Analysis of the most harmful and lethal products' reports from CAERS dataset

Akvilė Višniauskaitė ir Ieva Pudžiuvelytė (AI)

2022-03-02

Introduction

A public institution that regulates sales of the products is willing to label dangerous products accordingly. The information about the products that cause the most serious symptoms was requested. More formally, answers to the following questions needed to be provided:

1. What are top 5 most serious (the most harmful to health) symptoms?
2. Which industry products cause death most frequently?
3. How many and what kind of products should be banned in order to reduce number of deaths by 20%?
4. What other interesting insight can be found?
5. What additional library to visualise data could be used (demonstrate the usage)?
6. Give answers to least five additional questions:

- What products were concomitant to cause death?
- Were there any other reports of folligen solution therapy?
- What are other symptoms caused by "Exemption 4"?
- Which gender uses product "Exemption 4" more?
- Which age group uses "Exemption 4" the most?

Comments about the report

- The first, second, third, and additional questions are written down in the report in *italics* font for a clear separation.
- Additional questions were formulated in a course of analysis, thus the sequential reading of the report will explain why certain additional questions were raised.
- Choice and usage of the additional library for data visualisation was described in a separate section named "Visualisation of proportions".

Description of data

Data was taken from a database CFSAN Adverse Event Reporting System (CAERS), which was collected in the period of January 2018 to December 2020 [1, 2, 3]. The database consists of compliants by health-care professionals and consumers from the United States and are related to foods, dietary supplements, or cosmetics.

The original data for analysis is presented in a form of a comma-separated value table. The original dataset [2] had few modifications for data consistency: duplicate report identifiers were removed, representation of an unknown gender was set to be consistent (by "U" letter), all values in the dataset changed to be lower case, whitespaces trimmed, age units converted to years, and a separate matrix indicating the symptoms was created. Consequentially, the main dataset of reports, which was used for analysis, consisted of 10 columns with the following information:

- id - an identifier of the report
- created - a date when the report was created
- date - a date when symptoms occurred
- type - a label whether the product is considered as "suspect" or "concomitant" to cause the reported outcome
- product - a name of the reported product
- code - a numeral identifier of an industry of the reported product
- industry - a literal identifier of an industry of the reported product
- age - age of a patient (years)
- sex - gender of a patient ("M" - male, "F" - female, "U" - unidentified)
- outcome - symptoms that the product caused (or was suspected to cause)

Harmful products

Which are top 5 most serious (the most harmful to health) symptoms?

After the data analysis, it was observed that there were cases of death reported, which could naturally be included as the most harmful symptom to one's health. However, the analysis regarding reported cases of death will be presented below, thus this section is an attempt to filter the most serious symptoms besides death of an individual.

Seriousness can be interpreted differently. In our view, any form of stroke, cancer, paralysis, or anything related to brain or heart are serious symptoms, or, more precisely, outcomes that appeared after consuming or using the product. Therefore, we extracted and listed out these groups, although, the symptoms within these groups are rather diverse.

There were reports of different types (number of types is provided after each dash):

- death (including sudden, foethal, cardiac, and brain) - 9
- stroke (including ischaemic, haemorrhagic, and brain stem) - 5
- cancer (including stage IV and metastatic forms) - 70
- related to brain - 15
- related to heart - 8
- paralysis - 4

In order to give a clear answer to the question, we picked out five most serious symptoms from these groups that were apparent in the dataset of the reports:

1. Brain death
2. Ischaemic stroke
3. Haemorrhagic stroke
4. Metastatic cancer forms
5. Stage IV cancer forms

Lethal products

The overall number of lethal outcomes was 5391. These outcomes could be grouped to 26 different collections of outcomes that included death as one of them.

What products were concomitant to cause death?

Only 1 of all reports had a type **concomitant**, the remaining ones were only suspected to cause death. The concomitant case was about folligen solution therapy from cosmetics industry that caused death for a male patient, who was 29 years old.

Were there any other reports of folligen solution therapy?

The concomitant report regarding the death of the 29 years old male was the only one with folligen solution therapy included.

Interesting insight about this report

After checking the original data (before modifications) another report with the same identifier, outcome, and type "concomitant" was found about "Emu Oil-S", therefore folligen solution therapy might not necessarilly be the main cause of this death case.

Reducing number of deaths

How many and what kind of products should be banned in order to reduce number of deaths by 20%?

In order to answer the question, which product group should be banned to reduce the number of deaths by 20%, we would have to find the product or several products that makes up 20% of all deaths reported.

- Number of all death cases: 5391
- Number of death cases that makes up 20%: 1078.2

Formally, the goal is to find a group of products that would be the suspected cause of at least 1078.2 death cases.

However, analysis of the data brought us to another *interesting insight* - it was apparent that such group could not be found, since there is one product named "Exemption 4" from cosmetics industry which is responsible for 5078 death cases, therefore it is responsible for 94.19% of all cases. Consequentially, there are two possible actions that could be suggested:

- Ban all other products except "Exemption 4" - this action would decrease the number of deaths by 5.81%
- Ban "Exemption 4" - this action would reduce the number of deaths by 94.19%

Therefore, if the goal is to reduce the number of deaths by at least 20%, "Exemption 4" should be prohibited.

Which industry products cause death most frequently?

Based on the analysis described above, the industry whose products cause death most frequently is cosmetics, since this is the industry to which "Exemption 4" belongs.

Visualisation of proportions

A pie chart below presents the proportions of the death causing products. For more clarity, "treemap" package was used as an alternative to visualise proportions, since the most common argument against pie charts is that varied angles complicate the process of detecting differences between displayed areas [4].

The pie chart shows only two products separately, since these products had at least 100 reports ("Exemption 4" 5078 and "Kratom" 122). The third most frequently reported product regarding death cases was "Centrum silver men's 50+" with 7 reports, thus products with fewer than 100 cases were grouped and displayed as "Other products".

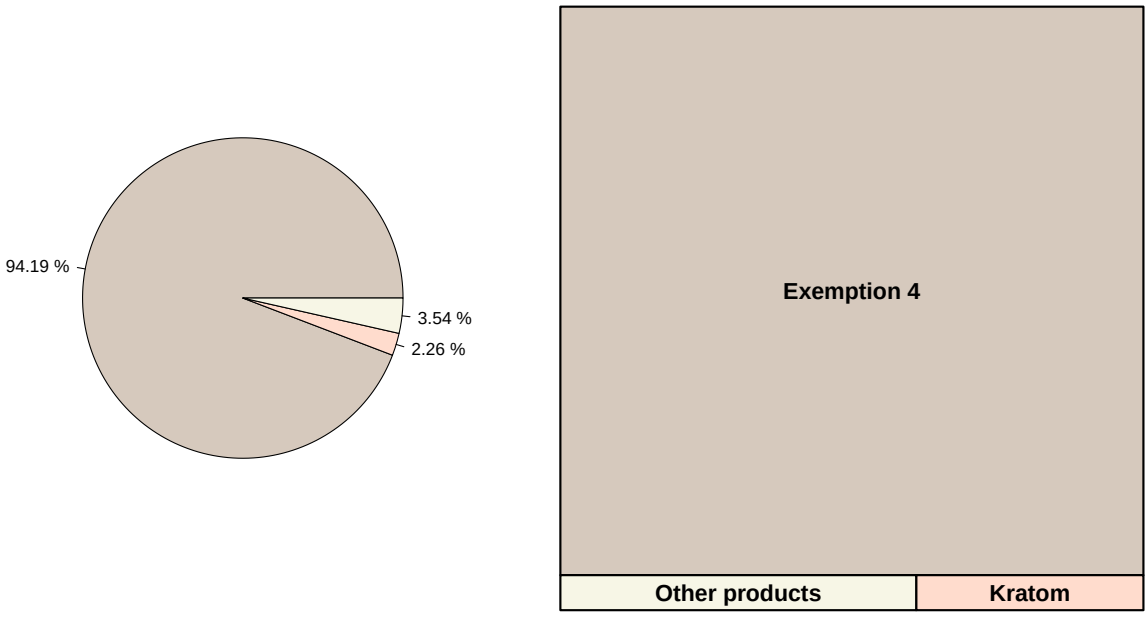

Fig. 1. The proportions of products that were reported with death as an outcome

What are other outcomes caused by "Exemption 4"?

There were 10 other outcomes listed for "Exemption 4":

1. hospitalization
2. medically important
3. patient visited healthcare provider
4. other outcome
5. other seriousness
6. patient visited ER
7. disability
8. life threatening
9. required intervention
10. congenital anomaly

Which gender uses the product "Exemption 4" more?

The gender group that uses "Exemption 4" more is the female group, which makes up 93.91% of all reports of this product.

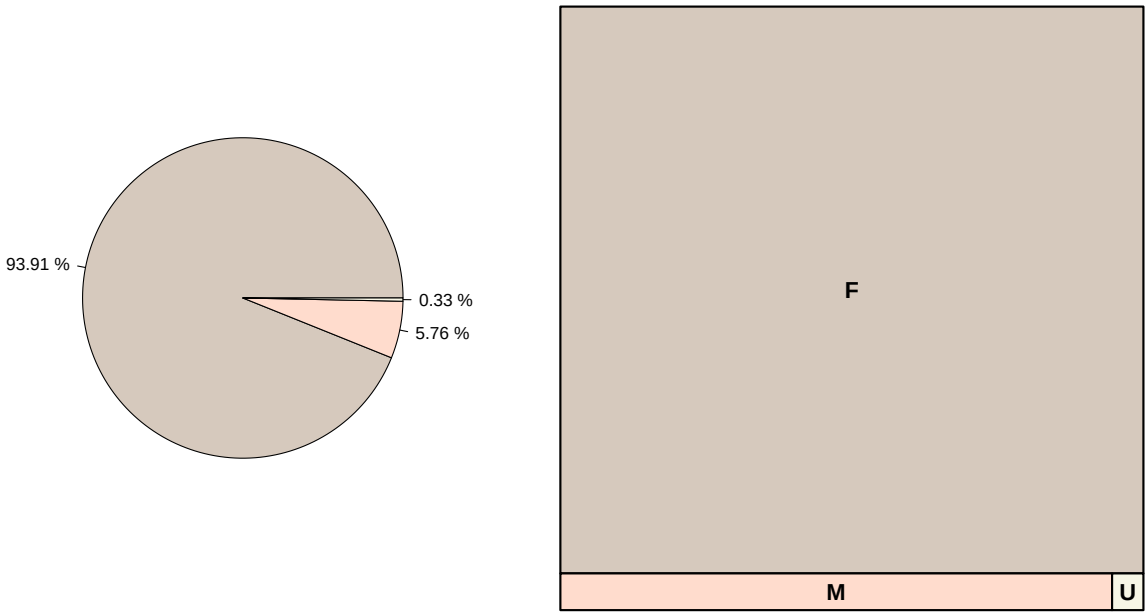

Fig. 2. The proportions of gender groups that used product "Exemption 4"

Which age group uses "Exemption 4" the most?

"Exemption 4" is mostly reported by 40-80 year old people. The exact age with the most reports was 60 years and the majority of cases reported by this age group were female (315 out of 326).

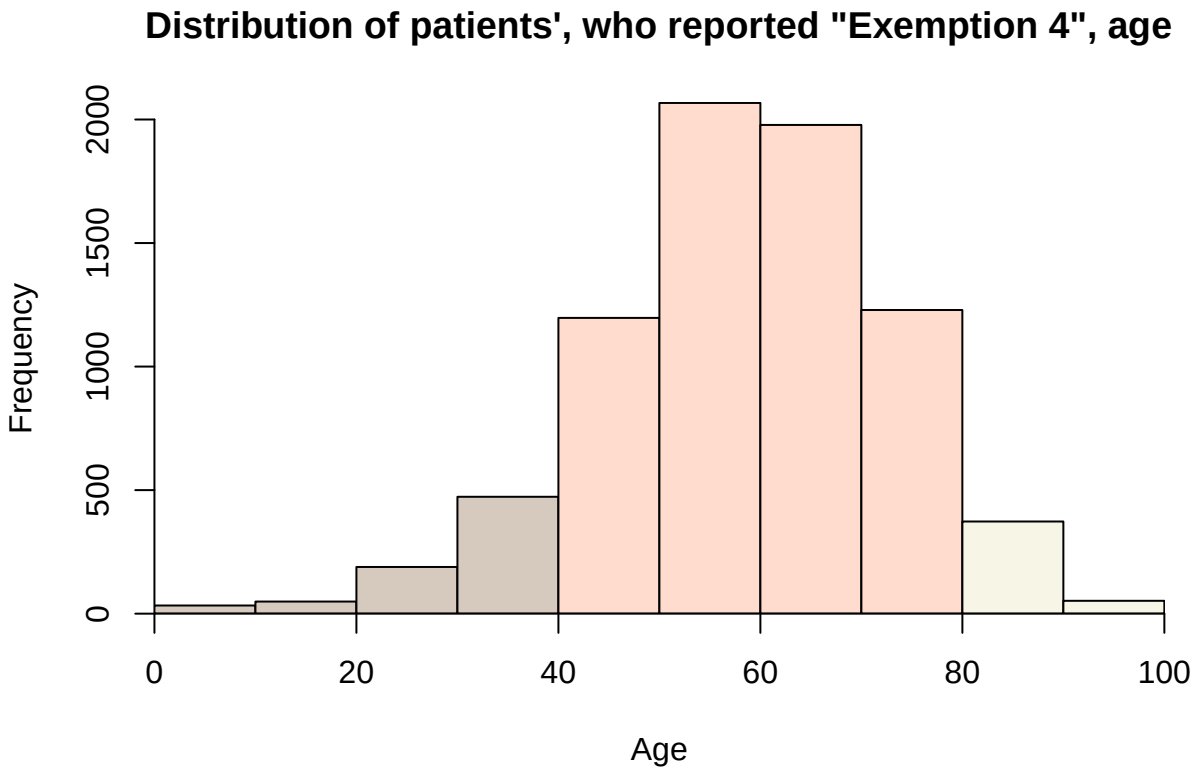

Fig. 3. The histogram displaying distribution of age among patients who reported "Exemption 4" age

Conclusions

To conclude, in the period of 2018 - 2020 only one product was considered as concomitant to the causial of death, although it was reported along with other products used at the time. The majority of death cases reported included "Exemption 4", which is cosmetics product - no additional information about it was provided, yet it might be responsible for around 95% of reported death cases.

References

1. U.S. Food and Drug Administration. (2022, January 31). *CFSAN Adverse Event Reporting System (CAERS).* https://www.fda.gov/food/compliance-enforcement-food/cfsan-adverse-event-reporting-system-caers.
2. U.S. Food and Drug Administration. (2020, December). *CAERS ASCII.* https://www.fda.gov/media/140575/download
3. U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition (CFSAN), Office of Analytics and Outreach (OAO). (2022, January). *README file for the Quaterly Data Extract (QDE) from the CFSAN Adverse Event Reporting System (CAERS).* https://www.fda.gov/media/97035/download
4. Huestegge, L., & Pötzsch, T. H. (2018). Integration processes during frequency graph comprehension: Performance and eye movements while processing tree maps versus pie charts. Applied Cognitive Psychology, 32(2), 200-216. https://doi.org/10.1002/acp.3396