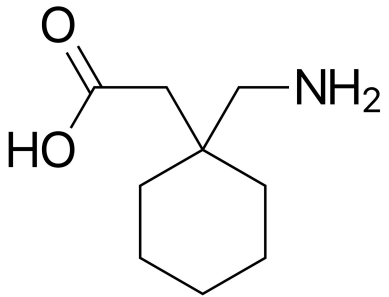
“Gabbies” on Streets :
Data Analysis on Street Price of Gabapentin

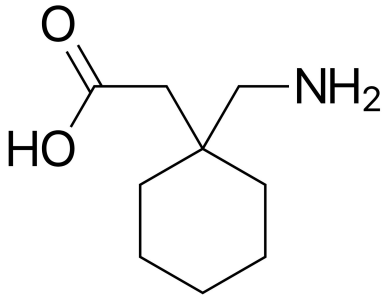
Group 5: Yunran Chen, Lingxi Song, Siqi Fu

Introduction: Gabapentin

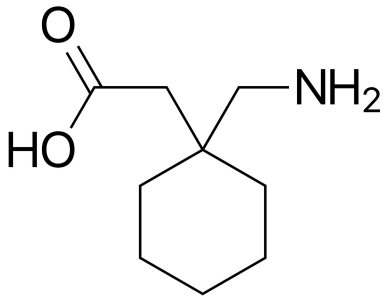


- is used to treat certain types of seizures and nerve pain.
- is first approved for use by FDA in 1993 .
- has been approved as a **generic** medication in the USA on **September 23th, 2014**.
- is cheap, and ranked 4th **popular** prescribed medication in USA
- Has euphoric effects -- being abused for **getting high**

Introduction: Gabapentin



- is used to treat certain types of seizures and nerve pain.
- is first approved for use by FDA in 1993 .
- has been approved as a **generic** medication in the USA on **September 23th, 2014**. -- I(policy)
- is ranked 4th **popular** prescribed medication in USA -- stable price
- Has euphoric effects -- being abused for **getting high** -- data preparation

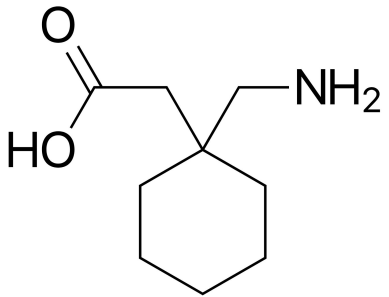


Goal: what influence the price ?

- Which **factors** have potential effect (source, reason, bulk, dosage strength)?
- Whether there are **regional** variance? (city, state, USA region)
- Whether there is a **trend** over time or seasonal effect? (date, quarter of purchase)
- Whether a **policy** takes effects?
- Where can I buy the cheapest Gabbies (**Rank states by the price**)?

Dataset: StreetRx

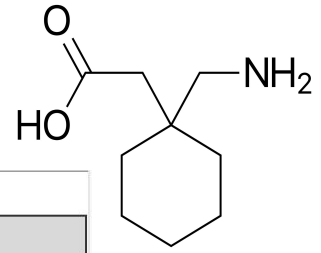
- **1578** self-reported street price (per mg) records by anonymous users from 2013/02/16 to 2019/3/30 in the **United States**



Data preparation

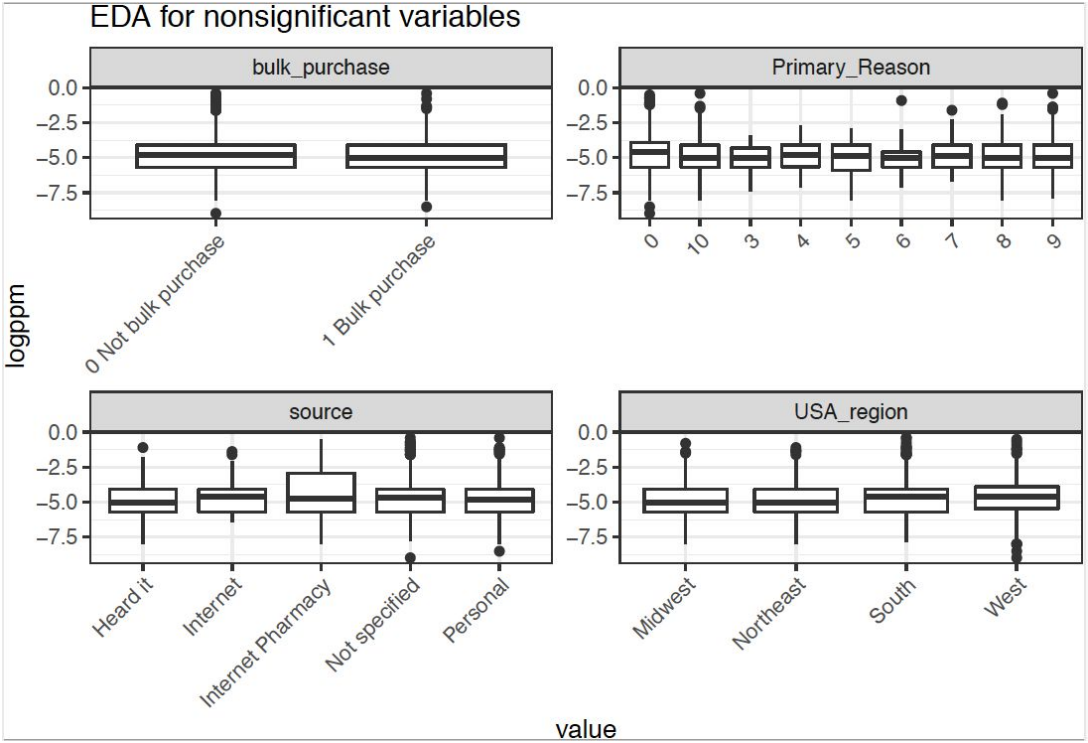
- **Missing value**
  - **Exclude “city”** since it contains 40% missing value and provides similar info as state.
  - Imputed “Not specified” for NA in `Source`.
  - Imputed “Not answered” for NA in `Primary Reason`.
- **Unreasonable Records**
  - Exclude records with PPM <= 0
  - Exclude records with `Primary Reason` as “calm down”.
- **Transformation**
  - Apply **logarithm** to PPM, denote as “logppm”.
- **Combine Categories**
  - Combined "Usatoday.com", "Reddit.com" ,"Reddit", "Quora" records in the `Source` column into "Internet" since they are similar.

EDA: factors without significant influence



Bulk purchase,
reason of purchase,
source,
USA region

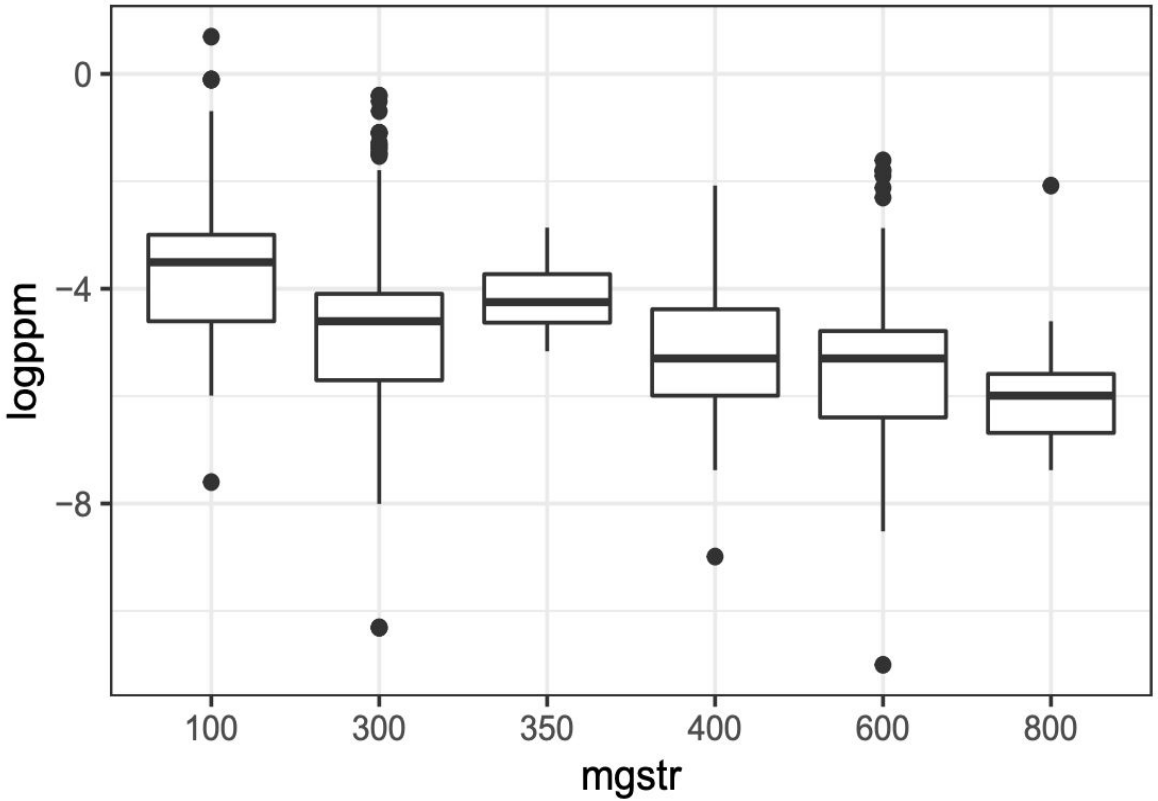
do not have significant effect on price.



EDA: dosage strength has significant influence

The dosage strength in mg of the units

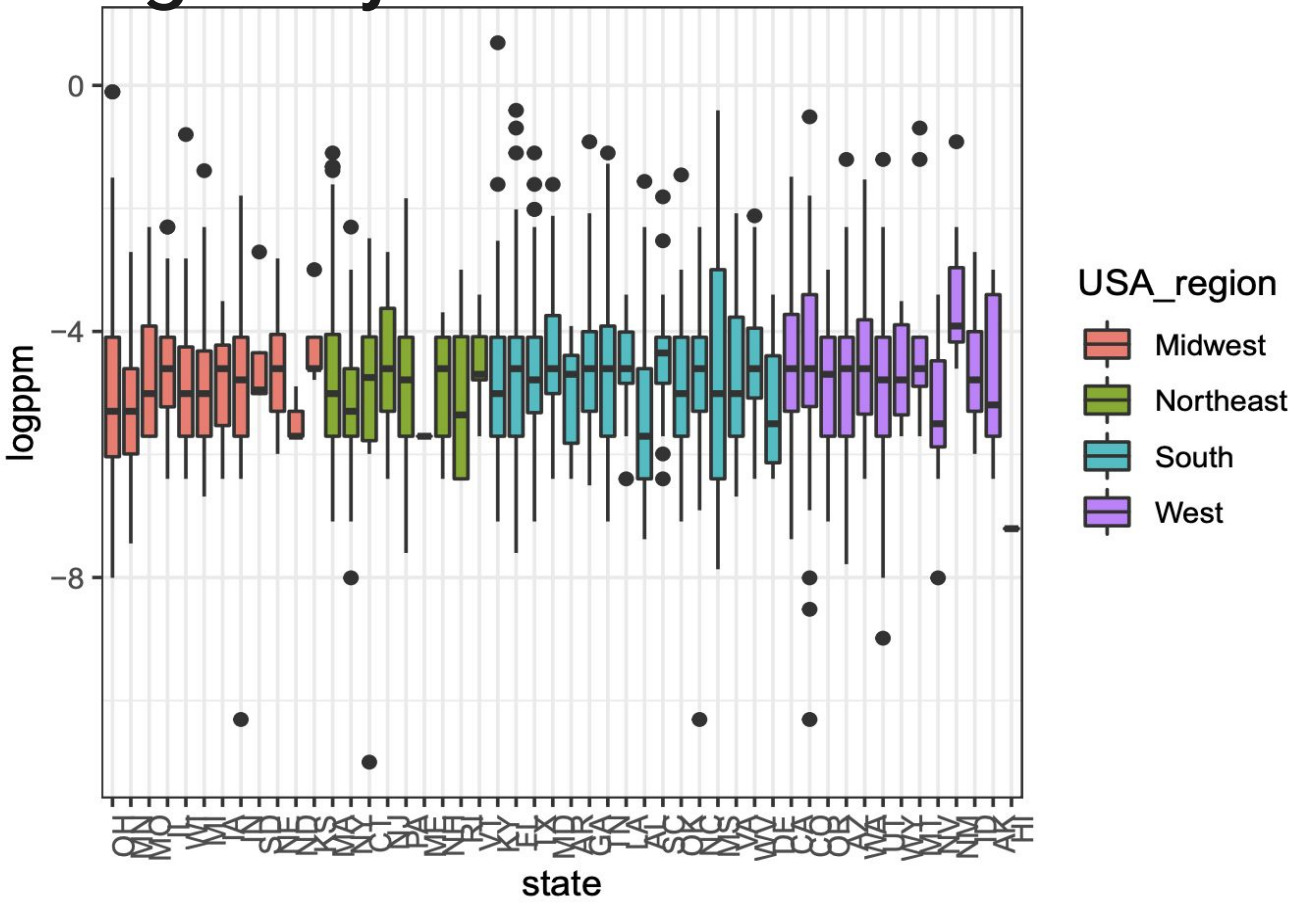
has a **nonlinear** effect (-log).



EDA: Regional heterogeneity exists

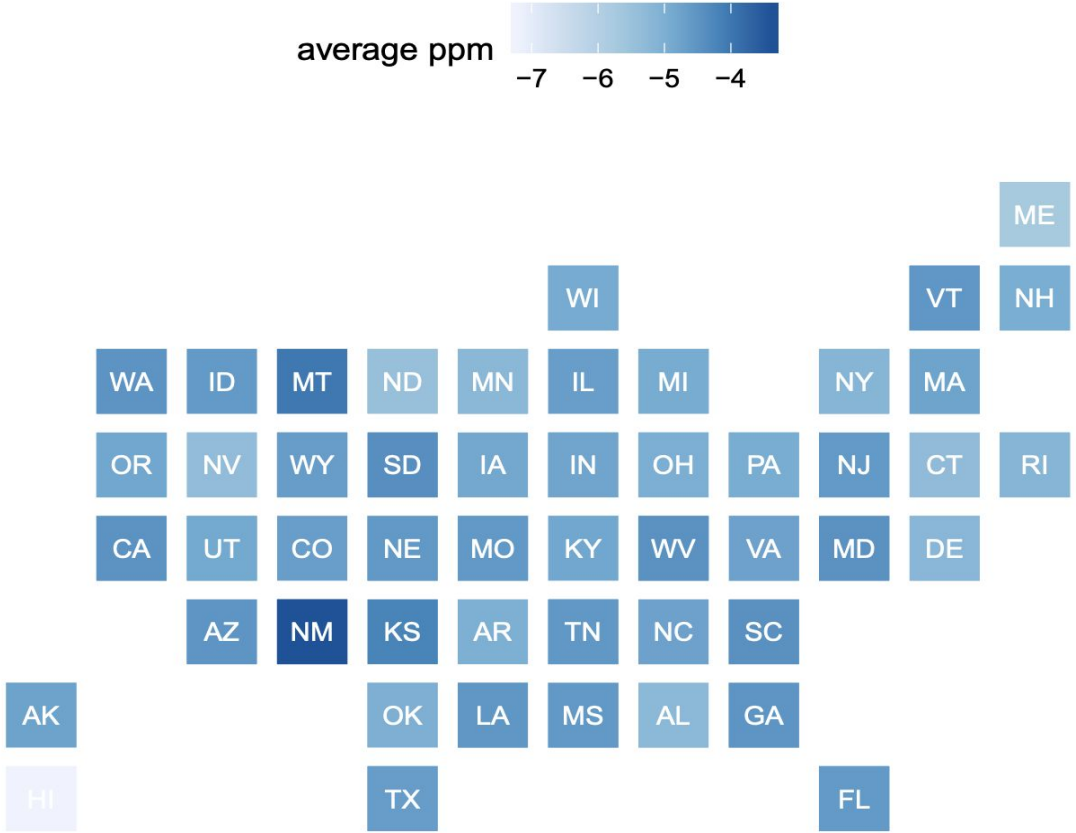
Heterogeneity across states

exists.



EDA: Regional heterogeneity exists

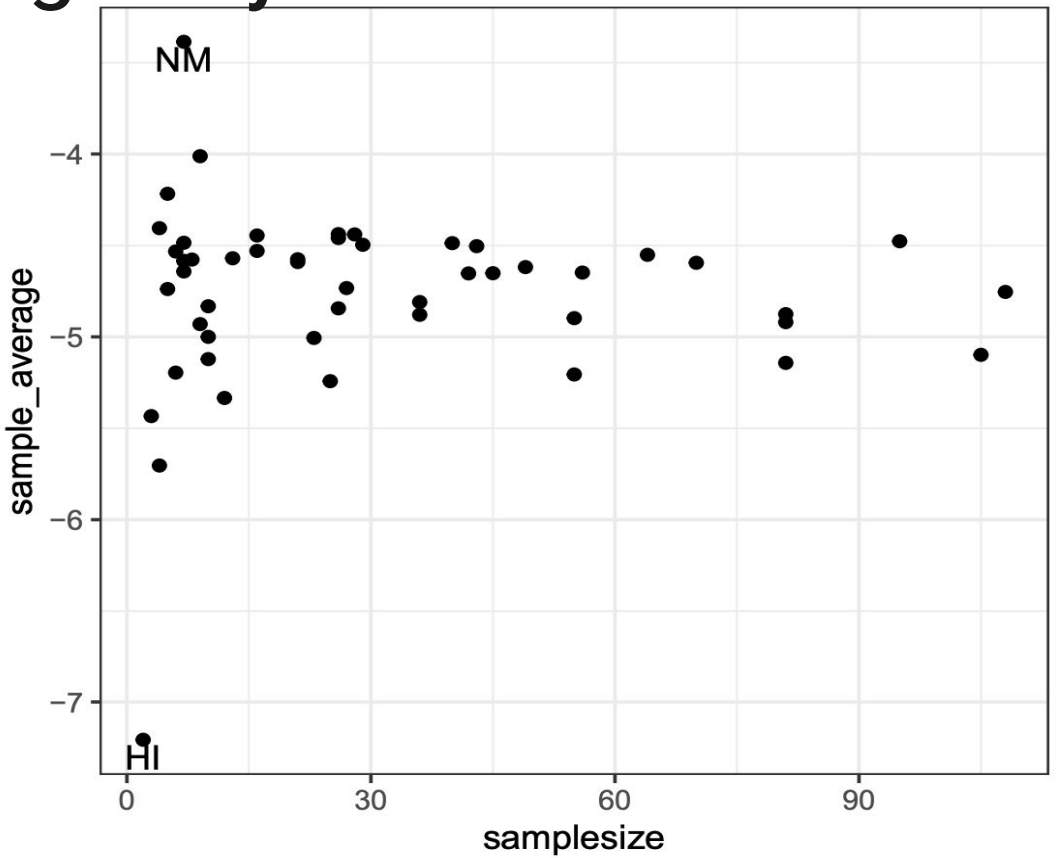
No significant connection among states **geographically**



EDA: Regional heterogeneity exists

Less sample size tends to have larger variance.

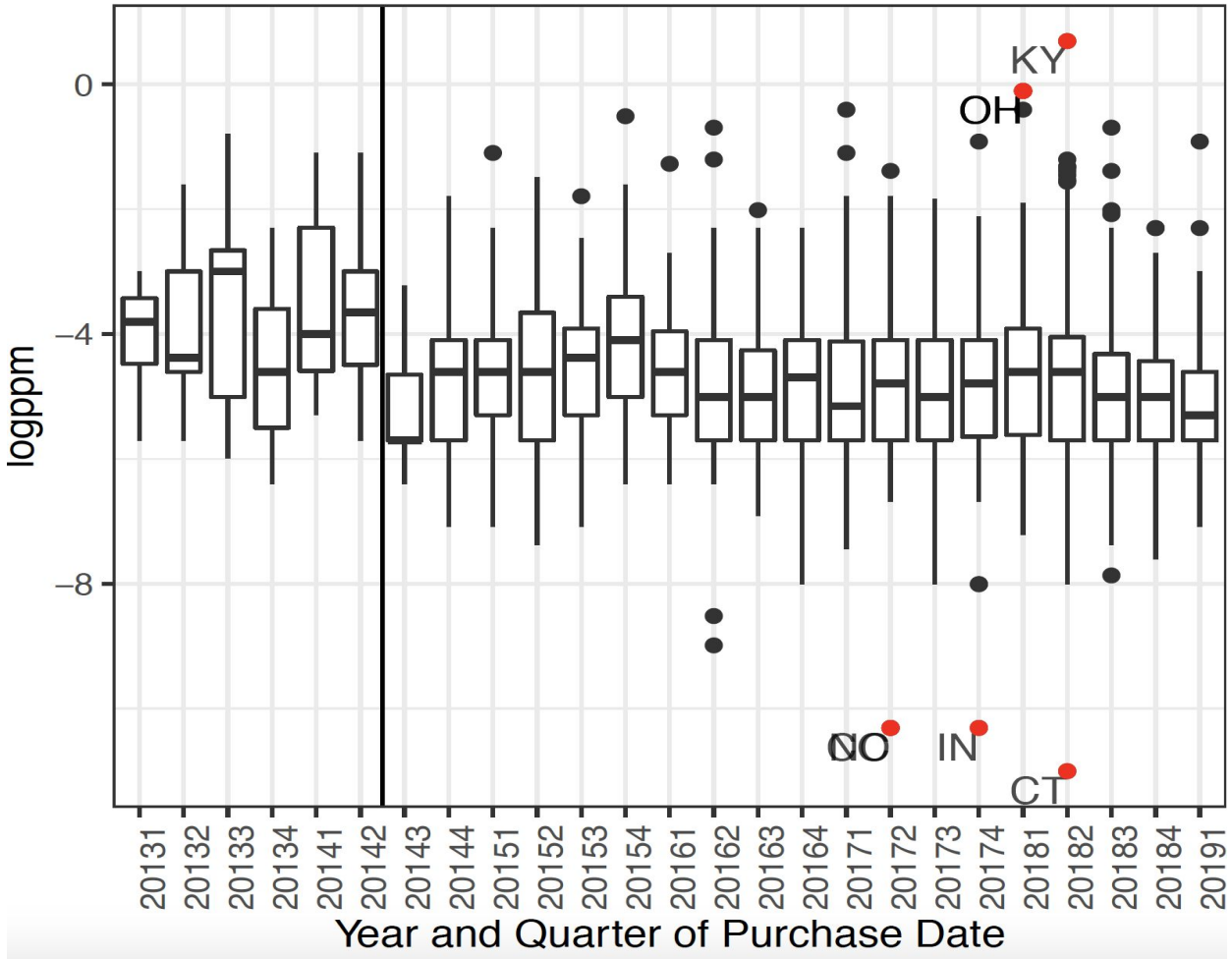
-- **Hierarchical structure needed**

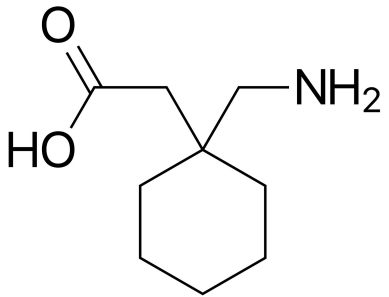


EDA: The policy took effects

Relatively stable over time

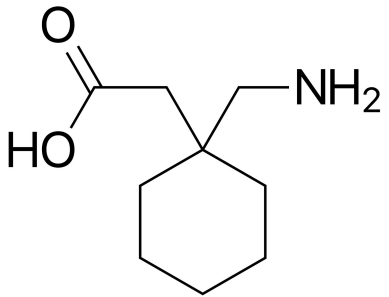
A **significant drop of price** in **2014 Q3** (Policy)





Modeling Process

- **Base model**: **I(policy) + log(dosage strength) + (1|state)**
  - Incorporate time effect:
    - Date, quarter or **I(policy)**
    - Continuous vs categorical
    - For continuous: polynomial, spline, log, …
  - In corporate dosage strength:
    - Continuous (polynomial, spline, **log**, …) vs categorical
- **Full model**: whether to incorporate other factors
- **Pre-final model**: whether to include interaction or consider other random effects
- **Criterion: BIC, LRT**



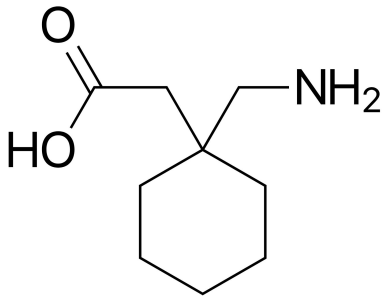
Results

Consider 0.005 threshold.

Table 1: Diagnosis Table for Potential Predictor

Variable	df	p_value	BIC	BIC.full.	BIC.best.
USA_region	3	0.0185872	4909.542	4931.734	4802.573
Source	4	0.1288671	4899.558	4931.734	4802.573
Bulk Purchase	1	0.8085181	4899.558	4931.734	4802.573
I(Policy)	1	0.0001294	4936.878	4931.734	4802.573
log(Dosage Strength)	1	0.0000000	5288.313	4931.734	4802.573
log(Dosage Strength):I(Policy)	1	0.7409263	4811.215	4931.734	4802.573

Final Model



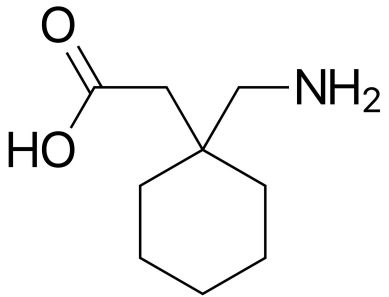
j: state i: individual

$$log(ppm)_{ij} = \beta_0 + b_{0j} + \beta_1 I(policy)_{ij} + log(mgstr)_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma^2)(\text{i.i.d})$$

$$b_{0j} \sim N(0, \tau^2)(\text{i.i.d})$$

$$\epsilon_{ij} \text{ is independent of } b_{0j} \text{ for any } i, j$$



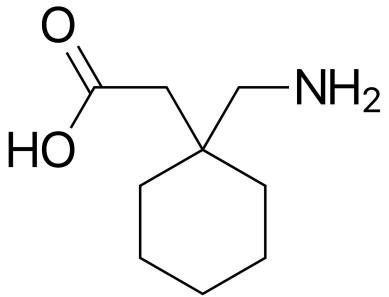
Results

Table 2: Estimation and 95% Confidence Interval

	Estimate	Std. Error	2.5 %	97.5 %	p-value of LRT
Intercept	1.592	0.296	1.012	2.174	
* I(Policy)	-0.64	0.14	-0.915	-0.365	5.432e-06
* log(Dosage Strength)	-0.991	0.048	-1.086	-0.897	1.302e-83
State	0.015	0.123	0.038	0.201	1
Residual	1.201	1.096	1.057	1.135	

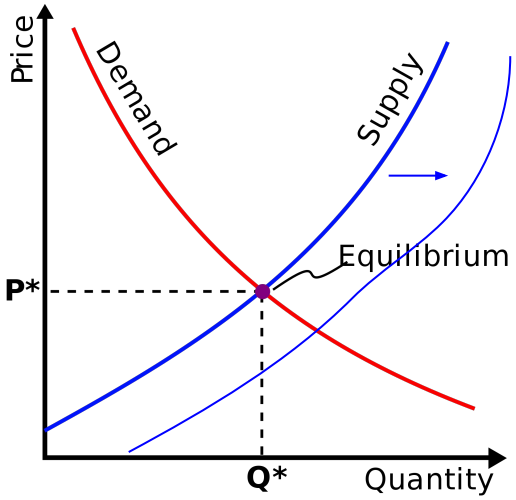
- After Gabbies become a **generic medication**, the price **decreases** by around **30.5% to 59.9%**.
- **Increasing** the dosage strength in mg per unit **by 10%** will **decrease** the price **by 8.2% to 9.8%**.
- **No significant heterogeneity across states**.
- **No significant effects from other factors**.

Interpretation



From Market:

- Generic and Nationwide Medication -> Sufficient Supply :
  - Sufficient Supply → Decrease its price
  - Nationwide Medication → No heterogeneity across states; No heterogeneity effect across states
- Sufficient Supply, Cheap, Popular → Relative Stable:
  - Sufficient Supply: No significant trend over time, across regions
  - **Cheap: not influenced significantly by the demand** (purchase reason or bulk purchase)
  - Popular: No need to lie (source of information)



From Producer:

- Increase dosage strength in mg per unit → Decrease Cost → Decrease Price

State Ranking

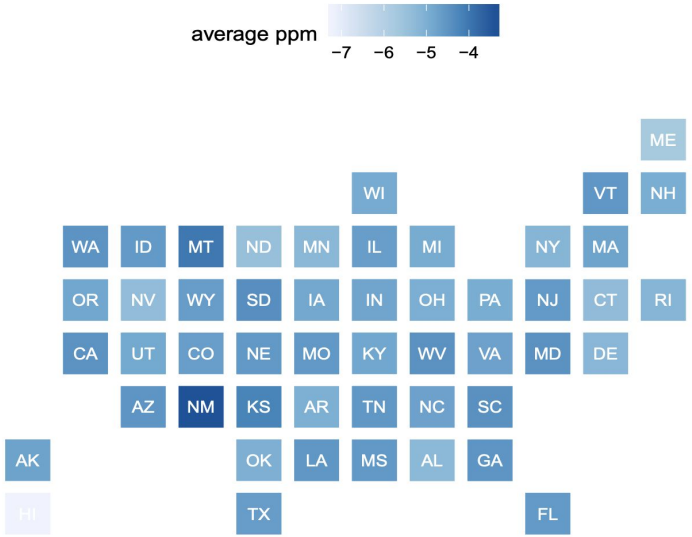


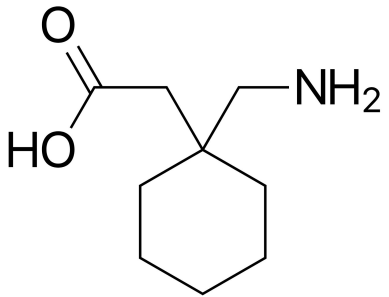
Table 3: Top 5 States with Lowest Average Price (Median)

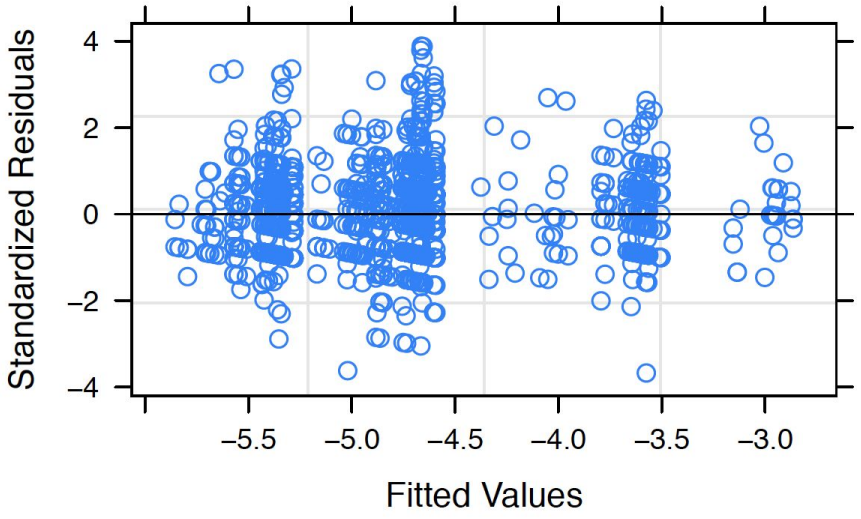
Region	State	Observed Mean	Estimation	5% CI	95% CI
West	HI	0.001	0.004	0.001	0.019
West	AK	0.006	0.005	0.001	0.019
South	MS	0.007	0.005	0.001	0.022
South	DE	0.004	0.006	0.001	0.023
Northeast	RI	0.005	0.006	0.001	0.024

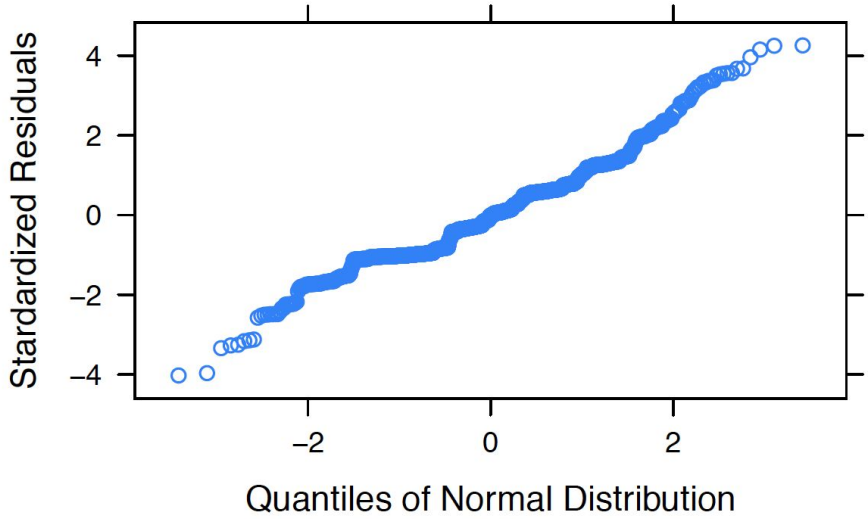
Table 4: Top 5 States with Highest Average Price (Median)

Region	State	Observed Mean	Estimation	5% CI	95% CI
Midwest	KS	0.01	0.026	0.006	0.098
West	NM	0.02	0.010	0.002	0.045
West	CA	0.01	0.010	0.002	0.040
West	MT	0.01	0.010	0.002	0.040
South	TN	0.01	0.010	0.002	0.040

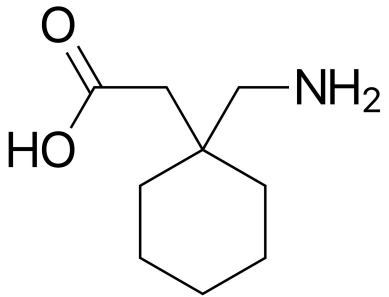
Model Diagnosis

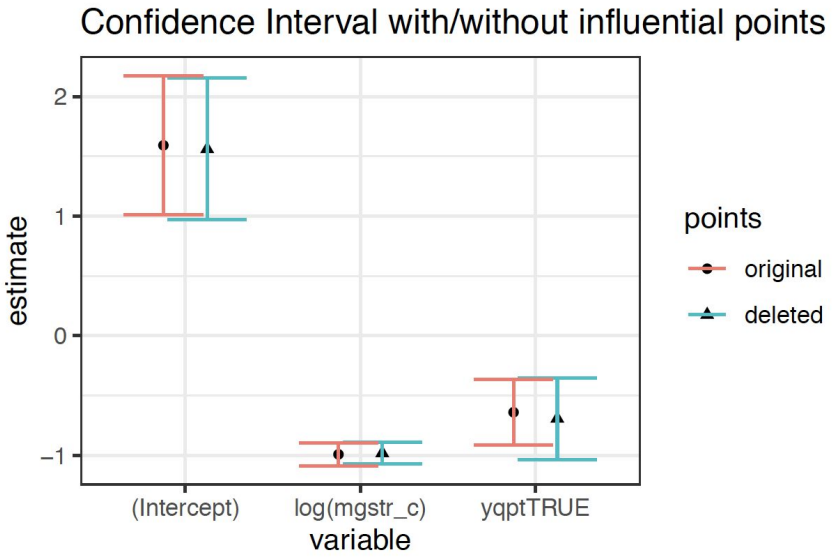


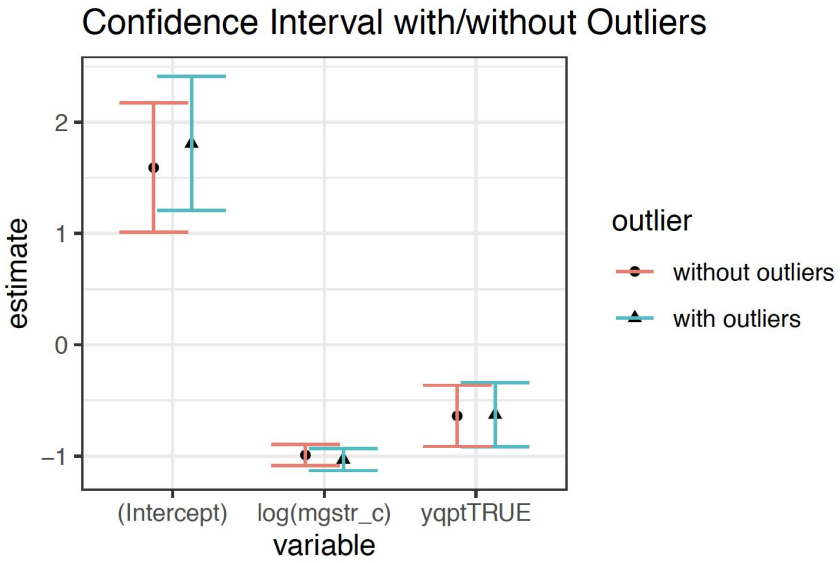


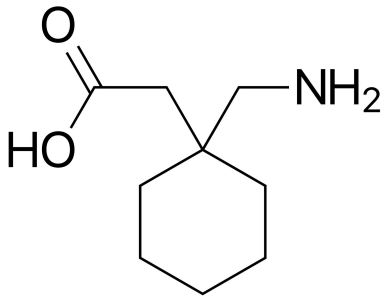


Sensitivity Analysis



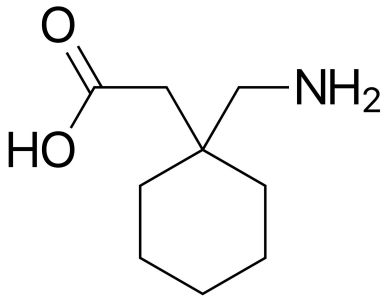






Conclusion

- Our model is **simple** and **effective**. It can be **perfectly explained** and **fit data quite well**. It contributes to provide us some insights on the street market of Gabbies.
- Our model is **quite robust** which is not sensitive to influential points and is slightly sensitive to the outliers.
- A heavy tailed t distribution on the error term may be an alternative choice if we want to emphasize the tail behavior of the dataset.



References

[1] Official website of GoodRx: <https://www.goodrx.com/drug-guide>

[2] Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Cesarini, D.(2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6.

[3] StreetRx: <https://streetrx.com>

Thanks for listening
Q & A