# Methods for Characterizing Temporal Patterns of Prescription Opioid Utilization

Elin Cho, Yuting Xu, Jaejoon Song, Yueqin Zhao, Rose Radin, Grace Chai, Yong Ma, Shekhar Mehta, Meilan Chen, Kyle Lee, Corinne Woods, Saranrat Wittayanukorn

U.S. Food and Drug Administration, Center for Drug Evaluation Research (CDER)

## 1. Background

While the overall national opioid analgesic (OA) prescribing rate declined from 2012 to 2019, the trend varies by geographical location. Numerous methods explore patterns in dispensed prescription data to identify high rates of prescription opioid use in certain areas across the country. Acknowledging the large influence of community-level factors on substance use and related health outcomes, increasing the geographical granularity of analyses can help identify the community-based factors influencing the opioid crisis and reveal geographic sub-populations in which to target surveillance. Several proprietary drug utilization databases are utilized by the U.S. Food and Drug Administration to better understand the scope of prescription OA use and evaluate patterns. One useful approach to enhance pharmacovigilance using these data may be through identification of areas with notable change over time in prescription OA dispensing. In this poster, we describe a two-step approach to: 1) identify geographic areas that had significant change over time in prescription drug dispensing, and 2) characterize clusters of geographic areas that share similar temporal change patterns.

## 2. Methods

### Step 1: Method to Identify Areas with Change Over Time

- We assume that weekly prescriptions dispensing data are discrete observations from a smooth underlying function, denoted as $\boldsymbol{X_i(t)}$, where $\boldsymbol{i}$ indicates a geographic area.
- We use a spline smoothing method (Green & Silverman, 1993; Ramsay & Silverman, 2005) to the data to obtain an estimated function, $\widehat{\boldsymbol{X}}_{\boldsymbol{i}}(\boldsymbol{t})$.
- We identify a subset of geographic locations that change significantly with respect to time, using hypothesis tests (Wu & Wu, 2013) : $\boldsymbol{H_0}: \boldsymbol{X_i(t) = 0}$ vs. $\boldsymbol{H_a}: \boldsymbol{X_i(t) \neq 0}$.
- We use F-statistic to compare the null ($\boldsymbol{H_0}$) against the alternative hypothesis ($\boldsymbol{H_a}$).
- For each geographic location, p-values are obtained using a permutation test. False discovery rate (FDR) is controlled by the Benjamini-Hochberg (1995) method.

### Step 2: Method to Characterize Temporal Change

- We obtain clusters of geographic areas that share similar temporal change patterns using *Iterative Hierarchical Clustering (IHC)* (Carey et al., 2016).
- IHC implements the following three steps to the estimated smooth function, $\widehat{\boldsymbol{X}}_{\boldsymbol{i}}(\boldsymbol{t})$:
  1) **Initialization**: A hierarchical clustering is applied to determine the initial clusters
  2) **Merging**: Clusters are re-assessed by examining the cluster centers. Any clusters with pairwise correlations greater than a pre-specified threshold $\alpha$ are merged.
  3) **Pruning**: Cluster membership is re-examined. Any cluster member with correlations between the cluster center and member less than $\alpha$ is removed from each existing cluster and assigned into a single-element cluster.
  4) Steps 2-3 are repeated until a pre-defined convergence criteria is satisfied.

## 4. Conclusions

- A simulation study demonstrated that the proposed procedure (*step 1*) can quickly identify core-based statistical areas (CBSAs) that had notable changes in per capita drug dispensing over time.
- Clusters identified using the IHC algorithm (*step 2*) can illustrate common trajectories of change among the CBSAs.
- Future directions include using enriched data to explore macro-level policies and health indicators associated with prescribing patterns.

## 5. Acknowledgements

This research was supported by the FDA CDER *Safety Research Interest Group (SRIG)* program and by an appointment to the Research Participation Program at CDER administered by Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between Department of Energy and FDA.

## 6. Disclaimer

This presentation reflects the views of the authors and should not be construed to represent FDA's views or policies.

## 3. Simulation Studies to Evaluate the Methods Performance

### 3.1. Simulation Study for Step 1: Identifying Areas with Change

#### 3.1.1. Simulation Study Procedure (Method: Step 1)

❑ **Data generation**

- Data were generated under 6 different simulation scenarios, where the proportions of core-based statistical areas (CBSAs) with change were defined as $p$ = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 with respect to the total number.
- In each simulation scenario, we generated data for 900 CBSAs over 52 weeks (the U.S. Office of Management and Budget has defined 927 CBSAs for the United States, as of 2020). For non-differential CBSAs, $\boldsymbol{X_i(t) = 0}$; for differential CBSAs, $\boldsymbol{X_i(t)}$ was a linear combination of sine and cosine functions. For each CBSA, noisy signals were generated by adding random errors generated from a Normal distribution.

❑ We identified CBSAs with change over time using Step 1 descried in "Methods"

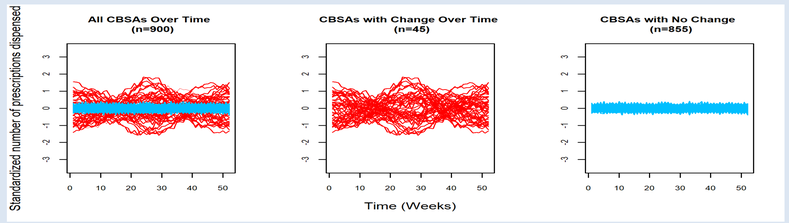


**Figure 1. Illustration of Simulated CBSAs with/without Change Over Time.** This is an example data set for 900 CBSAs over 52 weeks where $p$ = 0.05.

#### 3.1.2. Simulation Study Results (Method: Step 1)

We examined the performance of the proposed method in comparison to a conventional method using ANOVA (which used time as a single factor in the model) in 100 simulated datasets. Results indicate that the proposed method had higher true positive and lower false positive rates compared to ANOVA (Figure 2).

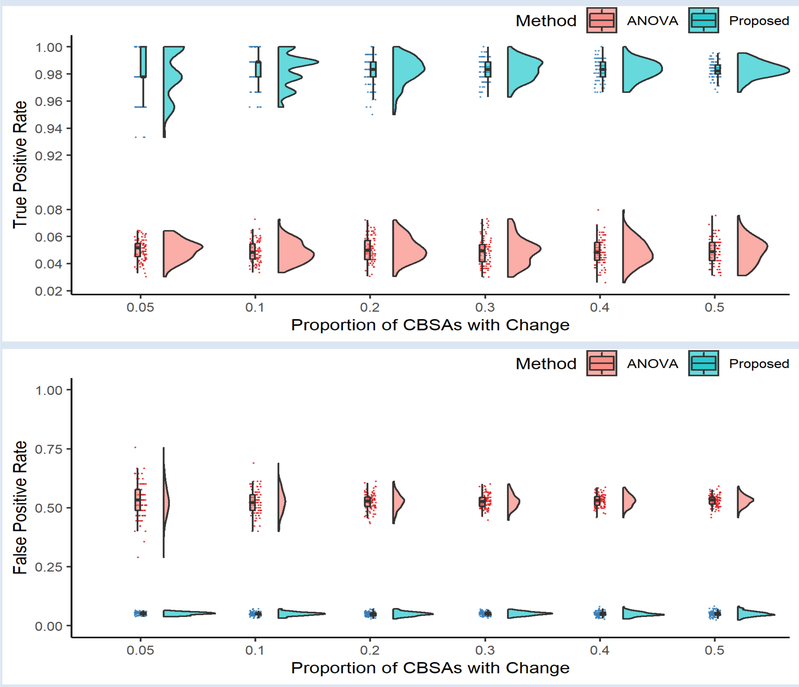


**Figure 2. Boxplots and Density Plots Representing True Positive Rates and False Positive Rates Across the 100 Simulated Datasets.**

### 3.2. Simulation Study for Step 2: Characterizing Temporal Changes

#### 3.2.1. Simulation Study Procedure (Method: Step 2)

❑ **Data generation**

- We simulated data for 900 CBSAs over 52 weeks arising from 3 true clusters
- The simulated data is illustrated in Figure 3. The true change over time pattern for each cluster was generated from a linear combination of sine and cosine functions, specific for each cluster.
- We considered 5 different scenarios, where noisy signals were generated by adding independent and identically distributed Normal random errors with mean 0 and differing variances $\sigma^2$, where $\sigma = 0.1, 0.5, 1.0, 1.5, 2.0$.
- Further, we evaluated the performance of clustering procedure by: i) applying the IHC procedure directly to the observed data; or ii) applying the IHC procedure after spline smoothing.

❑ We characterized CBSAs that share temporal change patterns using IHC algorithm descried in "Methods" section step 2.

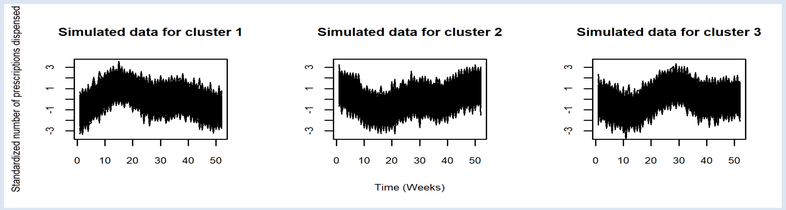


**Figure 3. Illustration of a Simulated Dataset.** This is an example of a simulated dataset with 900 CBSAs over 52 weeks arising from 3 true clusters.

#### 3.2.2. Simulation Study Results (Method: Step 2)

We examined the performance of IHC applied to the estimated smooth function ( $\widehat{\boldsymbol{X}}_{\boldsymbol{i}}(\boldsymbol{t})$ in "Methods") in comparison to IHC applied to the observed data (measured with error) in 1000 simulated datasets. Performance of the proposed method was superior across the 5 different simulation scenarios (Figure 4). The performance differed by cluster (i.e., shape of the true change over time pattern affected the clustering performance).

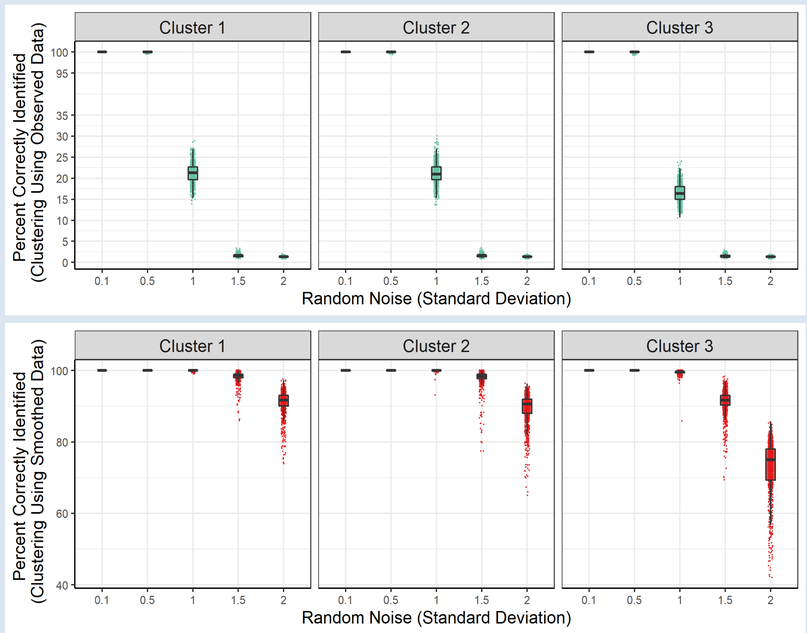


**Figure 4. Boxplots Representing the Percent of CBSAs Correctly Identified Across the 1000 Simulated Datasets for Each Cluster.**