



PROJECT MUSE®

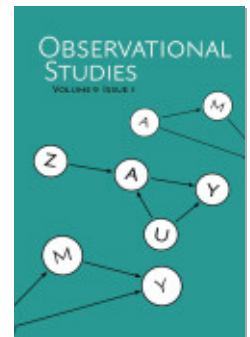
A Visual Diagnostic Tool for Causal Inference

Lucy D'Agostino McGowan, Ralph B. D'Agostino Sr., Ralph B. D'Agostino Jr.

Observational Studies, Volume 9, Issue 1, 2023, pp. 87-95 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2023.0008>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/877890>

🔗 *For content related to this article*

https://muse.jhu.edu/related_content?type=article&id=877890

A Visual Diagnostic Tool for Causal Inference

Lucy D'Agostino McGowan

*Department of Statistical Sciences
Wake Forest University
Winston-Salem, NC, 27109, USA*

mcgowald@wfu.edu

Ralph B. D'Agostino, Sr.

*Department of Mathematics and Statistics,
Boston University
Boston, MA, 02215, USA*

ralph@bu.edu

Ralph B. D'Agostino, Jr.

*Department of Biostatistics and Data Science,
Wake Forest University School of Medicine
Winston-Salem, NC, 27109, USA*

rdagosti@wakehealth.edu

Abstract

Rosenbaum and Rubin (1983) suggested a visual representation, that can be used as a diagnostic tool, for examining whether the relationships between confounders and outcomes are sufficiently controlled, or whether there is a more complex relationship that requires further adjustment. This short commentary highlights this simple tool, providing an example of its utility along with relevant R code.

Keywords: causal inference, data visualization, regression diagnostics

Despite numerous analytical methods available, the most common method for confounding control is covariate adjustment, therefore straight-forward, user friendly diagnostics for determining the adequacy of this adjustment are valuable. The final section of Rosenbaum and Rubin (1983) suggests a visual representation, that can be used as a diagnostic tool, for such an adjustment that we have not seen receive focus in research workflows or pedagogical methods. Specifically, Rosenbaum and Rubin (1983) point out that research studies that implement propensity score methods where the outcome is continuous and the propensity score is estimated using a linear discriminant analysis (\hat{e}_i), there are many useful properties for evaluating the fit of the final outcome model conditional on the propensity score. In this situation, the point estimate for the treatment effect in an ordinary least squares model fit with the adjustment covariates will be exactly equal to one fit adjusting for the propensity score, assuming the same functional form of all covariates is used in the fully adjusted model and the propensity score model. Therefore, a plot of the outcome (y_{0i} and y_{1i}) or residuals ($y_{ti} - \hat{y}_{ti}$) versus the estimated propensity score (\hat{e}_i) will provide a two-dimensional display of the multivariate adjustment. In the case of an ordinary least squares model, the latter is a function of the “residuals versus fits” plot with the key distinction that the points are visually distinguished by treatment arm (for example with a different color or shape or even faceted separately). A simple plot such as this could be a useful diagnostic tool

to assess whether the relationships between confounders and the outcomes are sufficiently controlled, or whether there is a more complex relationship that requires further adjustment, for example non-linear relationships or heterogeneous treatment effects.

Much attention in the literature has been given to the selection of confounders (VanderWeele 2019; VanderWeele and Shpitser 2011; Shrier 2008; Rubin 2008, 2009, 2007; Pearl 2000; Sjölander 2009; D'Amour et al. 2021; Häggström 2018; Greenland and Pearl 2011), however once selected, the functional form of the selected covariates appears to be less emphasized. Even under the assumption that all confounders have been adjusted for in some capacity, if the correct functional form is not used substantial bias can be seen in the estimated treatment effect. Similarly, the presence of non-parallel response surfaces for outcomes within treatment groups or heterogeneous treatment effects can also bias a result if not accounted for. These propensity score residual plots recommended by Rosenbaum and Rubin (1983) can provide one method for evaluating whether the estimated treatment effect is distorted by the presence of non-linear or nonparallel response surfaces. Of note, if the variance of the covariates differs by treatment, covariate adjustment (either directly or via adjustment of the propensity score) without properly accounting for non-linearity will yield substantially biased results (Rubin 1973).

In this variation of the “residuals versus fits” plot, the proposed plot maps the residuals ($y_{ti} - \hat{y}_{ti}$) to the y-axis and the estimated propensity score (\hat{e}_i) to the x-axis. Alternatively, one could plot the “standard” residuals versus fits plot (residuals on the y-axis and fitted values on the x-axis) stratified by treatment such that any differences are more clearly visible in the facets. Below is a simple example of such a plot along with the R code used to create it. The data used for the plot was simulated such that there is a heterogeneous relationship between a summary measure of the confounders (such as the true propensity score), X , and the outcome, Y , based on treatment, T as follows:

$$Y(1) = 0.5 \times X \tag{1}$$

$$Y(0) = -0.5 \times X \tag{2}$$

$$Y_{obs} = T \times Y(1) + (1 - T) \times Y(0) \tag{3}$$

As an illustration, we fit the following ordinary least squares model, and then created the standard residuals versus fits plot (Figure 1). We added a loess line to aid in visualizing any potential relationships.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 T + \hat{\beta}_2 X$$

Notice when examining this standard plot alone there is not an obvious violation of assumptions, if anything there appears to be a non-linear relationship between X and Y , rather than a heterogeneous effect. Using this plot alone could lead to the investigator erroneously adding a non-linear term rather than the appropriate interaction.

Based on the recommendation in Rosenbaum and Rubin (1983), we will create this same figure, stratified by treatment assignment. We can either do this by selecting a different

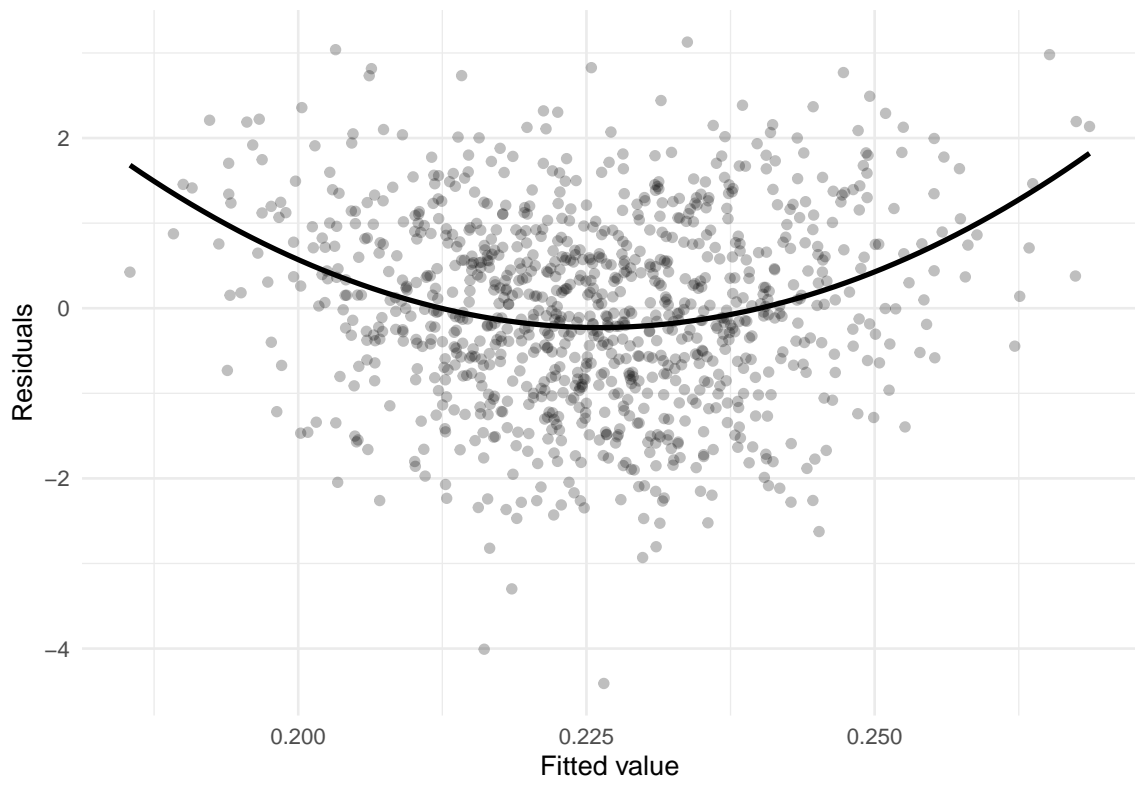


Figure 1: Standard Residuals versus Fitted Plot (Misspecified Model)

color for treatment arm and overlaying the plots (Figure 2) or faceting the plots, creating a separate one for each treatment group (Figure 3).

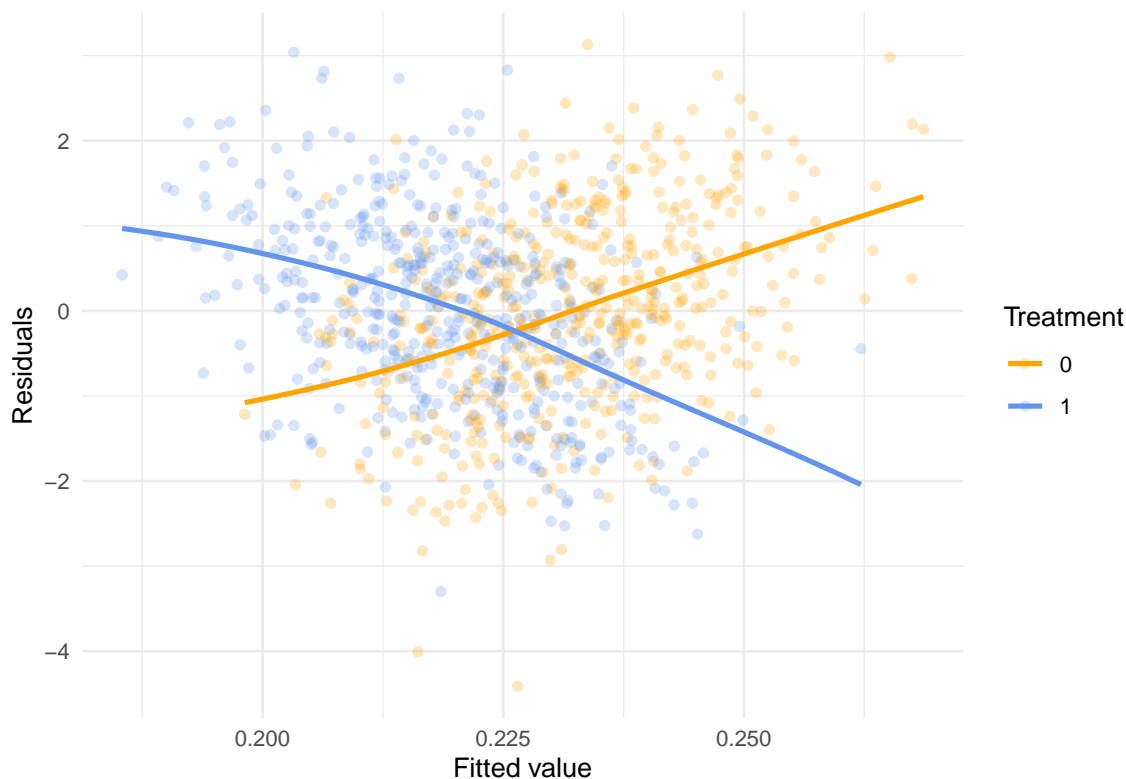


Figure 2: Residuals versus Fits Plot, colored by treatment arm (Misspecified Model)

Observing these figures, the lack of fit is much more obvious and would likely lead the researcher to re-fit the model accounting for the heterogeneous relationship. Additionally, the estimated causal effect for the misspecified model ($\hat{\beta}_1$) is 0, which is a biased estimate, compared to the “true” causal effect, which is dependent on the value of X . We could now fit the following correct model and recreate the stratified residuals versus fits plot (Figure 4).

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 T + \hat{\beta}_2 X + \hat{\beta}_3 T \times X$$

. Summary

When estimating causal effects using covariate adjustment, a sensible and straightforward diagnostic plot to use is the residuals versus fits plot stratified by treatment assignment. Ideally these plots would be generated during the exploratory phase of the modeling process and once in the confirmatory phase the correct relationship between the treatment, confounders, and outcome would be well understood allowing the correct model to be pre-specified. The use of this graphical tool could also be incorporated into routine diagnostics

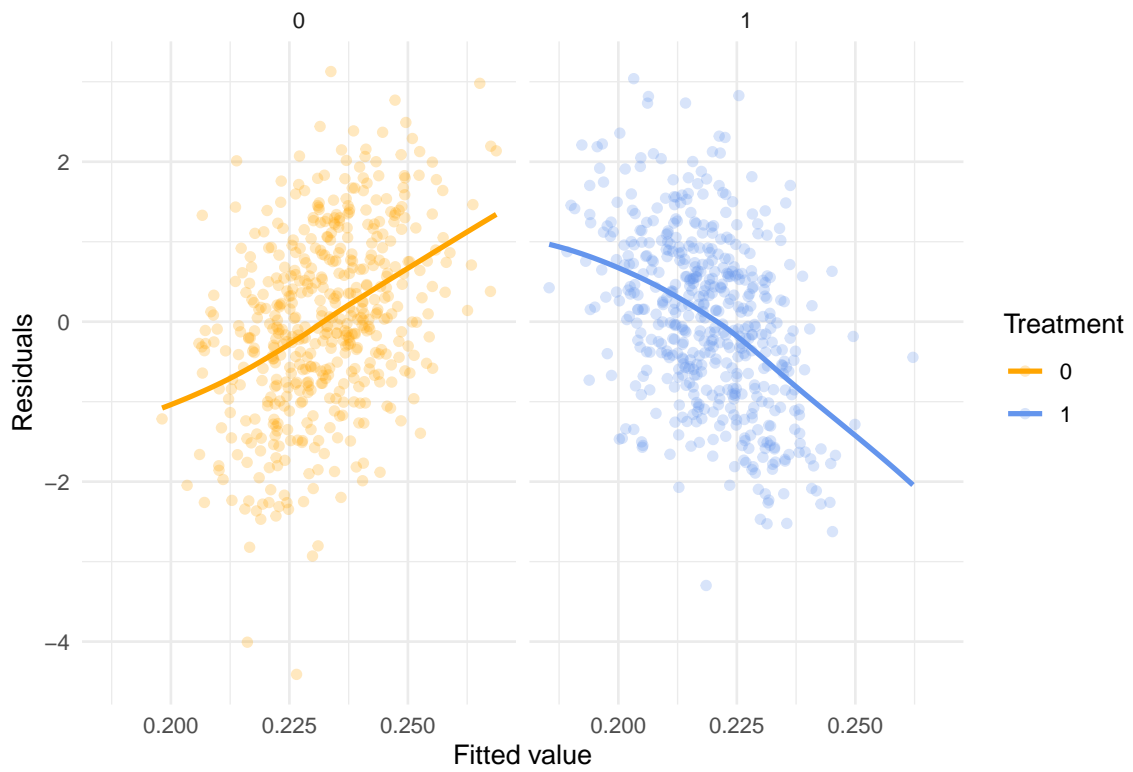


Figure 3: Residuals versus Fits Plot, faceted by treatment arm (Misspecified Model)



Figure 4: Residuals versus Fits Plot, colored by treatment arm (Correct model)

that are used when assessing the performance of propensity score models in statistical inference.

. Code

Below is the R code used to simulate the scenarios described in the paper as well as create the figures.

```
library(tidyverse)
## Simulate data

set.seed(8)
n <- 1000
x <- rnorm(n)
t <- rbinom(n, 1, exp(x) / (1 + exp(x)))
y1 <- 0.5 * x + rnorm(n)
y0 <- -0.5 * x + rnorm(n)
y_obs <- t * y1 + (1 - t) * y0

d <- tibble(
  y = y_obs,
  p = predict(lm(t ~ x)),
  t = as.factor(t),
  f = predict(lm(y_obs ~ t + x)),
  r = resid(lm(y_obs ~ t + x)),
  f_correct = predict(lm(y_obs ~ t*x)),
  r_correct = resid(lm(y_obs ~ t*x))
)

## Create standard residuals vs fits plot on the misspecified model

p_naive <- ggplot(d, aes(x = f, y = r)) +
  geom_point(alpha = 0.25) +
  geom_smooth(
    color = "black",
    method = "loess",
    formula = y ~ x,
    se = FALSE,
    span = 1
  ) +
  theme_minimal() +
  labs(x = "Fitted value",
       y = "Residuals")

p_naive
```



```
## Create residuals vs fits plot colored by treatment arm on the
## misspecified model (Figure 2)
```

```
p2 <- ggplot(d, aes(x = f, y = r, color = t)) +
  geom_point(alpha = 0.25) +
  geom_smooth(
    method = "loess",
    formula = y ~ x,
    se = FALSE,
    span = 1
  ) +
  theme_minimal() +
  scale_color_manual(values = c("orange", "cornflower blue")) +
  labs(color = "Treatment",
       x = "Fitted value",
       y = "Residuals")
```

p2

```
## Create residuals vs fits plot faceted by treatment arm on the
## misspecified model (Figure 3)
```

```
p3 <- ggplot(d, aes(x = f, y = r, color = t)) +
  geom_point(alpha = 0.25) +
  geom_smooth(
    method = "loess",
    formula = y ~ x,
    se = FALSE,
    span = 1
  ) +
  facet_wrap(~ t) +
  theme_minimal() +
  scale_color_manual(values = c("orange", "cornflower blue")) +
  labs(color = "Treatment",
       x = "Fitted value",
       y = "Residuals")
```

p3

```
## Create residuals vs fits plot colored by treatment arm on the
## correct model (Figure 4)
```

```
p4 <- ggplot(d, aes(x = f_correct, y = r_correct, color = t)) +
  geom_point(alpha = 0.25) +
  geom_smooth(
    method = "loess",
```

```

    formula = y ~ x,
    se = FALSE,
    span = 1
) +
theme_minimal() +
scale_color_manual(values = c("orange", "cornflower blue")) +
labs(color = "Treatment",
      x = "Fitted value",
      y = "Residuals")

```

p4

References

- D’Amour, Alexander, Peng Ding, Avi Feller, Lihua Lei, and Jasjeet Sekhon. 2021. “Overlap in Observational Studies with High-Dimensional Covariates.” *Journal of Econometrics* 221 (2): 644–54.
- Greenland, Sander, and Judea Pearl. 2011. “Adjustments and Their Consequences—Collapsibility Analysis Using Graphical Models.” *International Statistical Review* 79 (3): 401–26.
- Häggström, Jenny. 2018. “Data-Driven Confounder Selection via Markov and Bayesian Networks.” *Biometrics* 74 (2): 389–98.
- Pearl, Judea. 2000. “Causality: Models, Reasoning and Inference Cambridge University Press.” *Cambridge, MA, USA* 9: 10–11.
- Rosenbaum, Paul R, and Donald B Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70 (1): 41–55.
- Rubin, Donald B. 1973. “The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies.” *Biometrics*, 185–203.
- . 2007. “The Design Versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials.” *Statistics in Medicine* 26 (1): 20–36.
- . 2008. “Author’s Reply (to Ian Shrier’s Letter to the Editor).” *Statistics in Medicine* 27: 2741–42.
- . 2009. “Should Observational Studies Be Designed to Allow Lack of Balance in Covariate Distributions Across Treatment Groups?” *Statistics in Medicine* 28 (9): 1420–23.
- Shrier, Ian. 2008. “Propensity Scores [Letter to the Editor].” *Statistics in Medicine* 27: 2740–41.
- Sjölander, Arvid. 2009. “Propensity Scores and m-Structures.” *Statistics in Medicine* 28 (9): 1416–20.
- VanderWeele, Tyler J. 2019. “Principles of Confounder Selection.” *European Journal of Epidemiology* 34 (3): 211–19.
- VanderWeele, Tyler J, and Ilya Shpitser. 2011. “A New Criterion for Confounder Selection.” *Biometrics* 67 (4): 1406–13.