

## Unpacking Averages: Analyzing FDA Device-Related Citizens Petitions by Topic

Article By:

Bradley Merrill Thompson

---

Those who have been reading this blog know that I like to analyze collections of documents at FDA to discern, using natural language processing, whether, for example, the agency takes more time to address certain topics than others. This month, continuing the analysis I started in my [October post](#) regarding device-related citizens petitions, I used topic modeling on the citizens petitions to see which topics are most frequent, and whether there are significant differences in the amount of time it takes for FDA to make a decision based on the topic.

### Discerning the Topics

---

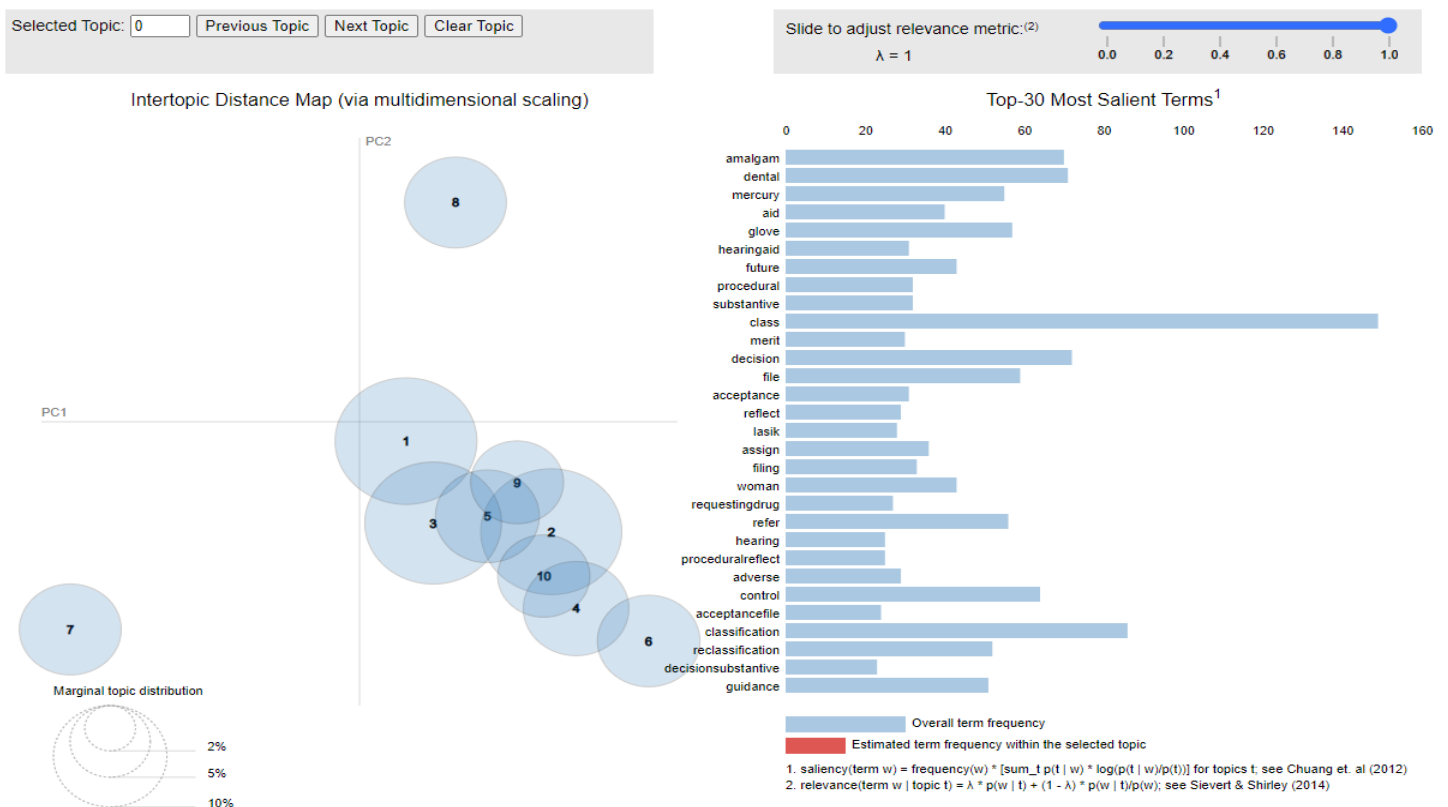
As you probably know, topic modeling is as much art as it is science. There are typically lots of different ways to slice and dice topics, and it takes some judgment to decide the ideal number of topics. The machine does not think like us, so it looks for word clusters that are repeated – word clusters that may not reflect the same way we categorize documents. That’s both good and bad. It’s bad in that sometimes it is difficult to make sense of what the machine learning model produces, but it can be good when the machine learning shows us a different way of looking at documents that we may not have thought of if we humans had done the work manually.

I decided to sort the petitions by 10 topics, recognizing that this way of sorting leaves a very large miscellaneous category. While a larger number of topics necessarily had fewer in the miscellaneous category, the topics themselves weren’t as recognizable at least to me. So here is what I came up with as 10 topics using the topic modeling techniques described below:

1. Classification/reclassification/exemption
2. Surgical and examination gloves

3. Software including clinical decision support
4. Radiation and adverse reports
5. Hearing aids
6. Dental issues including use of amalgam/mercury
7. General/future regulation. This is my miscellaneous category but it seems a little bit focused actually on future FDA decisions made on process basis
8. Medication safety
9. Implants
10. Lasik and otherwise involving the eye

You can visually see these topics in the interactive chart below where if you hover over the topic circle on the left, you will see the words that comprise the topic on the right.



*Click here to open the interactive graphic in a new browser window.*

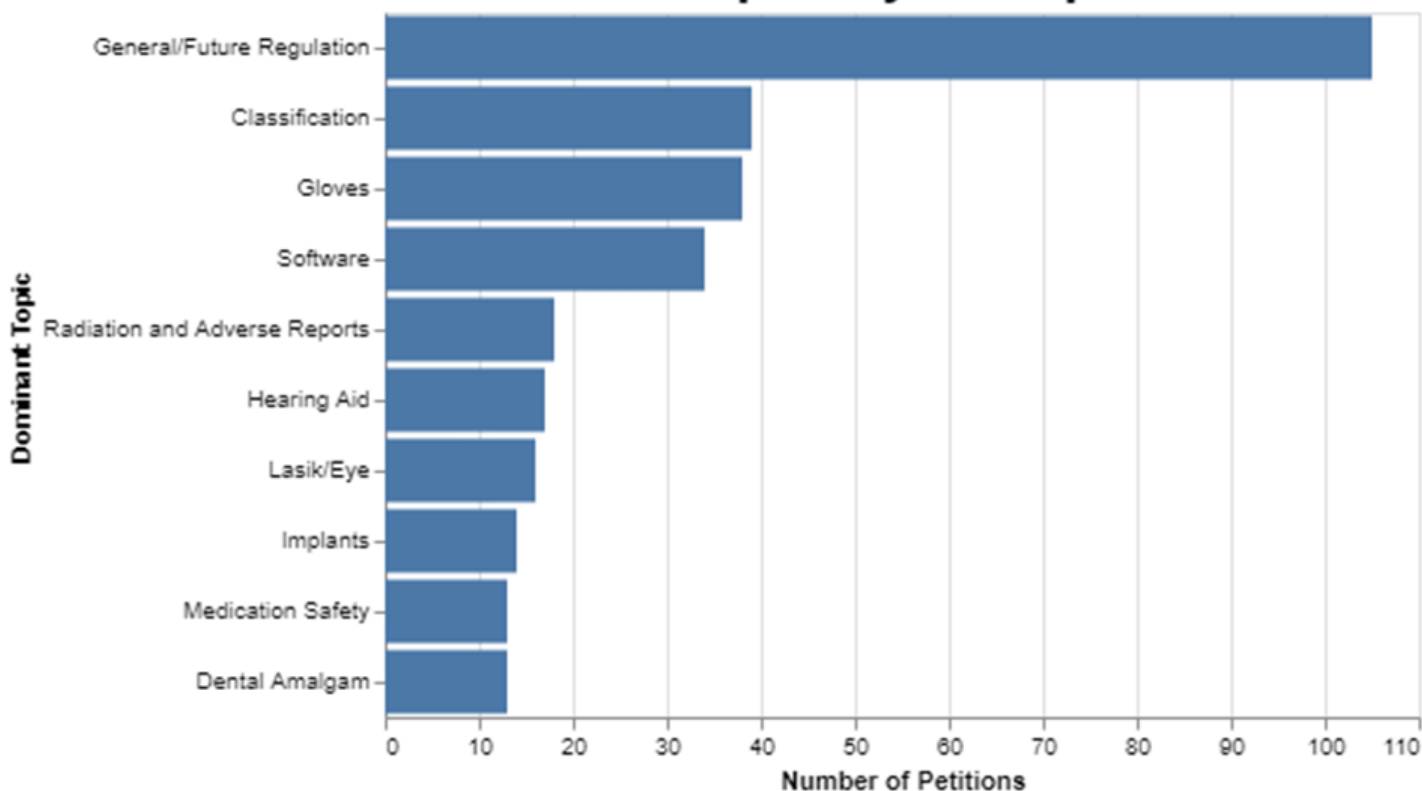
If you are interested from a technical standpoint, I used the spacy library to do the actual topic modeling. Before running the topic modeling algorithm, I used Phaser to create bigrams (two words together) because it is often important to look at words at least in pairs to catch, for example, the difference between “United States” and the words “united” and “state” separately. I used various coherence scores to arrive at 10 as the ideal number of topics.

I had to use my judgment in deleting uninteresting words. Uninteresting words simply dilute the topics in a meaningless way. The indefinite article “a”, for example, is just clutter and really doesn’t provide any information. I used a standard set of so-called “stop words” to get rid of trivial words, but I also came up with a tailored list of additional stop words to get rid of some of the clutter in citizens petitions. Thus, I took out 87 additional words such as 'request', 'fda', 'citizen', 'petition', 'cosmetic', 'cfr', 'united', 'states', 'commissioner', 'drugs', and 'federal' because frankly they appeared in just about every citizens petition in uninteresting ways like the “Federal Food, Drug and Cosmetic Act.”

## Analysis

I first wanted to see how frequent each of these topics was. As I mentioned already, selecting 10 topics means that the miscellaneous category is really big.

### Frequency of Topics



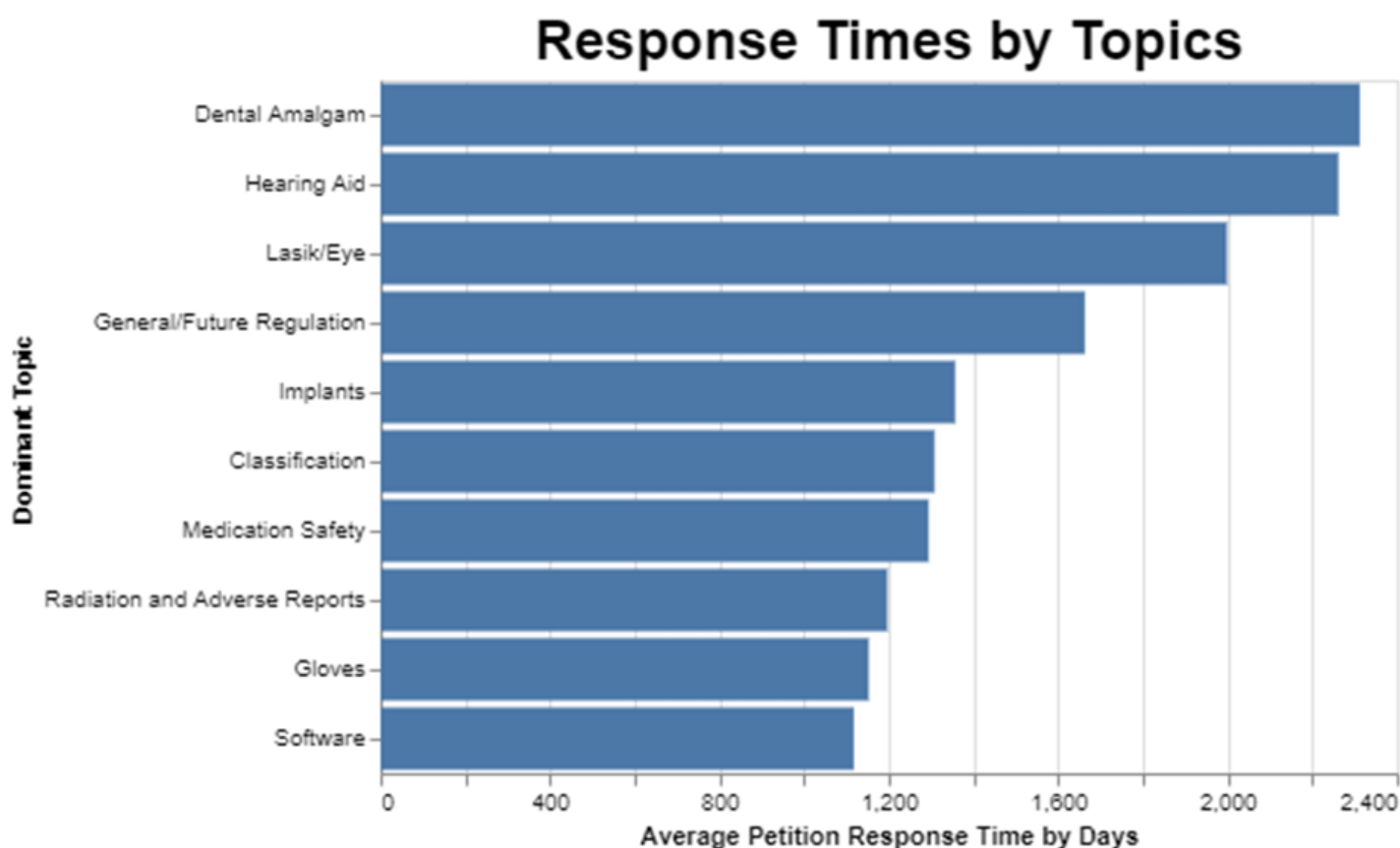
I could have gotten the miscellaneous category much smaller, but the trade-off would be that the other categories would not be nearly as meaningful. And I kinda liked the clarity of the other topics even though this approach left a lot in miscellaneous.

While the words “adverse” and “reports” showed up a lot in the category where radiation

was also the predominant topic, it's important to recognize that none of these topics are pure. They have dominant words and then they have less frequent words. This is to be expected because when you think about it, these petitions are each written to address whatever the particular petitioner wants to cover, so there's no reason to think that the topics would fall neatly into discrete categories.

Further, it's possible that a petition could address more than one topic like reclassification of an implant for the eye through a procedure that involves radiation, so I picked the dominant topic for each petition, which is just based on the mathematical prevalence of the topic within the petition. Thus, it's artificial to assert that each petition only addresses one topic. But I had to limit each petition to one topic to do the rest of the analysis, including, for example, figuring out how long it takes to address each topic. But just understand the artificial nature of the analysis.

After figuring out frequency, I wanted to understand how long it takes FDA to respond to petitions by topic.



The dominant topic appears to affect the review time substantially. Apparently, the dental filling topic was a thorny one for FDA to decide, because it took on average over 2200 days to resolve those petitions. On the other hand, and this got me excited because I filed a petition on clinical decision support software, the software topic seems to get resolved in, on average, half that time, or about 1100 days.

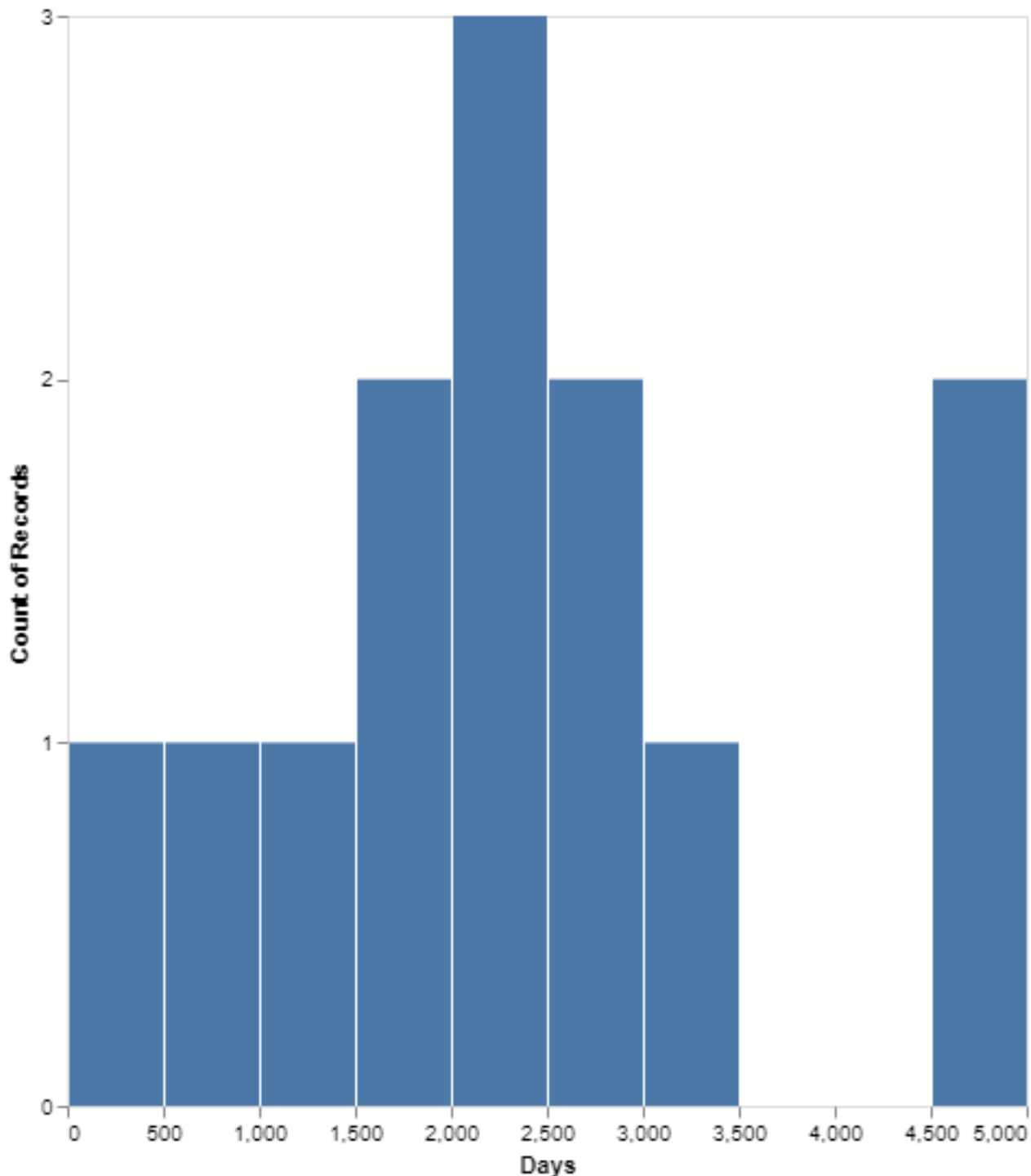
I have to catch myself when I talk about getting excited about a mere 1100 days of review time. The statute calls for 180 days, so the average software petition gets resolved in more

---

than six times the amount of time it's supposed to take.

And finally, obviously it's possible to go deeper into any of these averages. I picked the dental amalgam topic because it is the longest on average, just to see what the range is for petitions in that category.

## Response Times For Topic Dental Amalgam



While the average in that category is just over 2200 days, the longest one that category took apparently close to 5000 days. That's a lot of days. In fact, it's almost 14 years.

Conclusion

---

---

I started this analysis because I wanted to figure out how long I should expect to wait for a response from FDA on my citizen petition filed in February 2023 on the topic of clinical decision support software.[1] I guess I have my answer. On average, for that category, it takes about 1100, *and that's the fastest category of any of them*. That's almost exactly 3 years. So if my petition represents the average, I should expect to hear in February 2026. Good grief.

I filed the petition because the *Final CDS guidance* from September 2022 is chilling innovation in the software space, especially the application of artificial intelligence to clinical decision-support. Patients need doctors to have better clinical decision support, because right now doctors on their own are making many mistakes. There are many statistics available, but by one estimate outpatient diagnostic errors affect roughly 12 million Americans every year.[2] I'm not criticizing doctors because all humans make mistakes, especially in areas of great complexity. To err is human, but fortunately we can make machines that reduce those errors by helping humans see more options and think more rigorously.

My point is that clinical decision support software perhaps powered by machine learning and perhaps generative AI in particular has the potential to help doctors think more expansively about possible diagnoses, and the FDA's final CDS guidance is chilling innovation in this space. Are patients simply supposed to wait until February 2026 for this unlawful chilling of innovation to be lifted? If the petition that I filed or the subsequent July petition filed by the University of Florida law professor[3] are obviously wrong or mistaken, then it would be easy for FDA to simply reject our petitions in a one-page refusal quite quickly. The data I shared in October show that FDA can, when a petition is obviously wrong, reject the petition quickly. I can only assume that FDA is not acting on the two petitions because they can't be rejected easily.

The system is broken, and FDA is not following the law, while patients suffer. That's not right.

*The Unpacking Averages™ blog series digs into FDA's data on the regulation of medical products, going deeper than the published averages. Subscribe to this blog for email notifications.*

## ENDNOTES

[1] <https://www.regulations.gov/document/FDA-2023-P-0422-0001>

[2] Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf.* 2014 Sep;23(9):727-31. doi: 10.1136/bmjqs-2013-002627. Epub 2014 Apr 17. PMID: 24742777; PMCID: PMC4145460.

[3] <https://www.regulations.gov/document/FDA-2023-P-2808-0001>

National Law Review, Volumess XIII, Number 339

Source URL: <https://natlawreview.com/article/unpacking-averages-analyzing-fda-device-related-citizens-petitions-topic>