



## Processing and analyzing data on Maternal Health Outcome in Nigeria

**HelpMum:** Dr Abiodun Adereni, Omotooke Afolabi, Esther Isedowo,  
Shakirah Ayinla, Banji Ibukun, Corinne Namblard

**Volunteers:** Talia Bird, Uzma Iffat, Olayile Ejekwu, Caterina Fuligni

# TABLE OF CONTENTS

ABSTRACT _____	3
INTRODUCTION _____	4
Context	4
The Project Partners	4
The Team and the Process	5
DATA PROCESSING _____	6
DATA ANALYSIS _____	7
Question 1: How many pregnant women have you registered in the past one year?	7
Question 2: How often do these pregnant women attend antenatal?	7
Question 3: What are the most common challenges they present or are faced with?	8
Question 4: Did you refer them to the hospital?	9
Question 5: If no, what are the solutions you preferred?	10
Question 6: What was the end result of the situation?	11
1) Which diseases were NOT referred to the hospital and what was the outcome?	12
2) Which diseases were referred to the hospital and what was the outcome?	12
3) Is there a relationship between the number of patients a maternity home sees and attendance of antenatal care?	14
4) Is there a correlation between attendance of antenatal care and cases referred to the hospital?	15
5) How often were the deliveries done safely?	18
6) What were some of the top solutions preferred for common challenges faced?	19
RECOMMENDATIONS _____	22
Data Collection	22
Risk Scoring Models for MMR	22

## ABSTRACT

According to a UNICEF report, Nigeria's 40 million women of childbearing age (between 15 and 49 years of age) suffer a disproportionately high level of health issues surrounding birth. The goal of this report is to illustrate the 3-month long project undertaken by HelpMum, a social enterprise born with the mission to reduce maternal and infant mortality in Nigeria, in partnership with Solve for Good, a platform where social good organizations can post data projects they need help with, with the aim of finding volunteers to work on them. The common goal was to take initial steps in remedying Nigeria's increasing Maternal Mortality Rate (MMR). The data was originally formatted in a Word document and was collected from 203 maternity homes owned by traditional birth attendants across Nigeria. During the project duration, the team of volunteers engaged in activities of data processing, questions-formulation and data analysis. Useful insights were drawn in relation to aspects, such as hospital referrals, challenges encountered, delivery outcomes and treatments. Recommendations for future research are provided at the end of the report.

# INTRODUCTION

## Context

According to a UNICEF report, Nigeria's 40 million women of childbearing age (between 15 and 49 years of age) suffer a disproportionately high level of health issues surrounding birth<sup>1</sup>. While the country has 2.4% of the world's population, it currently accounts for 10% of pregnant mother deaths globally.

As of 2008, the maternal mortality rate (MMR) was 576 per 100,000 live births, the fourth highest in the world, and increased to 814 deaths/100,000 live births by 2015. The Maternal Mortality Rate includes deaths during pregnancy, childbirth, or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, for a specified year. According to the World Health Organization, this was the second-highest number of annual maternal deaths in the world in 2010 and 14% of all maternal deaths globally<sup>2</sup>. The causes of these maternal deaths are pre-eclampsia, infection, post-abortal sepsis, with the leading cause found to be post-partum hemorrhage. This can be in part explained by the persisting low numbers of births occurring in health facilities and the low number of births attended by trained healthcare service providers.

The goal of this project is to take first steps in remedying Nigeria's increasing MMR. In this project, the team processed and analyzed data on births in Nigeria to help formulate practical actions for addressing rising MMR rates.

## The Project Partners

**HelpMum** is a social enterprise born with the mission to reduce maternal and infant mortality in Nigeria. Using the power of mobile technology, Helpmum delivers lifesaving health information to pregnant women on their cellphones. Such information ranges from antenatal care data and immunization sessions reminders, zoonoses awareness and vaccination schedules for newborn babies.

**Solve for Good** is a platform for social good organizations to post data projects they need help with, for volunteers to help scope those projects into well-defined problems, and to help

---

<sup>1</sup> <https://www.unicef.org/nigeria/situation-women-and-children-nigeria>

<sup>2</sup> <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.4197&rep=rep1&type=pdf>

solve those problems. Once the projects are submitted by the organization, they go through a scoping process (done by scoping volunteers and guided by Solve for Good's Data Science Scoping Process). Once a project scope is finalized and reviewed, it becomes available for data science volunteers to start working on. The finished work is reviewed by a QA team consisting of volunteers and staff of the organization that submitted the project and given back to the organization as well as shared publicly on github.

## **The Team and the Process**

The Solve for Good volunteers' team was composed by three data scientists, Uzma Iffat, Olayile Ejekwu, Talia Bird, and a project manager, Caterina Fuligni. On HelpMum's side, the person responsible for the project was Dr Abiodun Adereni, CEO of Helpmum, helped by his collaborators: Omotooke Afolabi, Esther Isedowo, Shakirah Ayinla, Banji Ibukun and Corinne Namblard. Helpmum and Solved for Good signed a Data Sharing Agreement to ensure that the data used for the project was used appropriately. The agreement started on October 1, 2021 and remained in place for 3 months, ending on January 1, 2022. Throughout the project, the volunteers' team met via video-conferencing on a weekly basis. The timeline for the project consisted of three major phases:

- Data processing: 26 Oct - 23 Nov;
- Data Analysis: 24 Nov - 14 Dec;
- Report writing: 15 Dec - 31 Dec

The team met on a biweekly basis with Dr Adereni via video-conferencing as well.

# DATA PROCESSING

The data obtained from HelpMum was originally formatted in a Word document. In order to make the data more tangible for analysis, the text was scraped using the AWS Textract library and placed in a Pandas dataframe. Each question corresponded to a unique feature, resulting in a dataframe with 6 columns and 203 rows, consistent with the 203 maternity homes owned by traditional birth attendants from which data was collected across Nigeria. The following is a list of the questions individuals at each maternity homes were asked:

1. How many pregnant women have you registered in the past one year?
2. How often do these pregnant women attend antenatal?
3. What are the challenges/danger signs they present or are faced with?
4. Did you refer them to the hospital?
5. If no, what are the solutions you proffered?
6. What was the end result of the situation?

After placing the data into a functional structure, each column was modified into a more usable format in relation to the nature of the question it corresponded to. The first column, relating to the first question, was modified to a number form that represented the number of pregnant women registered at each maternity home. Column two and column four were separated using a rule-based system into distinct categorical values. Column two, representing the question “How often do these pregnant women attend antenatal?”, was mapped to values “often” or “not often”. Similar to column two, column four was mapped to values “yes” or “no”. Columns three, five, and six were more complex in nature as maternity homes answers greatly varied. Different processing techniques such as word clouds were used when analyzing data for these columns/questions in the data analysis portion of the project.

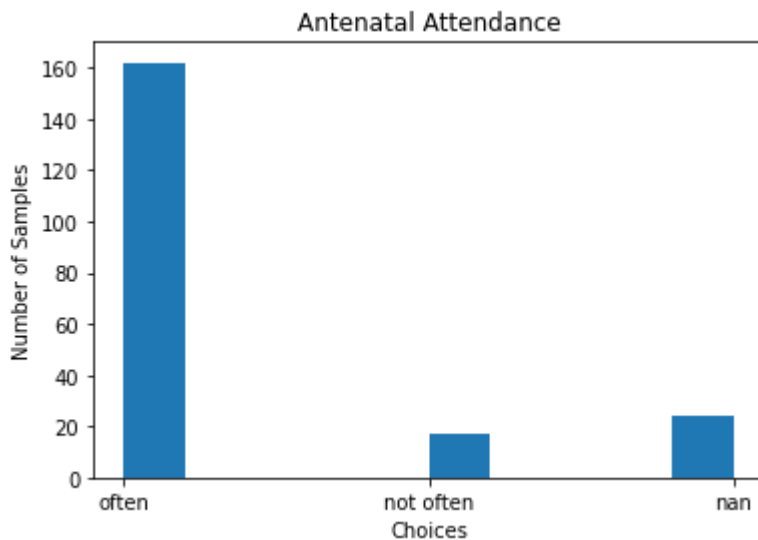
# DATA ANALYSIS

The team proceeded to first analyze each of the six questions of the survey.

## Question 1: How many pregnant women have you registered in the past one year?

Excluding all missing values from the data, the total number of pregnant women registered was 3335 births with an average of 18.324 for each maternity home. However, when considering that 22 maternity homes had missing information on the number of pregnant women registered, accounting for 10.344% of all maternity homes, 3335 births is likely a low estimate. Imputing missing values with the average number of births per maternity home, we get a higher estimate of approximately 3719.808 births.

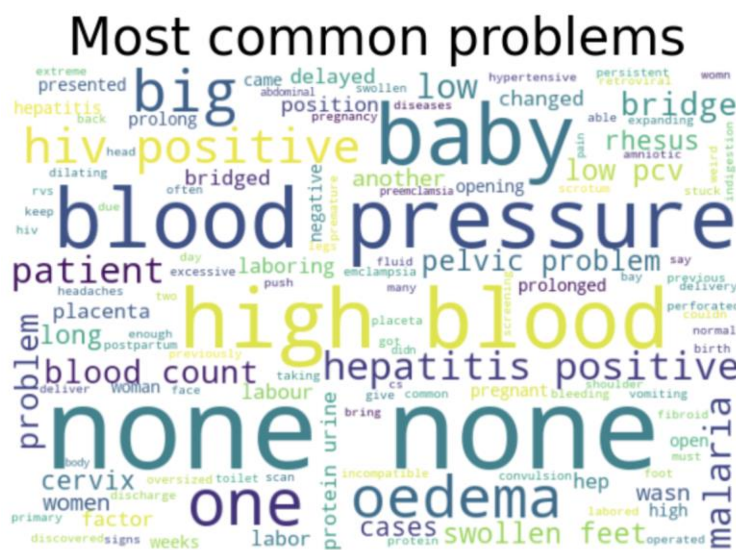
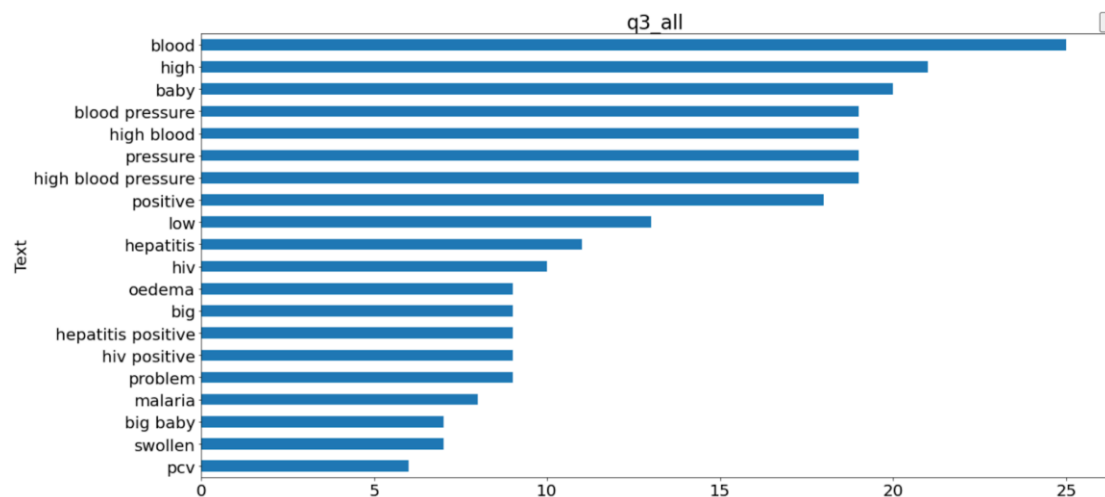
## Question 2: How often do these pregnant women attend antenatal?



The antenatal program is a “pay-out-of-pocket” program provided by HelpMum to mothers in Nigeria. Although the program is optional, it is warmly recommended as it helps detect complications that might hinder births and act upon the issue in advance. A lot of women were referred to antenatal, but many did not attend due to the out-of-pocket costs of the course.

The graph above shows in comparative proportions how often the maternity homes reported that women were attending antenatal care. It is clear from the graph that a very proportion (175/ 203) have reported that women attended antenatal care. The responses collected from maternity homes were in free form text and were matched to a binary choice of often/ not often to the best of the ability.

**Question 3: What are the most common challenges they present or are faced with?**



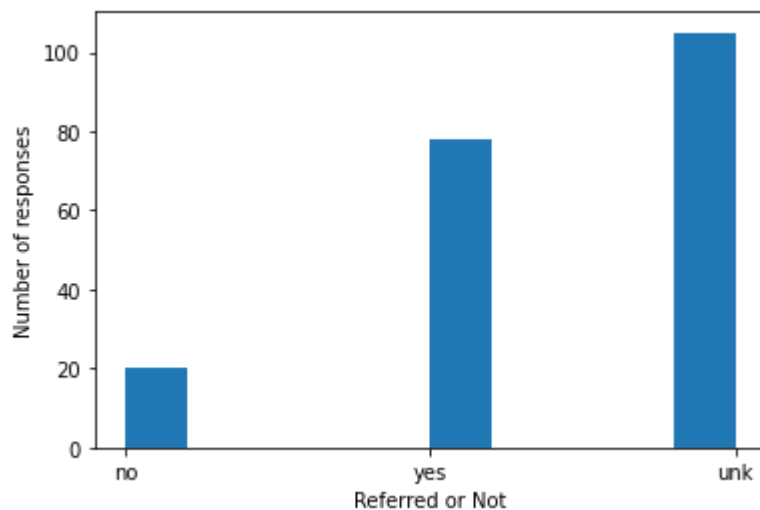
Two different representations of the findings, a graph and Wordcloud.

The graph above and word cloud represent the frequency of occurrence of words using n-



grams which are a sequence of  $n$  words with  $n$  being one, two or three. The top 20 most frequent words in 1,2,3 gram analysis. The frequency of occurrence showed that 'high blood pressure', 'hepatitis', 'hiv', 'oedema' and 'malaria' are challenges that occur most frequently in the text. The word cloud correlates with this showing similar results with the bigger texts representing higher frequency of occurrence of the word.

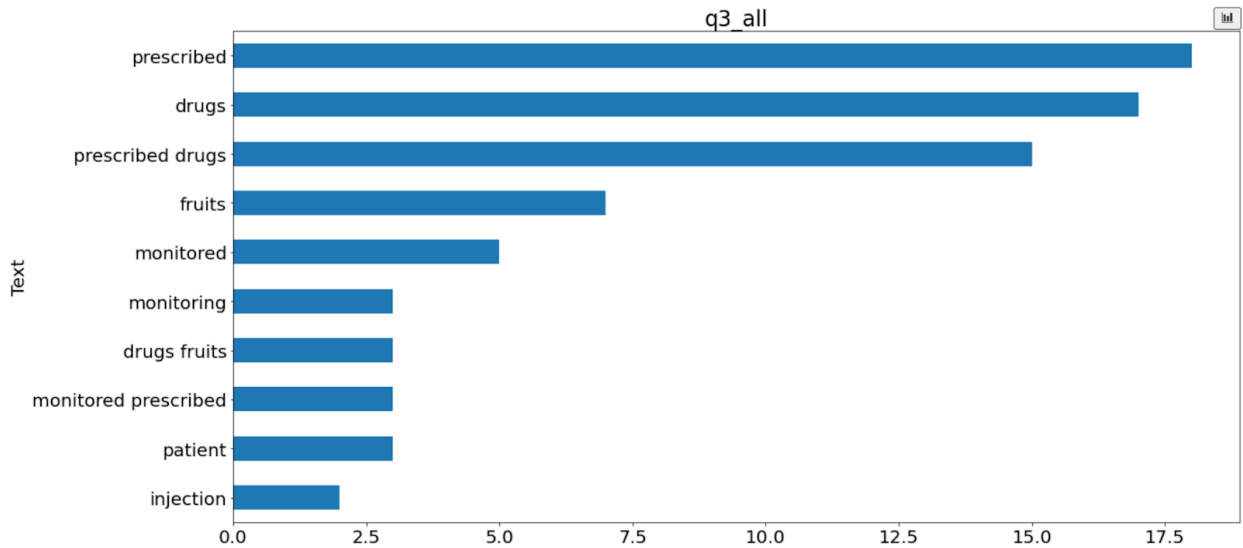
#### Question 4: Did you refer them to the hospital?



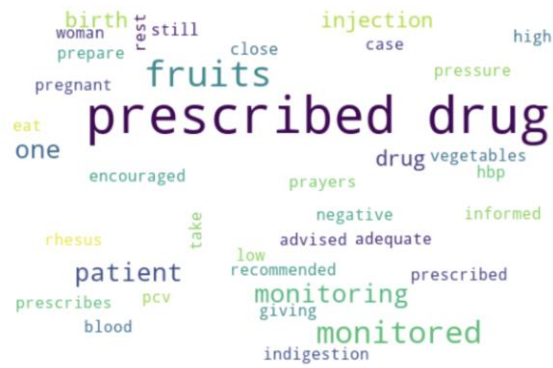
The above graph visualizes the response received from maternity homes when posed with the question 'Did you refer them to the hospital?' The responses were typically in a yes or a no. As clear from the graph, a large proportion (close to 50%) of the respondents chose not to answer this question (marked as 'unk').

For a few respondents who did not provide an answer, the answer was assumed to be 'No', if they had answered the follow up question of 'If no, what are the solutions you preferred?' Out of the respondents who did provide an answer, close to 80% said 'yes'

### Question 5: If no, what are the solutions you preferred?

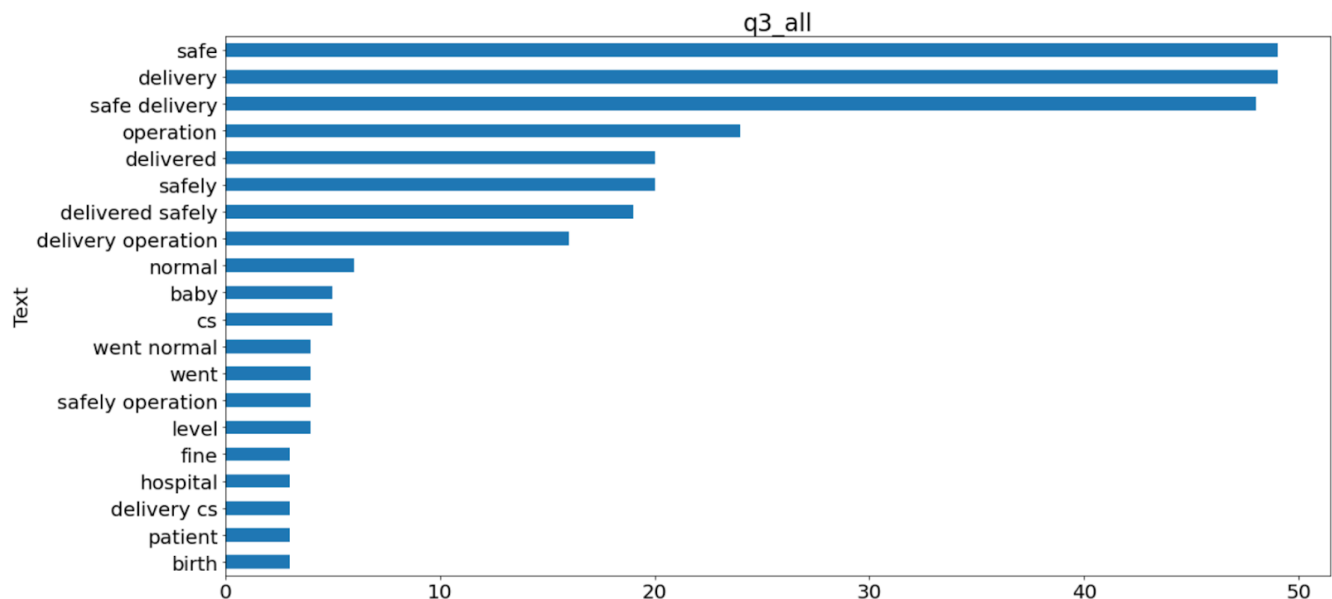


### Most common solutions



Although the text results in this answer were sparse, N-gram analysis was also used in this figure and showed that 'prescribed drugs', 'fruits', and 'monitored' were the most frequently used words to describe the solutions that were preferred. The word cloud also shows the same results.

### Question 6: What was the end result of the situation?



## Most common outcomes



The results showed that most reported 'safe delivery', 'operation', 'normal' as the end results. The word cloud also shows the same results.

**After the six questions have been analyzed, the team, along with Dr Adereni, proceeded to come up with several questions to deepen the analysis.**

## 1) Which diseases were NOT referred to the hospital and what was the outcome?

	Text	count	percentage
<b>Diseases not referred to hospital</b>	low pcv	5	29.411765
	malaria	4	23.529412
	low blood count	4	23.529412
	high blood pressure	3	17.647059

Note: Percentage value is from the number of occurrence of the text in the surveys that answered 'no'

<b>Outcomes of not referring to hospital</b>	delivered safely	7	41.176471
	went normal	4	23.529412
	blood count	2	11.764706
	pcv went normal	2	11.764706
	safe delivery	2	11.764706

The results of the analysis when the patients were not referred to the hospital showed that patients with 'low pcv', 'malaria', 'low blood count' and 'high blood pressure' came up most frequently as not being referred to the hospital. This dataset was quite sparse as 50 % of the data for referrals were unknown. The percentage shows the number of occurrences of the text in the survey when filtered answers for 'no'. The outcome shows high frequency of 'delivered safely', 'went normal' as shown in the table.

## 2) Which diseases were referred to the hospital and what was the outcome?

	Text	count	percentage
<b>Diseases referred to hospital</b>	high blood pressure	16	20.512821
	hepatitis	11	14.102564
	hiv positive	9	11.538462

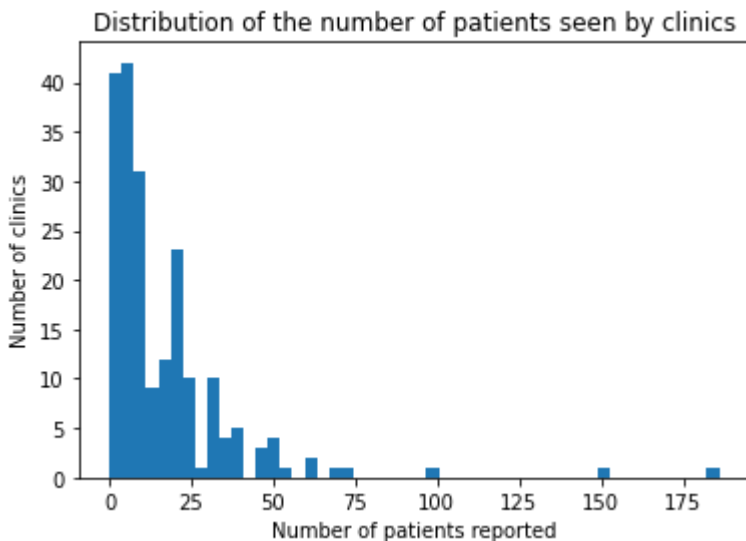
	hepatitis positive	9	11.538462
	oedema	8	10.256410
	big baby	7	8.974359
	swollen feet	6	7.692308
	pelvic problem	6	7.692308

Note: Percentage value is from the number of occurrence of the text in the surveys that answered 'yes'

<b>Outcomes of referring to hospital</b>	<b>Text</b>	<b>count</b>	<b>percentage</b>
	safe delivery	44	56.410256
	operation	24	30.769231
	delivered safely	11	14.102564
	cs	5	6.410256
	birth	3	3.846154
	hospital	3	3.846154

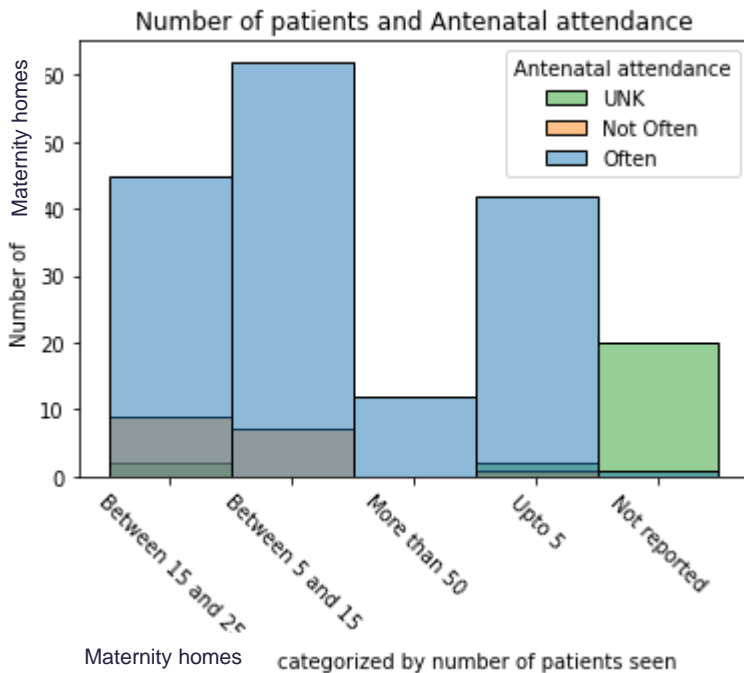
The results of the analysis when the patients were referred to the hospital showed that patients with 'high blood pressure' , 'hepatitis' , 'hiv positive' and 'oedema' were more likely to be referred to the hospital as they came up more frequently in the surveys that answered 'yes'. The outcome of referring to hospital shows high percentage occurrences of 'safe delivery', 'operation' 'cs'.

### 3) Is there a relationship between the number of patients a maternity home sees and attendance of antenatal care?



The figure above shows the distribution of the number of patients reported by the maternity homes. A large number of these reported seeing up to 25 patients. Very few maternity homes (around 10) reported seeing more than 50 patients.

Maternity homes seeing a larger number of patients can be indicative of the size and strategic location (Larger and more urban maternity homes could possibly see more patients). So it can be interesting to see if there is a relationship between the antenatal care attendance and number of patients a maternity home sees, such as smaller maternity homes having lower antenatal attendance.



In the graph above, each bar represents a category of maternity homes based on the number of patients reported. Maternity homes which reported 0 or no patients are categorized as 'Not reported'. Each bar for the maternity home category shows a further division (horizontally) to show the proportion of attendance frequency as reported by the maternity homes.

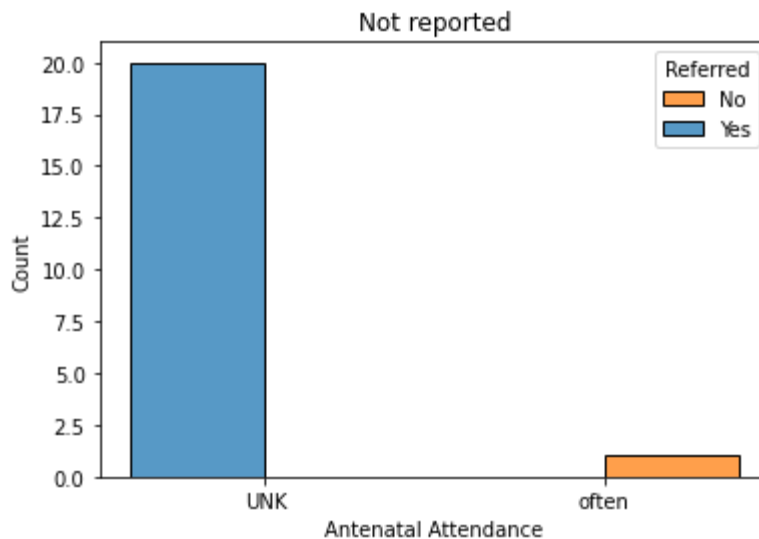
- All maternity homes with more than 50 patients reported antenatal attendance frequency as 'Often'.
- In maternity homes seeing 5 to 15 and 15 to 25 patients, about 20% of the homes reported that patients do not attend antenatal care often.

#### **4) Is there a correlation between attendance of antenatal care and cases referred to the hospital?**

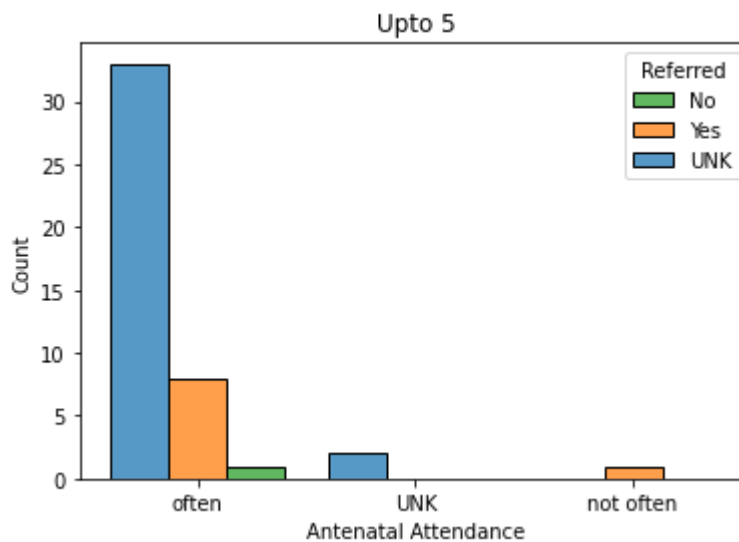
It could be hypothesized that the more often patients attend antenatal care, the lesser the cases referred to the hospitals. Since the data does not report the number of patients attending antenatal care and the number of patients getting referred to the hospital, correlation between the two cannot be measured in the traditional sense. However, the data can be visualized to see if any relationship exists.

The data of antenatal attendance frequency and hospital referrals are visualized for each category of maternity home. These categories are based on the number of patients reported by the maternity home.

1. For the **maternity homes which did not report the number of patients**, the antenatal attendance frequency is mostly unknown as well. However, almost all of such maternity homes reported patients being referred to a hospital as 'yes'. The missing data makes this case difficult to analyse.



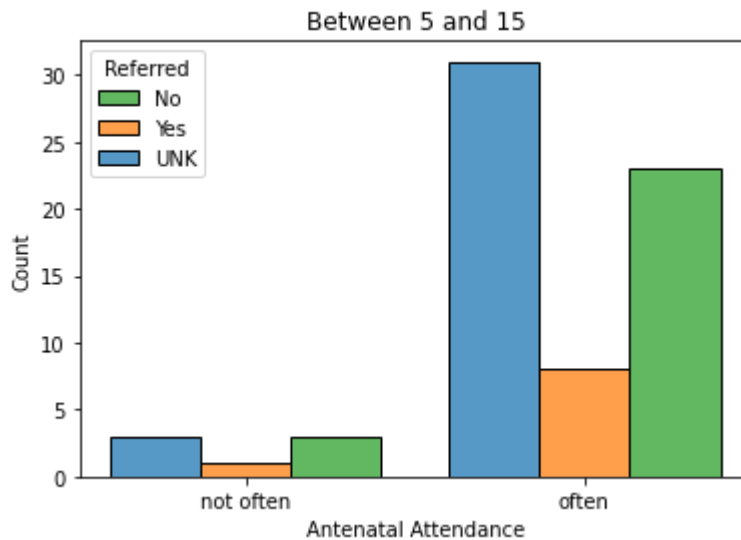
2. For **maternity homes which reported up to 5 patients**, most of them also reported as patients attending antenatal care often. But for the majority of these maternity homes, the status of being referred to a hospital is unknown. Among maternity homes which do report on the status of being referred to hospital, most maternity homes reported as 'Yes, referred to hospital'.



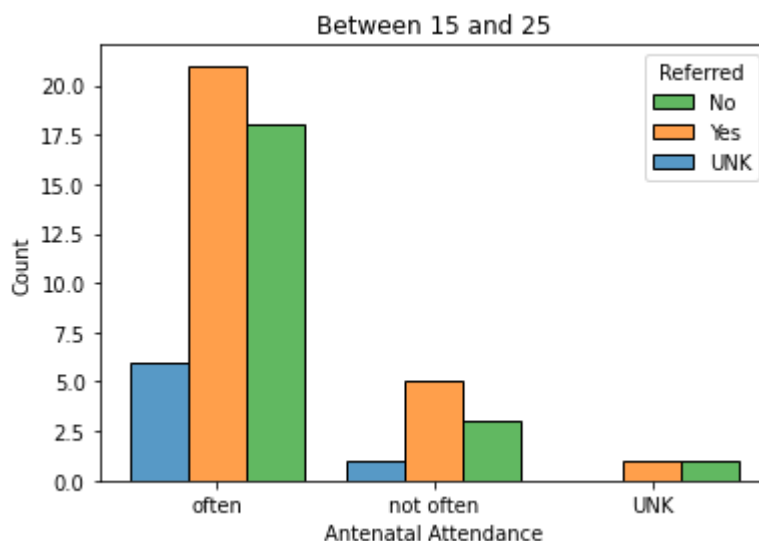
3. For the **category of maternity homes which saw between 5 to 15 patient counts**, the majority of them reported as not referring patients to the hospital.



The percentage of 'no' referrals is higher in the group attending antenatal care, as opposed to the group not attending often.

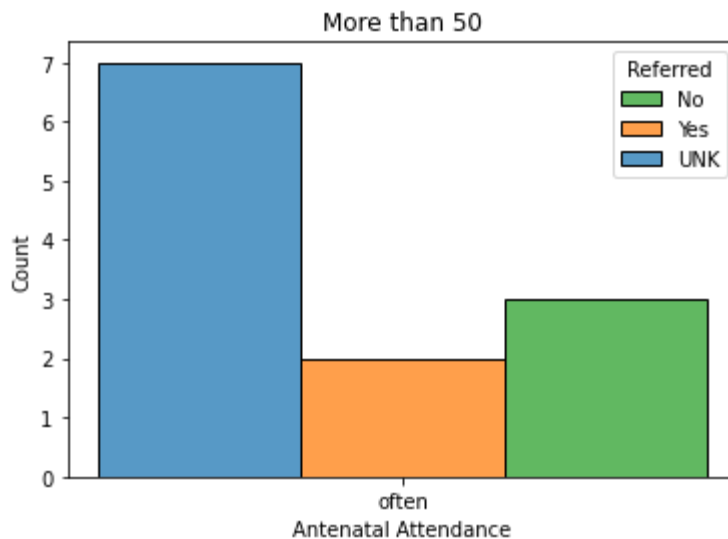


- For the **maternity homes which saw between 15 to 25 patient counts**, the percentage of homes reporting as having referred patients to the hospital is slightly higher than homes reporting as not. This could be justified due to the higher number of patients seen by the maternity home, which would raise the likelihood of the home having at least one patient referred to the hospital.



- The last **category of maternity homes are the ones which see more than 50 patients**. Among these homes, only 5 have reported if they referred patients to

the hospital or not, 3 saying 'No' and 2 'Yes'. No conclusions can be drawn from so few data points.

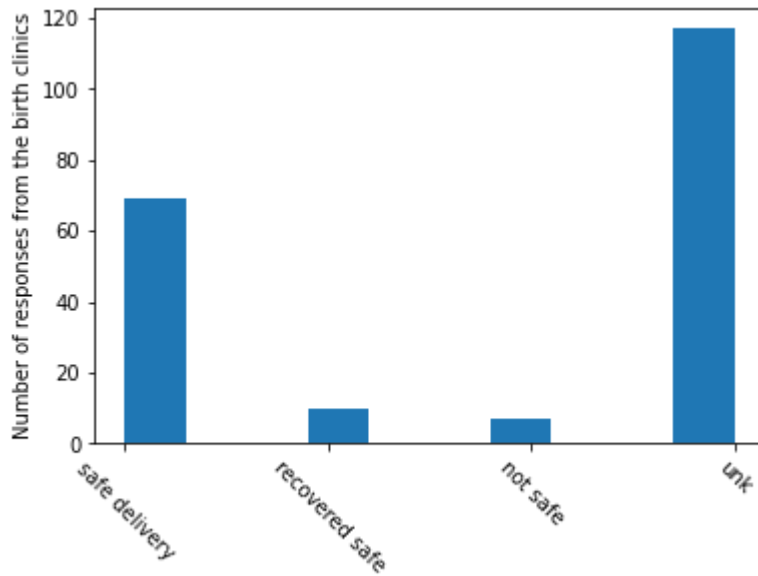


## 5) How often were the deliveries done safely?

The end result reported by the maternity homes summarized the outcome for the patients the homes handled. The responses indicated if the deliveries were safe or not. It also specified if the patients who were facing issues were able to recover and deliver safely.

Therefore, the open-ended responses of the maternity homes were mapped to three broad categories to better understand the data:

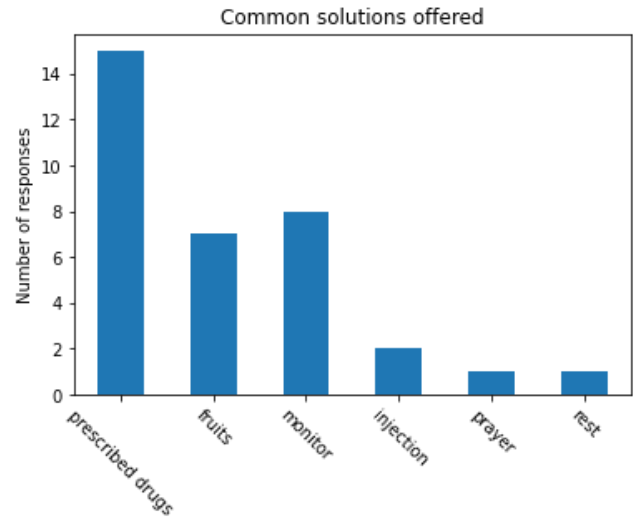
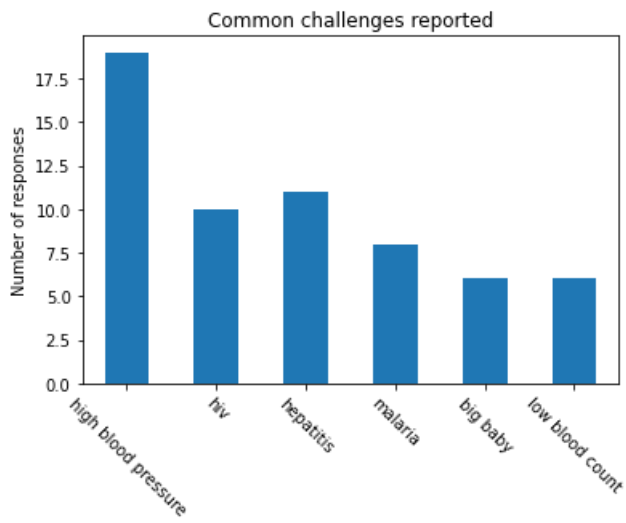
1. safe delivery - patients delivered safely
2. recovered safe - patients recovered from the condition and delivered safely
3. not safe - the delivery was not successful
4. unk - no response



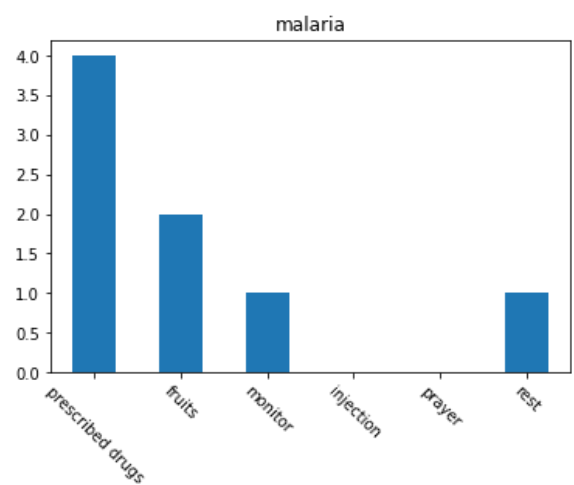
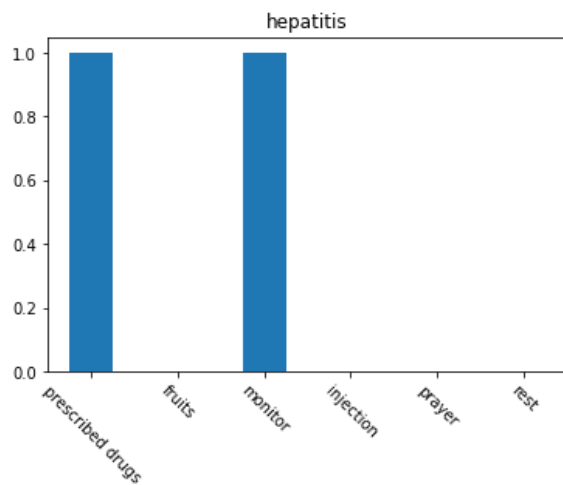
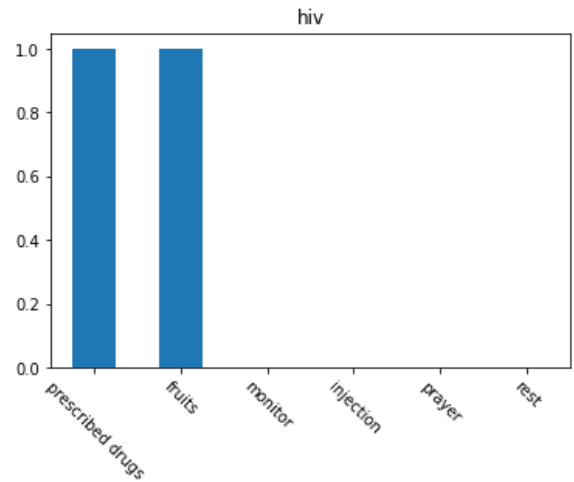
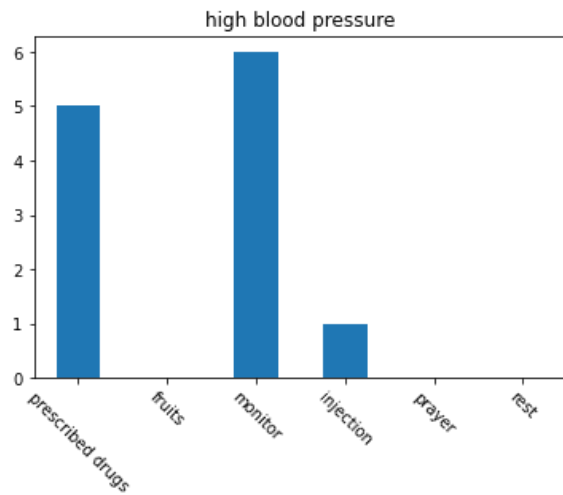
## 6) What were some of the top solutions preferred for common challenges faced?

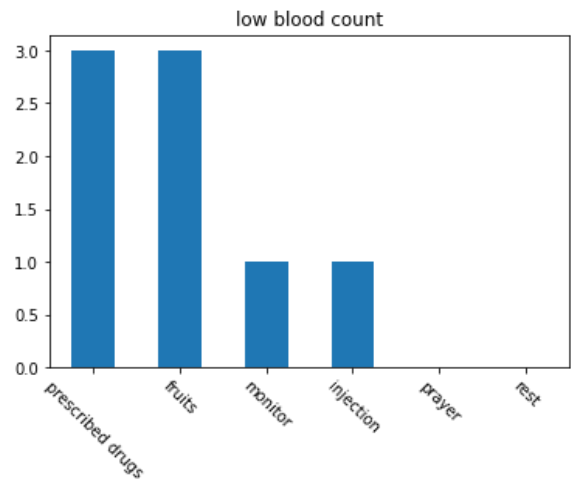
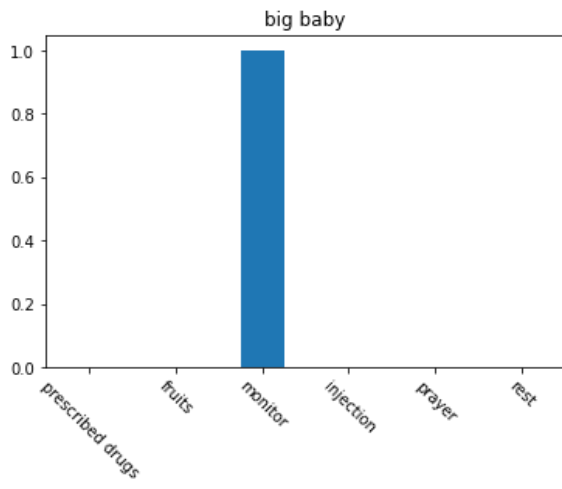
The following graphs show the top few challenges and solutions reported by the maternity homes.

The height of the bar for each challenge/ solution indicates the number of responses which mention it.



It is interesting to see which of the solutions were preferred for each of these challenges by the maternity homes.





# RECOMMENDATIONS

## Data Collection

The data collected from the maternity homes by the volunteers is based on 6 questions. Most of these questions in the current data collection effort are open-ended or broad, inviting a variety of answers. This results in problems in the data collected. **A properly designed survey based questionnaire with standardized responses to choose from will help in ensuring that the data is valid and accurate and make it easier to pull insights.**

**Example** - Question : How often do women attend antenatal care ?

Below are few examples of answers received -

'not regular', 'they do come often', 'yes , do come often', 'yes', 'not all do come often', 'do attend often', 'some do come often', 'often', 'as often as positive', 'not all of them come regularly', 'very often'

For data analysis, it is desired to match the responses to a few standard options such as 'very often', 'often', 'not often', 'not at all'. The ambiguity in the free form answers make this a very difficult task, often dependent on the analyst's decision, and therefore introducing certain biases.

**A better approach would be to pose this question with a set of choices enlisted in the survey.**

## Risk Scoring Models for MMR

Predictive models can be helpful in improving maternal mortality rates (MMR). For example, a model can predict a binary decision of whether a patient would need to be hospitalized or not. Such a model would take into consideration several data points about the patient and output a score between 0 (no need for hospitalization) to 1 (needs hospitalization). This score can then be treated as a risk score associated with the patient.

Predictive machine learning models are data driven. They require a history of data with outcomes for training purposes. Data points describing the pregnancy are called feature

variables. The quality and coverage of the feature variables greatly impact the predictive performance of the model.

The following pointers can be helpful in collecting the right data for such a model:

- If the prediction needs to be done for each patient, the data collected should also be at the patient level;
- Data will have to be collected over the period of the pregnancy, at various stages. For example:
  - Initial set of feature descriptors (collected at the onset of the pregnancy)
    - Patient descriptors - Such as age, weight, number of births
    - Existing medical conditions - Diabetic (yes/no), HIV positive (yes/no), High blood pressure (yes/no)
  - If the patient visits the maternity homes for check-ups, then data pertaining to any issues identified should also be considered. The initial survey data can be used to identify most common issues. For this list of issues, the patient should be evaluated and data recorded.
  - The outcome data such needed hospitalization or delivered safely is crucial for the model
- The data used for the model should be objective and reliable. It could be tempting to include data collected from the patient, for example does the patient follow a certain diet or does the patient attend antenatal care. However, it is quite likely that the patient will not report correctly. This will negatively impact the predictive model training and performance.