Neonic Pesticides and Honey Bees

Cari Comnick and Annie Shapiro

May 25, 2018

Introduction

This case study seeks to examine the effects of neonic pesticides on the production of honey from honey bee colonies in the United States. Compared to other pesticides, neonics (also known as neonicotinoids), are reported to cause less toxicity in birds and mammals than insects. As of 2013, neonics have been used on about 95 percent of corn and canola crops, the majority of cotton, sorghum, and sugar beets, and about half of all soybeans in the US. They are used as seed coatings on most corn and soybean seeds. Neonics are also used on the vast majority of fruit and vegetables, from apples, cherries, peaches, oranges, berries, leafy greens, tomatoes, and potatoes, to cereal grains, rice, nuts, and wine grapes. They were developed in the early 1990s and became widely used across the United States by the early 2000s. Imidacloprid, a specific type of neonic pesticide examined in this study, is currently the most widely used insecticide in the world.

Neonic pesticides are quite controversial, as they have been linked in a range of studies to adverse ecological effects, including honey-bee colony collapse disorder (CCD) and loss of birds due to a reduction in insect populations. However, findings on this issue have been conflicting, and thus the pesticides have not yet been directly linked to the decline in the honey bee population. In March 2012, the Center for Food Safety, Pesticide Action Network, Beyond Pesticides and a group of beekeepers filed an emergency petition with the Environmental Protection Agency asking the agency to suspend the use of clothianidin, another type of neonic that was examined in this study, for the sake of protecting the bee populations. The agency denied the petition. In 2013, the European Union and a few non EU countries restricted the use of certain neonicotinoids and in 2018, the EU fully banned the three main neonics (clothianidin, imidacloprid and thiamethoxam, all three of which were examined in this study) for all outdoor uses. Several states in the United States, including Minnesota in 2016, have began to restricted usage of neonicotinoids out of concern for pollinators and bees.

Methods

The goal of our analysis is to observe changes in honey production, controlling for variation from different states and years, and consider if these changes are associated with pesticide use. This is a longitudinal study. The elements of this study are time and the primary sampling units are the individual states. The response variable is the total honey production in each state in each year.

This data is from the US Geological Survey (USGS) and the US Department of Agriculture (USDA), merged by a user on Kaggle. It combines state-level pesticide data with data about honey production and honeybee colonies. This is a longitudinal data set, with information on years 1998 to 2017. However, as there was no pesticide data recorded for 2016 and 2017, we omit these years and thus our data set spreads from 1998 to 2015. Year was transformed to be the number of years into the study, so 1998 became year 0, and so on. The data set did not include Alaska, Delaware, Maryland, Connecticut, New Hampshire, or Massachusetts. Additionally, we omitted the state of Hawaii since it did not report any pesticide levels; this left us with with 43 states to analyze. After this cleaning, our data set consisted of 727 observations and 16 variables. 5 states - South Carolina, Maryland, Oklahoma, Nevada, and New Mexico - were lacking observations from various years; however, this is not concerning for our analysis.

Variables about honey production include the number of honey-producing colonies, the average honey yield per colony (in pounds), the total production of honey in pounds per state per year (in pounds, divided by 10,000), the honey stocks - a product of honey production - held by producers (in pounds), the average price per pound in each year and state (in dollars), and the production in dollars, which is the total honey production times the price per pound. As many of these variables are highly correlated, we carefully considered which were independent enough to be necessary in our model. We chose to include only the price per pound variable, as this would control for outside market effects.

The data set also includes pesticide data for 5 different neonic pesticides. Each pesticide is recorded as the amount in kilograms used each year in each state. The pesticides investigated in this study are clothianidin, imidacloprid, thiamethoxam, acetamiprid, and thiacloprid applied at each location and time. An additional variable of the sum of all 5 of the pesticides used per year per state was also provided. As this variable is clearly correlated with others and we wanted to see if specific pesticides were associated with honey production, it was not included in analysis. For each pesticide, we noted a high amount of 0's reported. This is because until 2003, pesticides of this type were less commonly used, and each state did not use every of pesticide examined in this study, depending on what its major crops were.

As all variables but year were heavily right skewed, log transformations were taken for all predictors (except for year). Because all pesticides had a high number of 0's, we added 1 to each of these observations so that it still remained 0 after we took the log. See the appendix for further discussion of this.

Results

Below are the summary statistics for the variables in our dataset.

Table 1: Summary Statistics

Number of Colonies (in 1000s)	Yield Per Colony (lbs)	Total Honey Production (10,000s of lbs)
Min. : 2.0	Min. : 19.00	Min. : 8.4
1st Qu.: 9.0	1st Qu.: 46.00	1st Qu.: 46.8
Median : 27.0	Median : 58.00	Median : 157.5
Mean : 62.6	Mean : 60.03	Mean : 422.8
3rd Qu.: 65.0	3rd Qu.: 71.00	3rd Qu.: 424.4
Max. :510.0	Max. :128.00	Max. :4641.0

Table 2: Summary Statistics

Stocks (lbs)	Price Per Pound (Dollars)	Product Value (10,000s of Dollars)	Clothianidin (kg)
Min. : 8	Min. :0.490	Min. : 16.2	Min. : 0.3
1st Qu.: 125	1st Qu.:1.020	1st Qu.: 85.3	1st Qu.: 1.0
Median : 437	Median :1.440	Median : 204.6	Median : 372.6
Mean : 1298	Mean :1.626	Mean : 542.6	Mean : 11339.9
3rd Qu.: 1454	3rd Qu.:2.020	3rd Qu.: 556.7	3rd Qu.: 7085.9
Max. :13800	Max. :5.530	Max. :8385.9	Max. :278498.8

Table 3: Summary Statistics

Imidacloprid (kg)	Thiamethoxam (kg)	Acetamiprid (kg)	Thiacloprid (kg)	All Neonic Pesticides (kg)
Min. : 3.2	Min. : 0.30	Min. : 0.1	Min. : 0.1	Min. : 3.2
1st Qu.: 924.9	1st Qu.: 25.05	1st Qu.: 1.0	1st Qu.: 1.0	1st Qu.: 1607.4
Median : 3713.1	Median : 1153.90	Median : 12.2	Median : 1.0	Median : 8561.6
Mean : 10131.1	Mean : 6367.28	Mean : 731.5	Mean : 122.0	Mean : 28690.1
3rd Qu.: 10623.3	3rd Qu.: 7981.05	3rd Qu.: 314.6	3rd Qu.: 1.0	3rd Qu.: 34681.4
Max. :150569.3	Max. :64834.60	Max. :36480.3	Max. :4273.2	Max. :403011.6

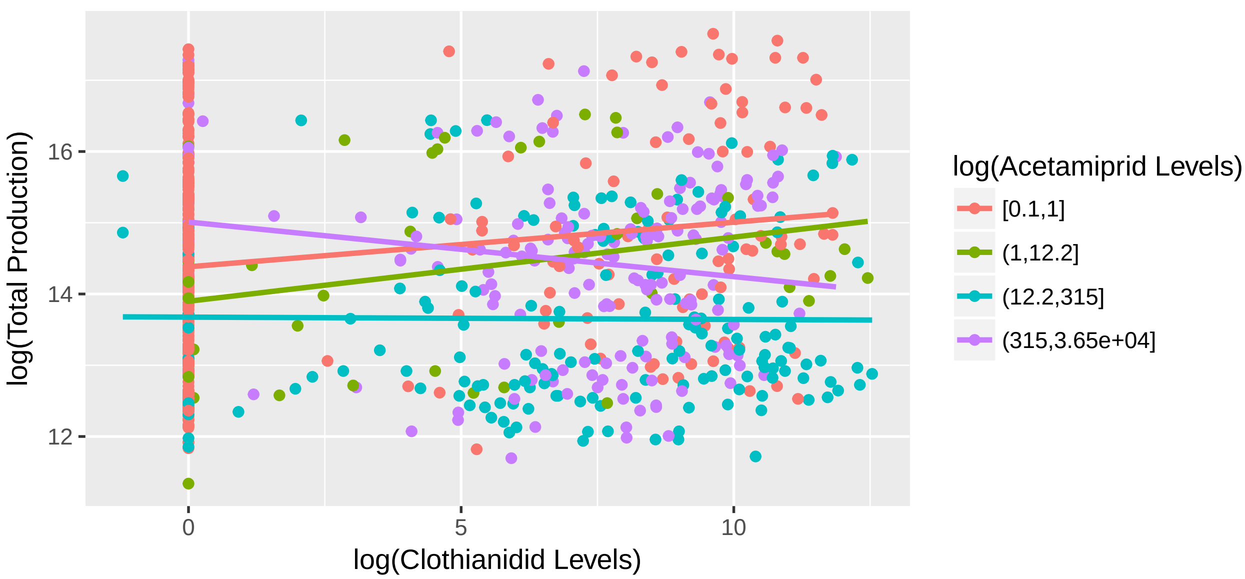
The following are some trends we observed within our exploratory data analysis. North Dakota and California tended have the highest number of colonies as well as total production, while Mississippi and Louisiana have most of the top yield per colony values. South Carolina and Kentucky have the lowest numbers of stocks, while South Dakota has the most. The Midwest states tend to have the higest rates of clothianidin use, while California tends to use the most imidacloprid, followed by Midwest states. The Midwest and Texas use the most thiamethoxam, and California uses the most acetamiprid by far. New York and Washington have the highest thiacloprid usage. Finally, overall, the Midwest states have the highest total use of neonics.

Through modeling, we found the following covariates to be significant in predicting the log of the total production: log of price per pound, year since the study began, log of the weight of clothianidin used, log of of the weight of acetamiprid used, an interaction term between the log of price per pound and year, and an interaction between the log of acetamiprid and the log of clothianidin.

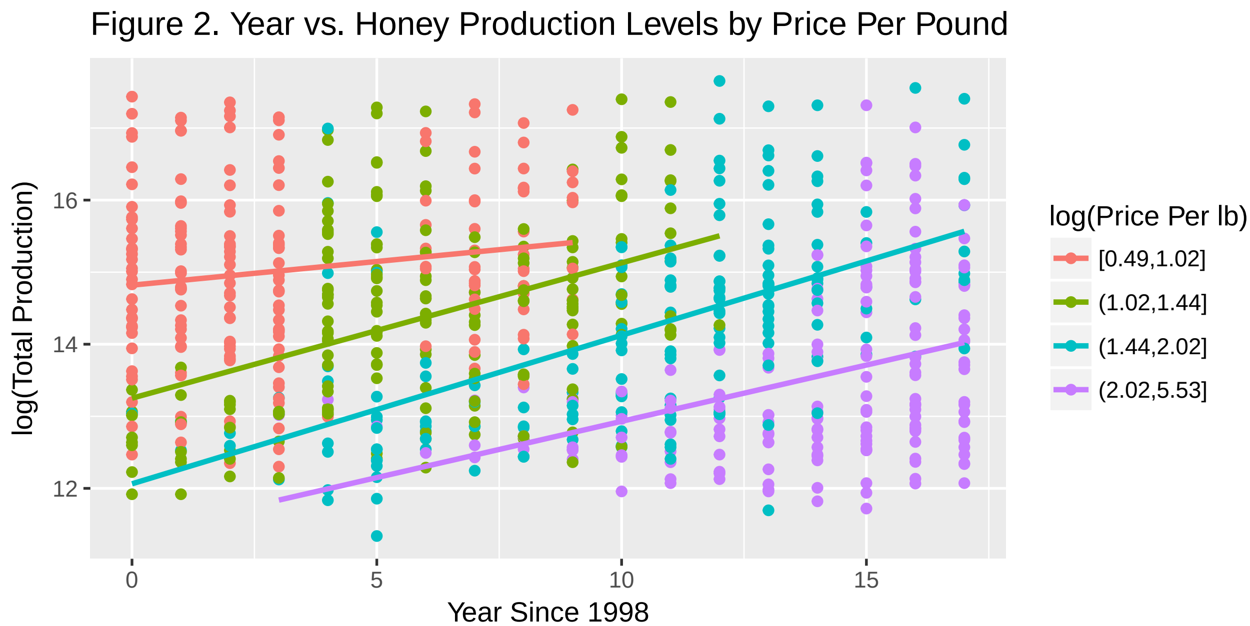
Additionally, a random intercept was added for state and a random slope was added for year to account for variance from those two factors. 97.5% of the variation in the model comes from the state level in the first year of the study, 0.0005% of the variation comes from rates of change in the total production over the 17 year observation period, and 2.5% of the variation comes from within-state deviations. Despite such little variation coming from time, a p-value of 0 confirms that the random effect for time is still necessary.

In Figure 1, we see the interaction between acetamiprid and clothianidid. Because we see different slopes among groups of acetamiprid levels when plotting clothianidid levels against the log of total production, we suspect an interaction between clothianidid and acetamiprid.

Figure 1. Clothianidid levels vs. Honey Production Levels by Acetamiprid Levels



In Figure 2, we also see a potential interaction between price per pound of honey and the number of years since 1998. We see that the slopes of best fit lines for the four categories of price per pound we created differ when plotted against the log of total honey production, especially at the low prices. When creating our model, we included these interactions as well as others suggested by EDA, although the two shown above were the only two that proved to be significant.



Below is our final model. We use a two-level model with state i and year j:

Level 1:

$$\log(TotalProduction_{ij}) = a_i + b_i yrsince1998_{ij} + \gamma \log(priceperlb_{ij}) + \delta \log(clothianidin_{ij}) +$$
$$\lambda \log(acetamiprid_{ij}) + \eta \log(clothianidin_{ij}) \log(acetamiprid_{ij}) + \theta \log(priceperlb_{ij}) yrssince1998_{ij} + \epsilon_{ij}$$

Level 2:

$$a_i = \alpha + u_i$$
$$b_i = \beta + v_i$$

Composite:

$$\log(TotalProduction_{ij}) = \alpha + \beta yrsince1998_{ij} + \gamma \log(priceperlb_{ij}) + \delta \log(clothianidin_{ij}) + \lambda \log(acetamiprid_{ij}) +$$
$$\eta \log(clothianidin_{ij}) \log(acetamiprid_{ij}) + \theta \log(priceperlb_{ij}) yrssince1998_{ij} + u_i + v_i yrsince1998_{ij} + \epsilon_{ij}$$

Where $\epsilon \sim N(0, \sigma)$ and

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix}^T \sim N(\vec{0}, \Sigma)$$

and

$$\Sigma = \begin{bmatrix} \sigma_1 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2 \end{bmatrix}$$

Our estimates for the fixed effects are summarized below:

Parameter	Estimate	Standard Error	t-Statistic
α	5.201	0.210	24.800
β	-0.029	0.008	-3.621
γ	-0.184	0.056	-3.282
δ	0.005	0.005	1.005
λ	0.006	0.007	0.865
η	-0.003	0.001	-2.809
θ	0.016	0.005	3.390

The 95% confidence intervals for the estimates:

Parameter	2.5%	97.5%
α	4.786	5.617
β	-0.045	-0.013
γ	-0.296	-0.075
δ	-0.004	0.014
λ	-0.008	0.011
η	-0.005	-0.001
θ	0.007	0.026

Our estimates for the variance components:

Parameter	Variance	Standard Error
σ_1	1.851	1.361
σ_2	0.001	0.031

Finally, our estimate for the correlation of the random effects is $\rho = .12$.

Interestingly, we found that most of the neonic chemicals were not significant in our model. We saw a significant interaction between kilograms of clothianidin and acetamiprid, although neither chemical on its own proved significant. This is seen in the confidence intervals for each coefficient. The confidence interval for both log(clothianidin) and log(acetamiprid) contain 0; however, their interaction has a negative effect on the log of total production of honey.

Individually, year since 1998 and price per pound both have negative coefficients, suggesting that an increase in either one results in a decrease in the total production of honey, but their interaction has a positive coefficient. This makes it difficult to interpret the the effect of each covariate generally for the model. However, in fixing one at a specific number we can see how the other would affect the total production of honey.

Given that neonic pesticide use increased in the US around 2003, we wanted to see the effect of year and its interaction with price per pound in 2002, 2003, and 2004. Holding pesticide use constant, we find that in 2002 (4 years since 1998), a doubling of the price per pound is associated with a $2^{-.18}2^{.016*4} = .923$ multiplicative change (or an 7.7% decrease) in the median number of bee colonies. Similarly, for 2003, this doubling is associated with a 6.7% decrease and in 2004, it is associated with a 5.6% decrease.

For the interaction between clothianidin and acetamiprid, we observe what happens if we fix the level of one pesticide and double the amount of the other one. We chose to fix each at its most common logged value, 9 for clothianidin and 7 for acetamiprid. Doubling acetamiprid while fixing log(clothianidin) at 9 results in a 1.2% decrease in the median number of colonies. Doubling clothianidin while fixing log(acetamiprid) at 7 results in a 0.97% decrease in the median number of colonies.

It should be noted that when diagnostics were checked, a few minor outliers and influential points were found. However, as each type of diagnostic checked showed different outliers or influential points (none were flagged among multiple diagnostics), no points were taken out. See figures 3 and 4 in the appendix and further explanation.

Discussion

This study finds the following to be significant in predicting the total production of honey after accounting for variance from time and different states: year, price per pound of honey, and interactions between year and price and the amount of acetamiprid and clothianidin used. We set out to determine the effects of neonic pesticides on total honey production, but our analysis points to price per pound and year as being more important factors. It is not entirely surprising that our results do not provide a clear answer to our question, as earlier studies have also found conflicting results. Future research might focus on using different models for each state, since the effects of pesticides could differ based on which state they are used in.

Although price per pound is not a factor directly linked to the production of honey, it is indicative of the amount of honey being produced at any given time. As the principle of supply and demand suggests, when there are limited goods to be sold, their price increases. When there is less honey, we would expect the price to increase. Thus, price makes sense as a predictor of levels of honey production.

There is one limitation to this study. It is well-known that climate change, particularly the rising global temperatures, are related to the decline in bee populations in the past two decades. Declining bee populations would lead to a decline in the total production of honey. This study examined only bee colony levels in relation to neonic pesticide use, not accounting for these climate factors. Further research done controlling for these other variables would be a necessary next step in fully understanding the effects of these pesticides on the numbers of bees and the amount of honey being produced. A potentially interesting approach could be to compare the numbers of honeybees between countries that support the use of neonics and those that do not.

References

https://www.kaggle.com/kevinzmith/honey-with-neonic-pesticide

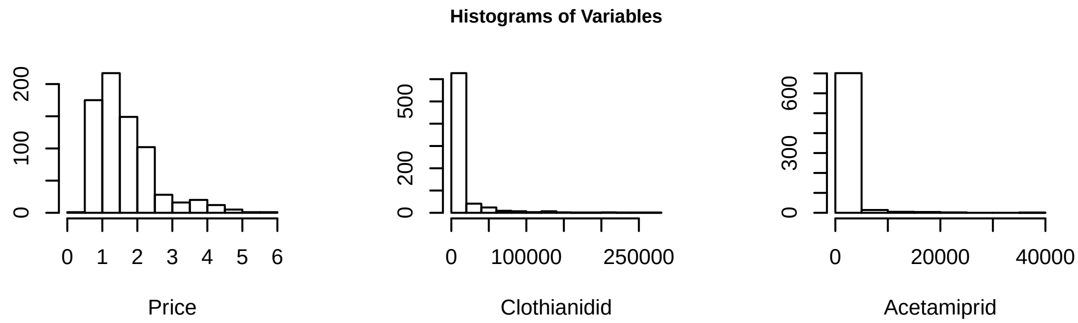
https://www.npr.org/sections/thesalt/2016/08/31/491962115/minnesota- cracks-down-on-neonic-pesticides-promising-aid-to-bees

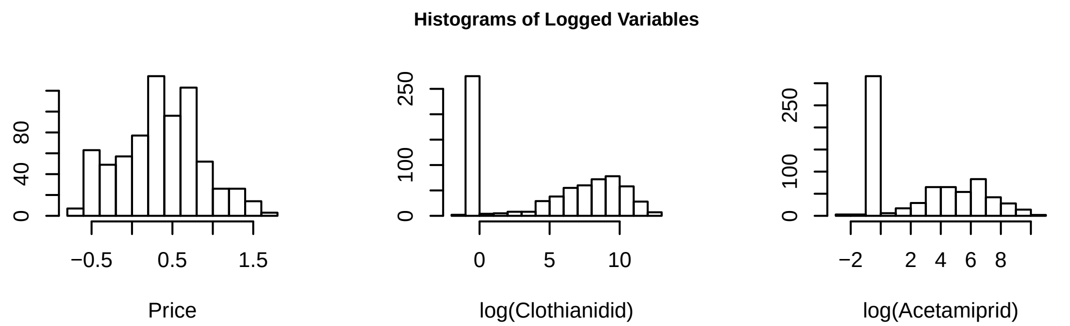
https://en.wikipedia.org/wiki/Neonicotinoid#cite_note-10

Appendix

Logged Variables

As mentioned above, we took the log of each chemical and other variables, as well as added 1 to each 0 obeservation to make it possible to do so. Below are histograms illustrating these transformations.





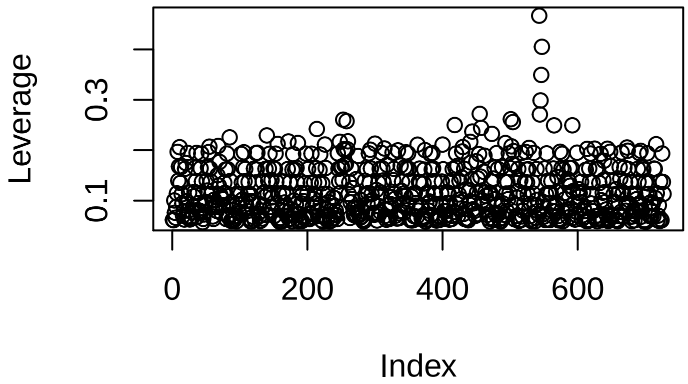
We can see that while logging the variables did make them less skewed and more normal in the mid-range values, we have huge numbers of zeros for each chemical. This is potentially an issue in our model building, and there perhaps may be a model out there that could better account for the high number of zero observations, as well as the very wide range in the weight of chemicals used.

Outliers and Influential Points

Diagnostic plots showed some potential outliers or influential points in our model. However, as no observation was repeatedly flagged in more than one diagnostic and none of the points flagged had any clear relation to each other, we kept them in the model as our data set is already not very large.

Plotting the leverage of each point, we see two more points flagged as potentially influential. These two points were not flagged in either of the previous diagnostics nor did they have unusual numbers recorded for their observations.

Figure 3. Leverage Plot



Interestingly, we also see in the state level leverage plot, South Carolina and 4 other states tend to show up on the leverage plots as high, likely because they are missing some years of data as mentioned earlier.

Figure 4. Leverage Plot − State Level