Medical Prejudice across Gender

Sven Kuffer
University of Luxembourg
Email: sven.kuffer.001@student.uni.lu

Sajjad Davoudi
University of Luxembourg
Email: sajjad.davoudi.001@student.uni.lu

***Abstract*—This study addresses pervasive gender biases in healthcare by analyzing clinical trials overseen by the U.S. Food and Drug Administration. We compiled a dataset from the FDA's clinical trials API and applied a combination of statistical methods and NLP to uncover systemic biases that adversely affect healthcare quality. Our findings underscore the critical need for inclusive clinical trial methodologies and equitable guidelines to ensure comprehensive and fair medical research. By highlighting specific disparities in disease diagnosis and treatment efficacy across different demographics, our research advocates for targeted reforms to mitigate these inequities in healthcare outcomes.**

I. Ideation

In this project we try to draw attention to the significant impact of gender and ethnic biases, utilizing various sources to highlight the prevalence and implications of this disparity , arguing for the urgent need to address them. These biases have led to delayed diagnoses and treatments, especially in critical areas such as cardiovascular disease, where women are 50% more likely to receive the wrong diagnosis after a heart attack. Ethnicity, Black People are more likely to have diabetes, cost for treatment, insulin is high, despite publicly available patents and low manufacturing costs. (in the USA) [1]
These issues are rooted in historical bias and systemic inequality. Addressing this challenge is not only a matter of improving health outcomes for the majority of the population but also a critical step toward achieving equity and fairness in healthcare.

Unethical experiments are sometimes done on minority communities across history. Biased findings and reports are published that shape our ideal health model.(msg– unhealthy– Asia– bad). [2]
Symptoms show differences between genders. Females are more likely to be undiagnosed or misdiagnosed with Autism. [3]
Textbooks study the Caucasian male body as the standard model. [4]

II. Research Question

To which extend are there differences accessibility, safety and representation in medical research across gender?
The Food and Drug Administration (FDA) in 1993 updated it's guidelines for conducting clinical trials, to help improve the inclusion, safety and study of particularly the underrepresented female body, in clinical trials. To analyse our dataset, we need to know about these rules and guidelines.
Additionally, in 1994, the Office of Women's Health (OWH) was founded to further support the cause, with a joint updated road-map in 2016. As the US has some of the strictest rules in the medical field, with the FDA as a governing body, we are analysing their guideline changes and their effect on clinical trials with data from the FDA ClinicalTrials API.
The new guideline includes three major improvement points. [5]

a) (1): Lift a participation ban for women with childbearing potential from entering early stages of trials.

b) (2): Require sponsors to include a balance of both genders as participants in clinical trials to detect significant gender-related differences.

c) (3): Study the effects on the menstrual cycle, exogenous hormonal therapy including oral contraceptives and the effect of drugs on the pharmacokinetics of oral contraceptives.

- H0: Nothing has changed from before the introduction of new guidelines to now, to improve the representation, safety and study of women in medicine.
- H1: The lifted ban on women of childbearing age entering early trial changes has opened up the inclusion in conducted trials.1a
- H2: Women are more equally represented in trials, by number and lesser restrictions.1b
- H3: More studies focusing on the menstrual cycle and female hormonal contraceptives have been conducted since.1c

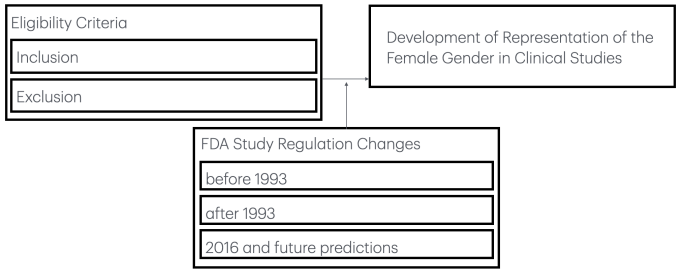
III. Dataset

We fetched our own data from the ClinicalTrials API by the FDA. Our dataset combines metadata about 67676 trials from 1916 to 2030. This means that not all trials have even been completed yet.
Study Status has 15 unique values, some of which are 'COMPLETED', 'Recruiting' and 'ACTIVE_NOT_RECRUITING'.

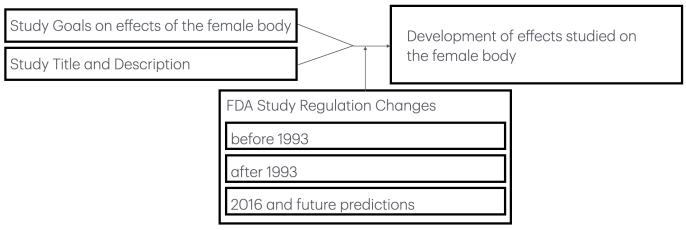
Most studies include both sexes, no distinction between any other sex/gender is being made.2 Studies are being conducted all around the globe, most of them in the USA, closely followed by France and Germany.3

A. Limitations

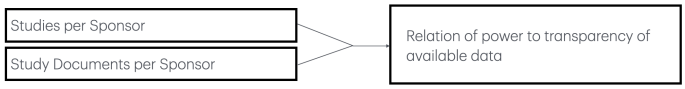
Our main dataset has only metadata about studies, but through the link associated, we can get more information. Unfortunately, not all studies have exact demographic numbers



(a) H1 Conceptual Model



(b) H2 Conceptual Model



(c) H3 Conceptual Model

Fig. 1: Conceptual Models for Hypotheses 1, 2, and 3

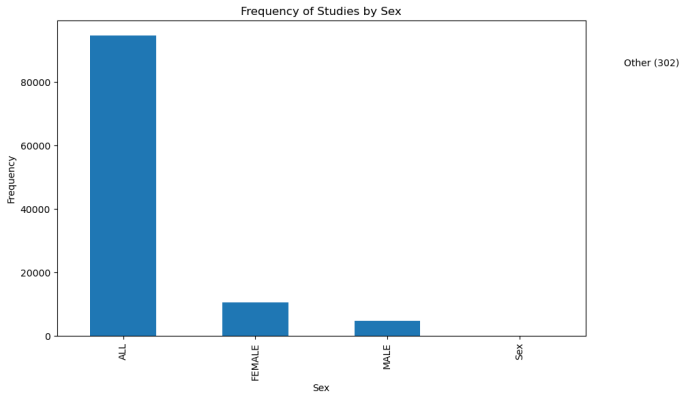


Fig. 2: Sex selection criteria

posted, mostly only those with posted results or published studies are useful to us. Generally, we can assume to only find relevant information for completed studies. Many studies that have started after the 2016 initiative and updated guidelines are still ongoing, so we don't have that information yet. We may need to apply nlp techniques and scrape the clinical trials page for selection criteria rules. Overall, it is feasible to automate, but we may not get as much data as we would have liked, as only a handful of trials have published results and reports.

Our dataset does not have valuable information about exact participant numbers or splits, these kind of information are hidden and we would have to search much deeper to extract them. Only 7587 trials include study documents, relevant to get this information. Only 8.96 % of completed studies come

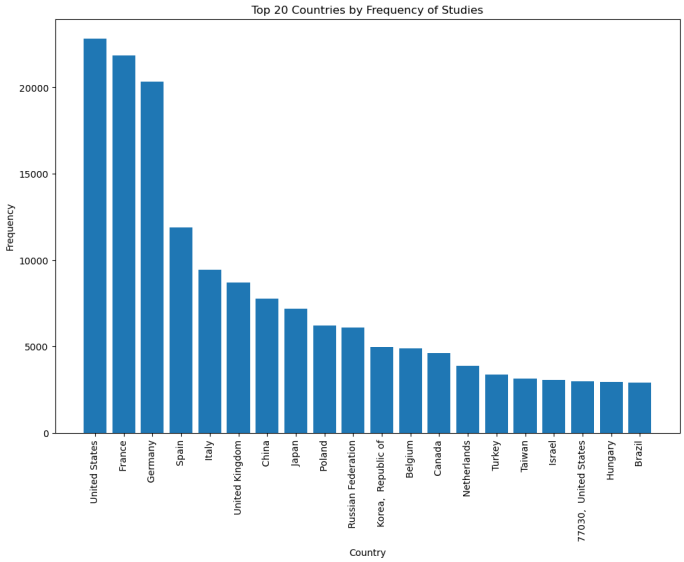


Fig. 3: Locations of trials being conducted

with documents. This is a huge bottleneck for our research in terms of availability. It was not possible for us to get that information. The data from before 1993 is sparse, we don't have enough entries to make reliable conclusions, but we can track the development of changes through the decades from the point where guidelines were introduced, as we do expect changes to happen gradually over time, rather than instantaneous between guideline changes. The data from 2016 onwards is incomplete, as it also includes future research that has not been completed yet. For most of our research, this was not a concern however, as the information we were interested in is already available at the moment a new study is tracked by the FDA. Our dataset thus includes future projections for upcoming studies. Prior to 2016, 77% of studies have the status completed, whereas after 2016 only 39% of studies are marked as being completed.

IV. Methods

We will apply a mixture of NLP methods, to extract information from nominal data and apply classification, clustering and topic modeling on eligibility criteria and study titles and descriptions. Our main dataset has been fetched by us from the FDA's clinical trials API in multiple steps. It contains metadata about clinical trials. We use the FDA's rules and guidelines for clinical trials to examine this data, as well as the revised rules specifically created to include more women in trials, in 2016. Data collection is very important, finding existing datasets, but also looking into methods to gather new data. Our main independent variables are sex, age, specific selection criteria, locations, sponsors, study status and date. For the date, we will look at different 3 periods , we are specifically interested in the cut prior 1993, after 1993 and after 2016. After 2016 will include future, not yet completed studies and enable us to make future projections. This cut is due to new FDA guidelines to include more women in trials.

The first hypothesis will be studied on eligibility criteria, we examine both exclusion and inclusion criteria and compare it with the dates were new regulations were introduced. We are particularly interested in the inclusion or exclusion of women with childbearing potential, as per the change in the guidelines. However, we do not limit us to that and review other words and phrases related to the female gender. We will train a classifier on eligibility criteria, using Support Vector Machines and Naive Bayes and select the best performing option. We will search for words related to 'childbearing potential' within the corpus, to determine, whether the lifted ban has had a positive effect.
The second hypothesis requires us to get quantitative numbers on the gendered split of participants, which we could not get. Instead, we will study trial sponsors. How much power the biggest players have and what it might mean in terms of equality. We are comparing the availability of study documents as opposed to the number of studies conducted for each sponsor and analyse, whether there is a trend in either direction. This will give us an insight in how transparent sponsors are with their data, data which we would like to have more reliable access to, but unfortunately don't. We will use k-means clustering and regression to plot and visualise trends. For the last hypothesis, we are studying the study titles and descriptions to see, whether these effects are studied more, with projections of the future. We will search for hand selected terms, related to the guideline statements, on a preprocessed corpus, to analyse the trend over time.

A. Considerations about data gathering

In case of patient specific data, this might be very difficult, as we deal with very sensitive personal data, that we have to treat ethically. For this and varying other reasons, patient data is often very sparse and hard to get too. Consider Minimal Mind State examination results for Alzheimer's patients or related disease. In this case, data is not readily available. The attempt has been made to train cnn and rnn ML models that work on smaller datasets to predict and classify MMST results. In this example case, patients are usually old and they might have varying other conditions with symptoms overlapping each other. It is often hard to use this kind of data.

V. Results

For each of the three hypothesis, we dedicate it's own subsection with the produced results.

A. H1

We trained a classifier on Eligibility Criteria, by splitting the field into exclusion and inclusion criteria and split statements. Our best performing model was an SVM model with 'rbf', it outperformed all other SVM or Naive Bayes models.

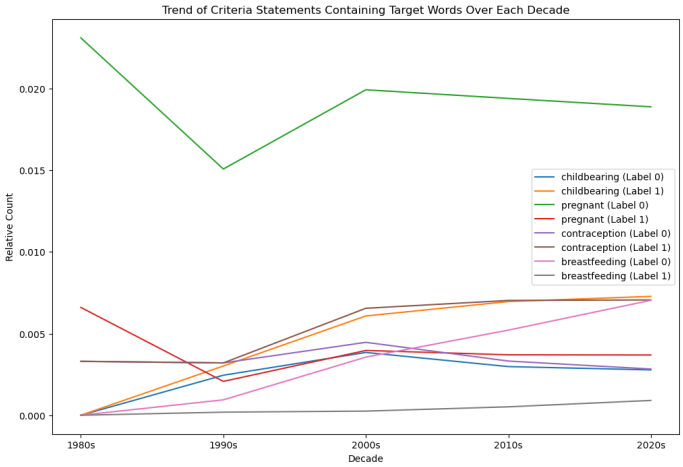


Fig. 4: Trend of Target Words Appearing as Eligibility Criteria

label	precision	recall	f1-score
0	0.89	0.90	0.90
1	0.85	0.83	0.84
accuracy			0.88

. The trend shows an influx of words related to pregnancy after 1993 as an exclusion criteria. This seems bad at first, as more women seem to be excluded. If we think further, however, it means that safety of pregnant women and their unborn children is a larger concern nowadays, as well as the increased consideration as a criteria overall.
The statements we are actually interested in, related to 'childbearing potential' show a slight positive trend for inclusion criteria, supporting our hypothesis. A similar trend is shown for 'contraception' and 'breastfeeding'. This means that the updated guideline to restrict participation for fertile women during early stages has had a positive effect. Performing a Friedmann test for statistical significance, we get a p-value of 0.0003, low enough to reject H0. We can therefore say, that there is a change over the decades.

B. H2

To verify this hypothesis, we wanted to look at participant numbers, split by gender. However, the metadata does not include this information. We would need to look through study documents, crawling through each entry with a dedicated automated task. We tried doing so, but concluded that the available sources are very sparse and heterogeneous. They are sparse, as sponsors often want to keep results and specifics under disclosure, for profit and to protect their copyrights. They are heterogeneous, as no documents are alike, they all come in pdf format, with tables that are not always available or structured very differently. This makes it extremely hard for us to process the data and produce accurate results. We concluded that this was not a feasible solution for us. Instead, we looked at sponsors in detail.
Studies have bodies that sponsor them, such as clinics, universities, pharmaceutical companies or sometimes individuals.

There are 18263 unique sponsors among our data. The five biggest ones are:

- (799) National Cancer Institute (NCI)
- (766) Assiut University
- (733) GlaxoSmithKline
- (679) Pfizer
- (672) AstraZeneca

5

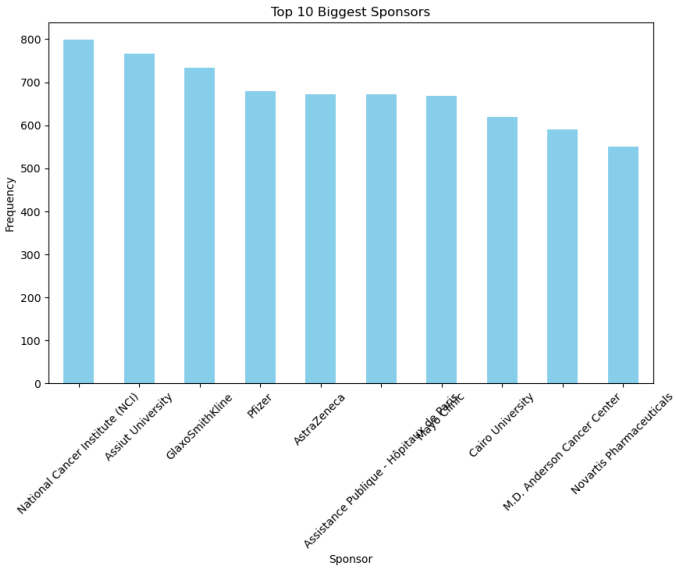


Fig. 5: 10 biggest Sponsors

On average a sponsor has 6.04 studies, the median being 1.0, which means that these big ones have quite a significant power, deviating so far from the mean.
This means that a very small group of sponsors, has control over more than half of all studies. This could potentially be bad, as it may lead to more important studies having less balanced sets of participants, that can be compensated with less important studies participants. We cannot say for sure, whether this is the case, as we do not have access to the data.

C. H3

To assess the third hypothesis, we analysed the topics of studies over the decades. We wanted to know, if studies were conducted on the menstrual cycle, contraceptives and the female body. We specifically looked at study titles and a summary description, although analysing study documents and results could have given even more insight. The words we were looking for were hand picked, based on the guidelines and what we thought was relevant to the topics. Using nlp techniques to find closely related vectors could have enhanced our set of words even further. We preprocessed our data by lemmatising and vectorising words, removing stopwords as well. Our results show a slight upward trend for most words, but they flatten out again, and the trend happens even before the updated 1993 guidelines, making us believe we should reject the hypothesis, as we can't find a correlation with the

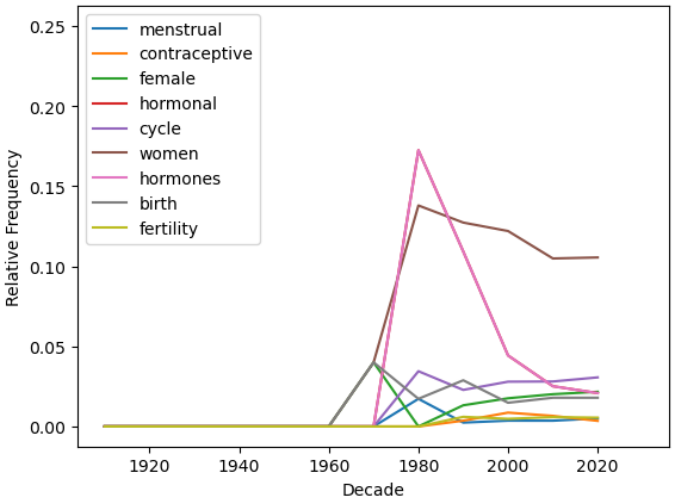


Fig. 6: Relative Frequency of Singled Out Words of Interest Appearing in Study Title or Summary

timeline. However, one word stands out as having a more extreme trend, it flattens out as well, but does not fall again. The word in question is 'women', making us conclude that women are after all more represented now, even if this does not directly correlates with the introduction of new guidelines. Words related to hormones are trending as well, but we can not necessarily pin these to female reproductive hormones only, however, as both follow a similar trend, it is possible to at least predict a possible correlation.
With a p-value of $7.072997015062785 * 10^{-51}$ in a chi-square test, our results are statistically significant. This means, that there is a high probability of an association between the keywords and the decades, signifying a change over time has ocured.

VI. Conclusion

In this research we tried to explore the multifaceted dimensions of gender biases within medical research, focusing on the implementation and impacts of U.S. FDA guidelines on clinical trials. In a detailed analysis based on data acquired from the FDA's ClinicalTrials API, we have highlighted three main hypotheses that examine the changes in clinical trial practices after the changes in FDA guidelines in 1993.
Firstly, our study has shown that there is a more nuanced approach to considering the childbearing potential for women, which represents a positive shift in the direction of inclusivity and safety in clinical trials. This is a positive consequence of increased awareness of the health needs and health conditions of women, which, in turn, have been neglected for many years. Secondly, The real quantitative information about the gender distribution of study participants is still missing, since access to detailed study documents is very limited. This raises a significant gap in transparency and data accessibility, which underscores the ongoing challenges in achieving full equity in medical research.

Finally, our review of study titles and descriptions, over the decades, has seen a steady rise in the inclusion of the terms of health issues related to women, which is not significantly correlated with the timeline for FDA guideline updates. This suggests that while there is an apparent shift in the trend of the research focus towards areas of women's health, the direct impact of FDA policies may not be as pronounced as would be expected.

Based on the results and evidence from this study, it is advocated that further reforms be undertaken in a targeted manner since these gaps are evidently persisting, leading to inequalities in the final health outcomes. Crucially, inclusive guidelines in the future should bring attention to greater transparency and data accessibility in a way that ensures true representation and therefore attends to the needs of all genders. By pushing for these reforms, we aim to contribute to a more equitable healthcare system where medical research and practice no longer reflect the biases of the past but are instead representative of and responsive to the diverse needs of the population.

REFERENCES

[1] L. J. Scheetz and J. P. Orazem, "The influence of sociodemographic factors on trauma center transport for severely injured older adults," *Health Services Research*, vol. 55, no. 3, p. 411–418, Jan. 2020. [Online]. Available: http://dx.doi.org/10.1111/1475-6773.13270

[2] M. Freeman, "Reconsidering the effects of monosodium glutamate: A literature review," *Journal of the American Academy of Nurse Practitioners*, vol. 18, no. 10, pp. 482–486, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-7599.2006.00160.x

[3] R. Brickhill, G. Atherton, A. Piovesan, and L. Cross, "Autism, thy name is man: Exploring implicit and explicit gender bias in autism perceptions," *PLOS ONE*, vol. 18, no. 8, pp. 1–16, 08 2023. [Online]. Available: https://doi.org/10.1371/journal.pone.0284013

[4] R. Parker, T. Larkin, and J. Cockburn, "A visual analysis of gender bias in contemporary anatomy textbooks," *Social Science Medicine*, vol. 180, pp. 106–113, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0277953617301879

[5] U. Food and D. Administration, "Evaluation of gender differences in clinical investigations," FDA Guidance Document, 2024, accessed: 2024-06-06. [Online]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/evaluation-gender-differences-clinical-investigations

[6] ——, "Definitions for clinical trials," FDA Clinical Trials Definitions, 2024, accessed: 2024-06-06. [Online]. Available: https://prsinfo.clinicaltrials.gov/definitions.html

[7] U.S. Food and Drug Administration, "Women in clinical trials: Research and policy," https://www.fda.gov/consumers/diverse-women-clinical-trials/women-clinical-trials-research-and-policy, 2024, accessed: 2024-06-06.