Harada, M., Dorie, V. & Hill, J. treatSens: A package to assess sensitivity of causal analyses to unmeasured confounding. (2024).

14. Therneau, T. & Atkinson, B. rpart: Recursive partitioning and regression trees. (2023).

15. Milborrow, S. rpart.plot: plot 'rpart' models: an enhanced version of 'plot.rpart'. (2024).

16. Hothorn, T. & Zeileis, A. partykit: A modular toolkit for recursive partytioning in R. *J. Mach. Learn. Res.* **16**, 3905–3909 (2015).

Appendix

Studies from which data were extracted to compile the RCT_1 dataset.

Bierman, R.E., Riechers, D.E., Sprague, C.L., Bollero, G. and Pedersen, W.L., 2006. Fungicide–herbicide interaction in soybean (*Glycine max*). Crop Protection, 25(2), pp.134-139. https://doi.org/10.1016/j.cropro.2005.03.013.

Bradley, C.A., 2008. Effect of fungicide seed treatments on stand establishment, seedling disease, and yield of soybean in North Dakota. Plant Disease, 92(1), pp.120-125. https://doi.org/10.1094/PDIS-92-1-0120.

Dorrance, A.E., Robertson, A.E., Cianzo, S., Giesler, L.J., Grau, C.R., Draper, M.A., Tenuta, A.U. and Anderson, T.R., 2009. Integrated management strategies for *Phytophthora sojae* combining host resistance and seed treatments. Plant Disease, 93(9), pp.875-882. https://doi.org/10.1094/PDIS-93-9-0875.

Gaspar, A.P., Marburger, D.A., Mourtzinis, S. and Conley, S.P., 2014. Soybean seed yield response to multiple seed treatment components across diverse environments. Agronomy Journal, 106(6), pp.1955-1962. https://doi.org/10.2134/agronj14.0277.

Guy, S.O. and Oplinger, E.S., 1989. Soybean cultivar performance as influenced by tillage system and seed treatment. Journal of Production Agriculture, 2(1), pp.57-62. https://doi.org/10.2134/jpa1989.0057.

Haafke, A.J., Kandel, Y.R., Batzer, J.C., Dangal, N.K. and Mueller, D.S., 2023. Effects of seed treatments and yield potential on infected soybean seed. Plant Health Progress, 24(4), pp.453-461. https://doi.org/10.1094/PHP-03-23-0019-RS.

Lueschen, W.E., Evans, S.D., Ford, J.H., Hoverstad, T.R., Kanne, B.K., Orf, J.H., Staricka, J.A., Stienstra, W.C., Warnes, D.D. and Hicks, D.R., 1991. Soybean production as affected by tillage in a corn and soybean management system: II. Seed treatment response. Journal of Production Agriculture, 4(4), pp.580-585. https://doi.org/10.2134/jpa1991.0580.

Orlowski, J.M., Haverkamp, B.J., Laurenz, R.G., Marburger, D.A., Wilson, E.W., Casteel, S.N., Conley, S.P., Naeve, S.L., Nafziger, E.D., Roozeboom, K.L. and Ross, W.J., 2016. High-input management systems effect on soybean seed yield, yield components, and economic break-even probabilities. Crop Science, 56(4), pp.1988-2004. https://doi.org/10.2135/cropsci2015.10.0620.

Rossman, D.R., Byrne, A.M. and Chilvers, M.I., 2018. Profitability and efficacy of soybean seed treatment in Michigan. Crop Protection, 114, pp.44-52. https://doi.org/10.1016/j.cropro.2018.08.003.

Schulz, T.J. and Thelen, K.D., 2008. Soybean seed inoculant and fungicidal seed treatment effects on soybean. Crop Science, 48(5), pp.1975-1983. https://doi.org/10.2135/cropsci2008.02.0108.

Scott, K., Eyre, M., McDuffee, D. and Dorrance, A.E., 2020. The efficacy of ethaboxam as a soybean seed treatment toward *Phytophthora*, *Phytopythium*, and *Pythium* in Ohio. Plant Disease, 104(5), pp.1421-1432. https://doi.org/10.1094/PDIS-09-19-1818-RE.

Computational environment

```
## [1] "R version 4.4.2 (2024-10-31 ucrt)"
```

```
## [1] "x86_64, mingw32"
```

```
## R version 4.4.2 (2024-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26200)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] grid      stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] treatSens_3.0           viridis_0.6.5           viridisLite_0.4.2
##  [4] tipr_1.0.2              moments_0.14.1          rnaturalearthdata_1.0.0
##  [7] rnaturalearth_1.0.1     sf_1.0-19               kableExtra_1.4.0
## [10] gridExtra_2.3           patchwork_1.3.0         lubridate_1.9.4
## [13] forcats_1.0.0           stringr_1.5.1           dplyr_1.1.4
## [16] purrr_1.0.4             readr_2.1.5             tidyr_1.3.1
## [19] tibble_3.2.1            ggplot2_3.5.1           tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.2.1              farver_2.1.2
##  [3] fastmap_1.2.0                 digest_0.6.37
##  [5] timechange_0.3.0              lifecycle_1.0.4
##  [7] terra_1.8-21                  magrittr_2.0.3
##  [9] compiler_4.4.2                rlang_1.1.5
## [11] tools_4.4.2                   yaml_2.3.10
## [13] knitr_1.49                    labeling_0.4.3
## [15] classInt_0.4-11               here_1.0.1
## [17] xml2_1.3.6                    KernSmooth_2.23-24
## [19] withr_3.0.2                   e1071_1.7-16
## [21] colorspace_2.1-1              scales_1.3.0
## [23] MASS_7.3-61                   cli_3.6.4
## [25] mvtnorm_1.3-3                 rmarkdown_2.29
## [27] reformulas_0.4.0              generics_0.1.3
## [29] rstudioapi_0.17.1             httr_1.4.7
## [31] tzdb_0.4.0                    readxl_1.4.3
```

```
## [33] minqa_1.2.8                 DBI_1.2.3
## [35] proxy_0.4-27                splines_4.4.2
## [37] parallel_4.4.2              cellranger_1.1.0
## [39] vctrs_0.6.5                 boot_1.3-31
## [41] Matrix_1.7-1                jsonlite_1.9.0
## [43] hms_1.1.3                   systemfonts_1.2.1
## [45] units_0.8-5                 rnaturalearthhires_1.0.0.9000
## [47] rematch_2.0.0               glue_1.8.0
## [49] nloptr_2.1.1                codetools_0.2-20
## [51] stringi_1.8.4               gtable_0.3.6
## [53] lme4_1.1-36                 munsell_0.5.1
## [55] pillar_1.10.1               htmltools_0.5.8.1
## [57] R6_2.6.1                    dbarts_0.9-30
## [59] Rdpack_2.6.2                rprojroot_2.0.4
## [61] evaluate_1.0.3              lattice_0.22-6
## [63] rbibutils_2.3               class_7.3-22
## [65] Rcpp_1.0.14                 svglite_2.1.3
## [67] nlme_3.1-166                xfun_0.51
## [69] pkgconfig_2.0.3
```