# The Impact of Pediatric Leukemia Outbreak on Housing Prices: A Difference-in-Differences Analysis

Luc Chen

## Abstract

This study investigates the causal impact of pediatric leukemia outbreak on housing prices in Churchill County, Nevada, using a difference-in-differences approach. We use a dataset of 10,204 housing transactions from 1990 to 2002 in Churchill and Lyon counties, with Churchill being the treatment group and Lyon being the control group. We find a statistically significant negative effect of leukemia outbreak on housing prices in Churchill County after controlling for housing characteristics. Our results have implications for real estate companies and policymakers, as it highlights the importance of considering disease outbreaks in residential location choices and public health interventions to mitigate their economic impact.

## Introduction

Housing prices are influenced by various factors, including macroeconomic indicators, housing characteristics, and neighborhood attributes (Cohen & Karpavičiūtė, 2017; Wittowsky et al.). The location of a residential property plays a critical role in determining its price, with factors such as accessibility, amenities, and neighborhood quality being key drivers. The presence of environmental health hazards, including disease outbreaks, can also influence housing prices as residents seek to avoid locations with a higher risk of infection. This paper investigates the impact of a pediatric leukemia outbreak on housing prices in Churchill County, Nevada, using a difference-in-differences approach. The study's findings have implications for real estate companies and policymakers, as it emphasizes the importance of considering disease outbreaks in residential location choices and public health interventions to mitigate their economic impact.

## Data description

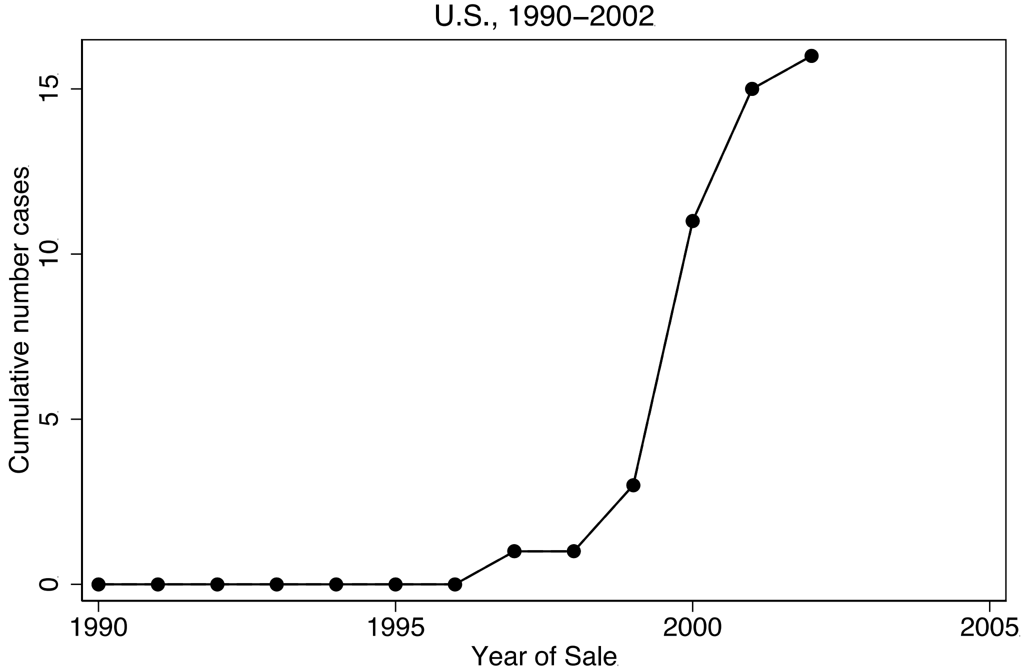
Table 1—Summary statistics for Churchill County and Lyon County

| | Churchill | Lyon |
|---|---|---|
| Mean inflation-adjusted sale prices in dollars | $115500.7 | $122395.4 |
| Mean year of Sale | 1,996 | 1,997 |
| Mean building age | 17.37 | 9.67 |
| Lot Size in Acres | 1.46 | 1.12 |
| House Interior Size in Square Feet | 1,523.07 | 1,503.30 |
| Mean condition Score | 1.89 | 2.39 |
| Observation | 3596 | 6608 |

The dataset comprises a total of 10204 observations that cover twelve years from 1990 to 2002. The dataset includes features County (Churchill or Lyon), year of sales, nominal sales prices, lot size, house interior size, building age, conditional score, and inflated-adjusted sale prices. According to Table 1, Churchill County and Lyon County contain a total observation of 3596, 6608 each. The overall mean inflation-adjusted sales price for two different counties are roughly the same, with 115500 dollars for Churchill County and 122395 dollars for Lyon County. The mean year when the house gets sold is 1996 for Churchill County and 1997 for Lyon County.

Regarding housing characteristics, the mean building age is 17.37 in Churchill County, much higher than in Lyon County, which is 9.67. For house interior size and lot size, houses in Churchill County have 1523.07 $ft^2$ and 1.46 ac respectively, and houses in Lyon County have 1503.3 $ft^2$ and 1.12 ac. The Lyon County houses have a relatively higher mean conditional score of 2.39 compared to those in Churchill County, which is 1.89.

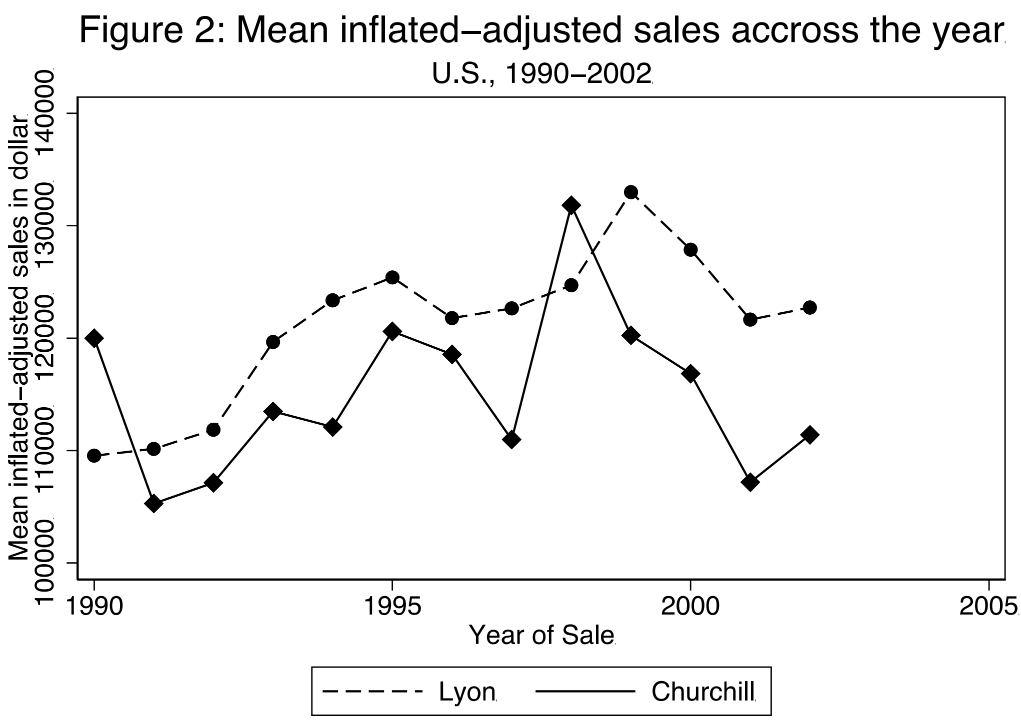
Figure 1: Cumulative Number of Leukemia Cases in Churchill County U.S., 1990–2002




*Notes:* Each data point is the cumulative number of pediatric leukemia cases in Churchill County at the end of each year

Figure 1 illustrates the cumulative number of pediatric leukemia cases in Churchill County from 1990-2002. The data reveals varying collection periods within specific years. For instance, in 2000, the number of cases ranged from 3 to 11, with 3 collected at the beginning of the year and 11 by the end. To enhance data visualization, we selected the highest number of cases for each year, assuming data collection occurred at year-end. It is important to note that each data point in Figure 1 represents the cumulative case count in Churchill County at the close of each year.

The graph indicates no leukemia diagnoses prior to the end of 1996. One case occurred in 1997, followed by two additional cases in 1999. From the end of 1999 to 2000, eight more cases emerged. Background information suggests that before 2000, leukemia received little attention in Churchill County. However, concerns about the county's safety grew starting in 2000. This provides a rough idea that 2000 could be the cutoff point for a disease outbreak, but a more detailed analysis is required, as discussed in the following paragraph with reference to Figure 2.




Our goal is to determine the causal impact of the disease on housing prices while accounting for inflation. Thus, we utilize inflation-adjusted sale prices as the dependent variable to accurately capture housing price differences over the years. Figure 2 presents the general trend of housing prices from 1990 to 2002 in both counties. The dashed and solid lines represent the mean inflation-adjusted house prices for Lyon and Churchill County, respectively, from 1990 to 2002.

To accurately determine the cutoff year, we observe that from 2000 onwards, the slope of Churchill County's line declines more significantly than Lyon County's. This suggests that the leukemia outbreak negatively impacted housing prices. Consequently, we select 2000 as the cutoff point, with 1990-1999 and 2000-2002 as the first and second time periods, respectively.

## Analysis and Result

$$realsales_i = \beta_0 + \beta_1 period_i + \beta_2 group_i + \delta(period_i * group_i) + X'_i\beta_3 + \epsilon_i \qquad (1)$$

We begin by estimating the impact of disease on house price using difference in difference model. $realsales_i$ is the outcome of interest and represents inflated-adjusted house price for $i$ th observations. $period_i$ is a dummy variable that is equal to one if the $i$th measurements of house was sold in post period (2000-2002) and 0 in pre period (1990-1999). $group_i$ is an indicator that equals to 1 if the $i$th measurement of house is in the treatment group (Churchill County) and 0 in control group. $(period * group)_i$ is the interaction term and it calculates the multiplication of the two indicator variables. We have a row vector of controls ($X'_i$) including lot size, house interior size, building age and house score. It's worth to notice that there are missing values in variable condition, but we still include it because it's important variable that contribute to the change in house price, and Stata regress command will automatically drop the row where having missing value. Finally, we have error term $\epsilon_i$ that give us the unobserved factors that may affect the housing price. The reason for adding covariates into the model because we think that house lot size, interior size, age, and conditional score have impact on the outcome housing price. Furthermore, we want to include them for both counties to identify the treatment effect and avoiding omitted variable bias. Also adding covariates can deflate the standard error even though they have no effect on the outcome, which make our statistical test more powerful, and we can determine $\delta$ significant or not more easily.

To validate the difference-in-differences model, we must check the parallel trend assumption, which suggests that the trend for control and treatment groups is roughly the same before the treatment effect occurs (before 2000 in our model). According to Figure 2, the first plot shown earlier indicates the average housing price over the 1990-2002 time period for Lyon and Churchill Counties. We cannot conclude that it violates the parallel trend assumption, even though the two lines diverge before the treatment effect, because we fail to account for sample variability and other housing characteristics that potentially influence the mean inflation-adjusted sales price during the pre-treatment period. However, the trend is roughly the same, and there is no significant difference between treatment and control groups during the pre-period. This supports the validity of our difference-in-differences model.

Table 2—Difference in difference: Estimated effect of disease on house price

| realsales | Coefficient | Std. err. | t-value | p-value | 95% Conf. | Interval | Sig |
|---|---|---|---|---|---|---|---|
| county | 8663.207 | 1012.305 | 8.56 | 0 | 6678.88 | 10647.535 | *** |
| time | -4312.62 | 945.902 | -4.56 | 0 | -6166.784 | -2458.457 | *** |
| DID | -15577.587 | 1759.408 | -8.85 | 0 | -19026.39 | -12128.783 | *** |
| acres | 730.597 | 48.663 | 15.01 | 0 | 635.209 | 825.986 | *** |
| sqft | 61.074 | 1.031 | 59.25 | 0 | 59.053 | 63.094 | *** |
| age | -135.513 | 24.374 | -5.56 | 0 | -183.29 | -87.736 | *** |
| condition | 25254.4 | 991.421 | 25.47 | 0 | 23311.01 | 27197.791 | *** |
| Constant | -27984.789 | 2175.294 | -12.86 | 0 | -32248.816 | -23720.763 | *** |

| | | | |
|---|---|---|---|
| Adj R-squared | 0.5219 | Number of obs | 9773 |
| F-test | 1525.138 | Prob > F | 0.000 |
| Akaike crit. (AIC) | 232645.482 | Bayesian crit. (BIC) | 232702.981 |

**** p<.01, ** p<.05, * p<.1*

Table 3—Difference in difference estimators table

| | 1990-1999 | 2000-2002 | Difference |
|---|---|---|---|
| Lyon County | -27984.789 | -32,297.409 | -4312.62 |
| | (2175.294 ) | (2306.098) | (945.9018) |
| Churchill County | -19,321.582 | -39,211.789 | -19,890.207 |
| | (1954.52) | (2487.399) | (1501.342) |
| Difference | 8663.207 | -6,914.38 | -15577.587 |
| | (1012.305) | (1537.146) | (1759.408) |

*Noted:* Estimated standard errors are displayed in parentheses.

To evaluate our model's performance, we note that the adjusted R-squared is 0.5219 from Table 2. The model explains 52.19% of the variability observed in our dependent variable. The p-value for the F-statistic is 0, indicating statistical significance. The model variables are jointly significant and provide a better fit to the data than an intercept-only model. This leads us to conclude that our model performs fairly well in fitting the data.

According to Table 2, we can tell that all the parameters are statistically significant with p value smaller than 1 percent significant level. Combining with Table 3, we can say that between 1990-1999 time period and 2000-2002 time period, when holding housing characteristcs constant, the mean inlfation-adjusted house price decrease by 4312.62 and 19,890.207 dollars at Churchill County and Lyon County respectively. During 1990-1999 time period, the mean inflation-adjusted house price in Churchill County is 8663.207 higher than Lyon county, and 6914.38 dollars lower during 2000-2002 time period when holding housing characteristcs constant. Given that estimated coefficient $\delta$ is -15577.587, I interpret it as holding housing characteristcs constant for two different counties, the average differential change in inflation-adjusted house price from the first (1990-1999) to the second (2000-2002) time period of Churchill county relative to the Lyon County is -15555.6 dollars. The difference in difference estimators shows there is a significant drop of housing price in Churchill County when leukemia was drawn into people's attention. It intuitively makes sense because people from other counties or states are concerned about the infection of desease, so they get intimidated and have no incentive to buy a house in Churchill County. A lack of demand for housing in Churchill County in turn contributes to a drop in housing price. Furthermore, as many residents live in Churchill County also consider leaving to avoid getting infection, the supply of houses increases considerably and drags down the housing price.

## Conclusion

In summary, we first identified the year 2000 as our cutoff period for the treatment effect by examining the cumulative number of pediatric leukemia cases in Churchill County during the

1990-2002 period and the trend of housing prices from 1990 to 2002 for both Churchill and Lyon Counties. We then investigated the causal impact of disease on housing prices by visualizing the trend of housing prices from 1990 to 2002, which suggested that after 2000, with the outbreak of pediatric leukemia, housing prices in Churchill County dropped significantly compared to Lyon County. To obtain more rigorous results, we applied the difference-in-differences model and used a linear trend model to validate the parallel trend assumption before the treatment effect. Ultimately, the results from the difference-in-differences model suggest that the increase in pediatric leukemia cases in Churchill County has a statistically and economically significant negative effect on housing prices.

Our findings have broad implications. In times of widespread disease, such as during the COVID-19 pandemic, people consider the prevalence of viruses in an area as a factor when choosing a place to live. This, in turn, can lead to a decrease in housing prices. Real estate companies should be aware of the potential impact of diseases on regional housing prices to optimize site selection and maximize their profits..

## References

Cohen, Viktorija Cohen, and Lina Karpavičiūtė. "(PDF) the Analysis of the Determinants of Housing Prices - Researchgate." *The Analysis of the Determinants of Housing Prices*, https://www.researchgate.net/publication/314143042_The_analysis_of_the_determinants_of_housing_prices.

Hoekveld, Josje, et al. "Residential Housing Prices: Impact of Housing Characteristics, Accessibility and Neighbouring Apartments – a Case Study of Dortmund, Germany." *Residential Housing Prices: Impact of Housing Characteristics, Accessibility and Neighbouring Apartments – a Case Study of Dortmund, Germany*, https://www.tandfonline.com/doi/full/10.1080/21650020.2019.1704429.