A Mathematical Analysis of Unplanned Pregnancies in the United States

Sonia Kopel, Amber Widmer, and Christopher Mullinax, College of Charleston

Abstract— The issue of unplanned pregnancy is among the more divisive in the United States. The mission of this project is to further investigate existing data sets to gain a new insight into this problem. Specifically, this study focused on targeting the causes of unplanned pregnancy and identifying risk groups in order to identify potential interventions that may result in fewer unplanned pregnancies nationwide. Furthermore, survey data was used to gauge how these high-risk groups feel about current public health policy that would likely serve to benefit them the most.

Index Terms— Public health, sexual education, unplanned pregnancy

I. Introduction

Nearly 50% of all babies born in the United States each year are the result of an unplanned pregnancy [5]. This proportion is even higher when considering pregnancies that result in other outcomes such as abortion or fetal loss. As displayed in Figure 1, as of 2010, 17% of pregnancies end in abortion and 90% of these are unplanned pregnancies [6]. A similar amount, 18%, of pregnancies end in fetal loss. We make the assumption that the same rate of unplanned pregnancies occurs in those that end in fetal loss as in those that result in live birth. Thus, we can estimate that 56% of all pregnancies in the United States each year are unplanned.

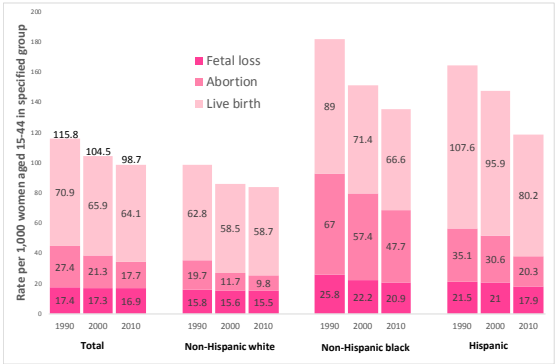

Figure 1: Pregnancy rate per 1,000 women aged 15-44 for the years 1990, 2000, and 2010 broken out by outcomes: fetal loss, abortion and live birth.

While pregnancy rates have decreased over time, the proportion of unplanned pregnancies has remained relatively constant.

Unplanned pregnancies are more common among women at the very beginning of their reproductive life-span as well as those at the very end [10]. Additionally, the instances of unplanned pregnancy are highest in southern and highly populous states [7]. As illustrated by Figures 2 and 3, heatmaps of the unintended pregnancy rates by state compared to poverty rates by state share some notable similarities, particularly in the South. It is well established that unintended pregnancies have higher instances among lower income demographics [9].

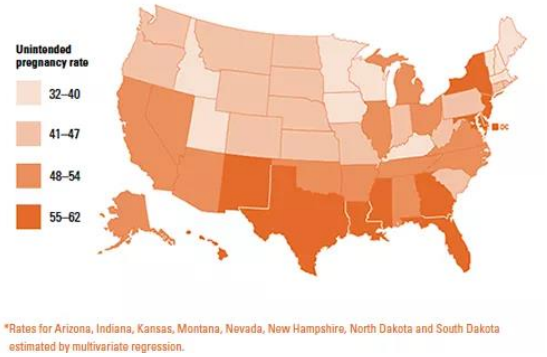

Figure 2: Heatmap of unplanned pregnancies in the United States, 2010.

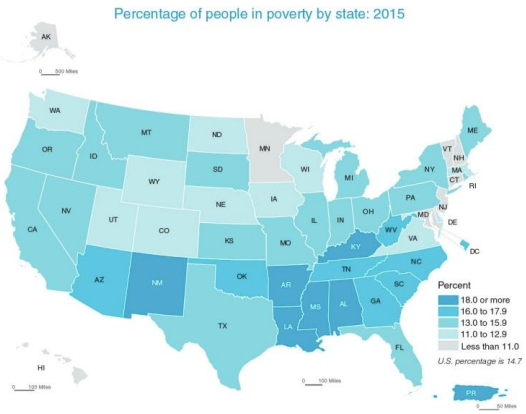

Figure 3: Heatmap of poverty rates in the United States, 2015.

Unplanned pregnancy can endanger a woman's financial stability, educational goals, and personal relationships, as well as pose additional risks. For women in their late 30s or 40s, an unplanned pregnancy could mean accidental subjection to the risks for mother and baby associated with a geriatric pregnancy.

In younger women, the primary risks are related to their financial stability. Of teenage mothers, only 1% graduate from college by age 27 [11]. Young mothers are also more likely to be single parents. If they do get married, these marriages are more likely to dissolve [12]. When a woman has her first child at a young age, she tends to have more children than those who delay childbearing until later [11]. With less education and larger families, these young mothers gain less work experience than their peers without children and are more likely to live in poverty.

Beyond the impact to the women, an unintended pregnancy could have health impacts on the baby and economic impacts on the country.

When a pregnancy is unplanned, the mother has not prepared her body for pregnancy. She may not be taking the vitamins that could help ensure her babies care, partake in prenatal care when she should or when she would like to, and may lack exposure to information that could aid in a healthy pregnancy. Once the baby is born, parental stresses could transfer to the baby. Additionally, if the baby is raised in a single parent home, he or she is more likely to drop out of high school and less likely to complete college than children from the same socioeconomic background who were raised by both biological parents [10].

Finally, each unplanned pregnancy is estimated to cost taxpayers $9,653 on average [13]. Because of the high rate of unplanned pregnancies in lower income individuals, these women are more likely to take advantage of government programs such as Medicaid or WIC. Every prevented instance of an unplanned pregnancy would save taxpayers $4,771 [13].

II. Data

This study analyzed public data gathered from online sources. None of the data sources has personal identifiers which ensured that all responses were anonymous.

A. Pregnancy Risk Assessment Monitoring System (PRAMS)

The PRAMS data is a surveillance system of the CDC which collects population-based data from participating states reflecting maternal attitudes before, during, and shortly after pregnancy. The surveillance currently covers about 83% of all U.S. births. The surveillance includes a questionnaire sent to a random sample selected from birth certificates issued in participating states.

The results of this questionnaire, specifically the pregnancy questionnaire, were used for this study. The years 2009-2011 were included, which are the three years for Phase 6 of the survey. The nature of the data is population-based, as mentioned, and thus includes information about the number of respondents in a subgroup which answered a question in a particular way. There are questions pertaining to the intendedness of the pregnancy as well as many socioeconomic characteristics of the population, so we were able to use this data to identify key differences between the intended and unintended pregnancy populations. However, because the raw data was highly summarized, no person-level analysis could be performed.

As the sample is selected from birth certificates, this data does not contain information on unintended pregnancies which were terminated or resulted in fetal loss.

B. Kaiser 1998

The Kaiser Family Foundation is a nonprofit focused on providing trustworthy non-partisan information on national health issues. We looked at a 1998 National Survey of Americans on Sex and Sexual Health. The survey covered American attitudes on sex education, sexually transmitted diseases, unintended pregnancy, and how people talk (or not talk) about sexual issues with children and partners. This data set has person-level data.

The survey coverage on unintended pregnancy was restricted to opinion-based questions about the potential causes of unplanned pregnancy. These questions generally focused on television, sexual education in schools, and societal norms. Thus, it was impossible to directly identify respondents who had experienced an unplanned pregnancy.

This data set was made available to us through Cornell University's Roper Center.

C. Kaiser 2017

The March 2017 Kaiser Health Tracking Poll: ACA, Replacement Plans, Women's Health was analyzed as well. Like the other Kaiser Family Foundation data set, the March 2017 Health Tracking Poll results were made available to us through Cornell University's Roper Center. This data set contains person-level data as well as demographic information such as education, income, age category, general political affiliation (moderate, liberal, conservative), marital status, and political party affiliation.

We analyzed questions related to pregnancy and maternity care. There were also several questions about Planned Parenthood – awareness of services provided, and support for funding – that were analyzed as well.

III. Methods

Our analysis was, for the most part, exploratory and varied with the format of each analyzed data set. We combined classical statistical tests with exploratory data science techniques. Most of our work was done using free packages from within the R community; however, we also ran Markov simulations in MATLAB. Our code is available in a public Github repository.

A. Variable Correlations

Correlations were used to determine the strengths of linear relationships between two variables in multiple data sets. They were performed on entire data sets to gain a top-level understanding of some of the trends within the data as well as on individual questions to gain a detailed understanding of

what variables were correlated with the proportions of certain responses.

These investigations were all done in R using default and public libraries. Any given correlation value ranges from -1 to 1 with values close to 0 representing weaker correlations (little or no linear relationship), values close to 1 representing strong positive correlations (as one variable increases, so does the other), and values close to -1 representing strong negative correlations (as one variable increases, the other decreases).

1) Full Data Correlations

In order to find questions with correlated responses, numerical weights were assigned to the qualitative responses in the Kaiser surveys. For example, <Strongly Disagree, Somewhat Disagree, Don't Know, Somewhat Agree, Strongly Agree> was translated to <-10, -8, 0, 8, 10>. For responses that covered ranges of values, such as for Age Groups and Income Brackets, the median was assigned. For Income, this was reduced down to the same order of magnitude. With a numerical interpretation of each response, correlations were calculated for each pair of questions/variables in the data set.

The visualizations that summarize these findings were created using the "corrplot" library in R.

2) Answer Level Correlations

Correlations were also done to show the strength of the linear relationships between grouping variables and responses to particular questions. For example, in the PRAMS data, one question asked if the respondent's pregnancy was unplanned with binary responses "yes" and "no." We analyzed how those responses were affected by the grouping variable "income". The incomes of respondents were broken up into several categorical income brackets. All of the brackets were converted into numeric values that approximated their respective central values. For each answer to the question, the correlation between the numeric incomes and the proportion of responses that gave the answer was calculated and analyzed.

In the PRAMS data set, the factor levels <Less than $10,000, $10,000 to $24,999, $25,000 to $49,999, $50,000 and above> were coerced to numeric values <5000, 12500, 37500, 99000>. In the 2017 Kaiser survey, the income brackets <Less than $20,000, $20,000 to less than $30,000, $30,000 to less than $40,000, $40,000 to less than $50,000, $50,000 to less than $75,000, $75,000 to less than $90,000, $90,000 to less than $100,000, $100,000 or more> were coerced to <10000, 25000, 35000, 45000, 62500, 100000, 100000>. The bracket "$90,000 to less than $100,000" was grouped with the "$100,000 or more" category to reduce sparsity and offset the effect of a heavy right skew.

B. Association Learning

In order to gain more insight into the person-level data, we ran a number of association learners from the "RWeka" package. The rules output from the association learner provided insight into how individuals were distributed across multiple groups. The rules also confirmed some of the findings from the correlation investigation.

These rules were of the following form:

$$\Sigma\,(\text{Question}_i = \text{Answer}_j) \rightarrow \text{Question}_x = \text{Answer}_y$$

which roughly translates to the following: if an individual responded to certain questions with particular answers, then she would likely answer Question_x with Answer_y.

The "arules" package was also used which allowed us to specify a certain right-hand or left-hand side of the rule. We applied association learners to knowledge-based questions from the PRAMS data to attempt to identify a profile for the persons who were answering each knowledge-based question incorrectly.

C. Modelling Birth Control Efficacy with Markov Chains

We used a mathematical technique called a Markov Chain to model the probability of various birth control methods failing after a certain number of years. This was done to test the assumption that if one truly wishes to avoid having an unplanned pregnancy, the only truly "safe" bet is abstinence.

A Markov Chain shows how transition probabilities change over time and eventually converge. In order for the method to work, we must have a stochastic matrix, meaning the sum of all entries in any row must equal one.

We simulated a number of environments with varying assumptions. In each environment we tested the same methods of birth control: withdrawal, family planning (FAM) (also known as the calendar method), intrauterine devices (IUD), condoms, injections, the pill, and implants. For all of these methods, the probability of having the birth control method fail within a year was provided by Planned Parenthood [2]. Those probabilities account for user error, so they were selected over the theoretical efficacy of a given method to simulate a more realistic environment in all but one simulation. However, Planned Parenthood only provided two decimals of accuracy, and rounded down the efficacies of the implant and the IUD to 0.99. To be thorough, we performed multiple trials using both a low (0.99) and a high estimate (0.999) Because of the 100% efficacy of abstinence, and the controversy surrounding how realistic this method is to put into practice, abstinence was not included in the Markov Chain. The time step for all Markov Chains was one year.

Probabilities of switching to a birth control method were proportional to the percentage of women who used the given method of birth control.

Once the transition matrices were created, the matrices were then raised to higher and higher powers until they eventually converged. However, because convergence occurs far beyond the fertility period (or even life span) of most women, this study was more interested in the intermediate states of the matrices particularly after five and 30 time steps. Thus, if $\mathbf{T}$ is a transition matrix, $\mathbf{T}^5$ would show how the probabilities changed after five time steps.

All simulations were performed entirely in MATLAB.

D. Population Independence Tests

Statistical tests were performed on data that could be subset into multiple groups to see if there were statistically significant differences between the grouped populations. These tests were all performed in R.

1) Two Proportion Tests

For the PRAMS data, we performed a chi-square test to determine whether there were significant differences between

the responses for each question in the Pregnancy survey for the group of respondents who said their pregnancy was unintended and those whose pregnancy was intended. Because of the nature of this test, this analysis covered only the questions with binary responses (180 of the 220 questions in the survey).

2) Tests for Association

In all of the data sets that were used in the analysis, the majority of the variables were nominal (categorical) in nature with two or more possible values. We tested the variables for independence for each question. The G-test of independence was used to test whether the proportions of one variable are different for different values of a second variable. The G-test of independence and the Chi-square test of independence provide almost identical results. The biggest difference is that the G-values are additive and can be used for more elaborate statistical designs.

Numerous variables were tested in all of the data sets. Those that produced a significant result were then investigated further. To do this, an odds ratio was calculated to describe the magnitude of the difference. For example, in the PRAMS data set we were looking for potential predictors of unplanned pregnancies. There was a noticeable difference between women who exercised three or more time per week versus those who did not exercise regularly. The odds ratio for the two groups was calculated to be 0.69. The interpretation of this value indicated that women who exercised three or more times per week were 31% less likely to have an unintended pregnancy. The results for the odds ratio comparisons could lead to potential paths for further research.

IV. Results

A. Top Level Data Analysis

1) Correlations

Within the Kaiser 98 survey data, there were not a lot of very strong correlations one way or the other, but we see that the block of questions 13a to 17 are frequently correlated with one another. These questions are related to teaching sex education in schools, whether or not certain topics should be taught, and whether high schools should provide condoms to students. Question 14a is the one in this block without a positive correlation, and asks whether high schools should teach students to wait until marriage. Question 3 regarding whether TV should mention safe sex when showing sex is also positively correlated with much of this block. We see a negative correlation between the questions asking whether respondents have kids and whether or not they feel the need to talk more about birth control. Those with kids think there is enough talk about birth control.

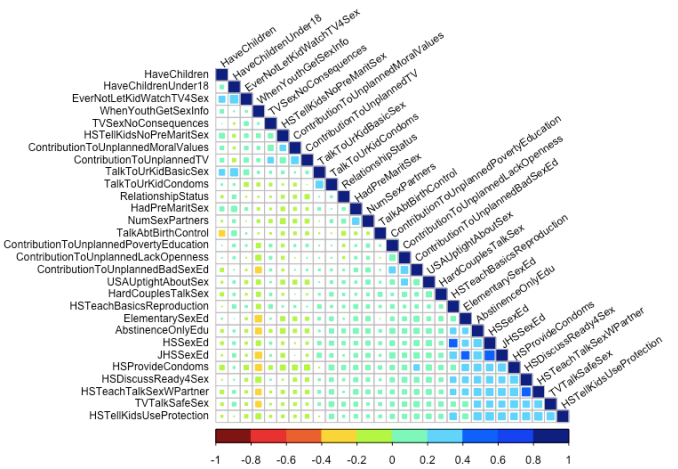

Figure 4: Plot of correlation between variables for the Kaiser 98 data set.

Looking at the correlation plot for the Kaiser data from 2017, we find the strongest correlations to be positive ones, and there tends to be a shared sentiment between being more liberal, Democratic, supporting laws that require private health plans provide birth control at no cost, and supporting Medicaid funding for Planned Parenthood. This group tends to be younger and "less married". They also tend to know well the services that Planned Parenthood provides, but they are not necessarily aware that Planned Parenthood provides abortions.

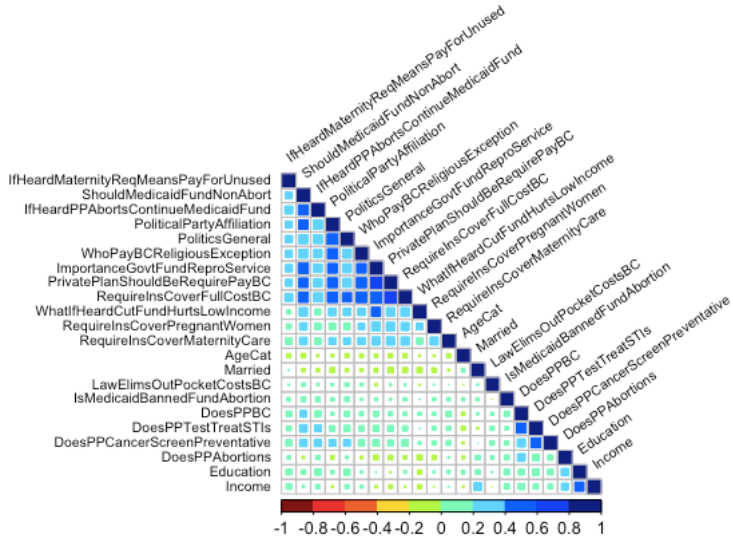

Figure 5: Plot of correlation between variables for the Kaiser 17 data set.

We see even a slight negative correlation between the Democrat respondents and having the knowledge that Planned Parenthood provides abortion services. This may indicate that the young, Democratic support for Planned Parenthood and its funding is coming more from a place of support for easy access to birth control, STI testing and treatment, and preventative care than support for access to abortion services.

B. Association Learning

There were 5 knowledge-based questions in the Kaiser 17 data set:

I. To the best of your knowledge, would you say that Planned Parenthood does or does not provide – Abortions?
II. To the best of your knowledge, would you say that Planned Parenthood does or does not provide – testing and treatment for STIs?
III. To the best of your knowledge, would you say that Planned Parenthood does or does not provide – Cancer screenings and preventative services?
IV. To the best of your knowledge, would you say that Planned Parenthood does or does not provide – contraception including birth control?
V. To the best of your knowledge, is there a ban on federal Medicaid funds being used for abortion?

We found several common characteristics for each question that appeared in the left-hand side of the rules, meaning that these characteristics helped predict an incorrect answer to the corresponding question.

For Question I the strongest rule for answering the question incorrectly was simply Education = less than high school (support = 0.0138, confidence = 0.6957, lift = 3.2958, count = 16). Education appeared in many of the rules, as did age and low income. Individuals with low levels of education, high or low age, and low-income were more likely to incorrectly answer Question I.

The profile for Question II was less clear, but we saw that having been married (states Married, Divorced, Widowed), more conservative politics, semi-low income, and low education were involved in many rules.

Question III had a clear profile – middle-aged (45-54) individuals with conservative politics (PoliticalPartyAffiliation = Republican or PoliticsGeneral = Conservative) were useful predictors for an incorrect answer. This could be explained by considering that these individuals will not need these services through Planned Parenthood, and being Republican or Conservative – are less likely to support Planned Parenthood, thus may have generally less awareness of all the services they provide.

Question IV was largely answered incorrectly by individuals with low income and low levels of education. This is the knowledge-based question which was most frequently answered correctly. The strongest rule was Education = high school incomplete, Income = less than $20,000, PoliticsGeneral = Moderate (support = 0.0035, confidence = 0.5, life = 6.8810, count =4).

Finally, for Question V, young age, low income, and never married helped us predict an incorrect answer. This is our first glimpse at a common theme – that individuals who stand to benefit most from knowledge or support of certain programs are less likely to have that knowledge or support. We've seen that younger women and lower-income women experience higher rates of unplanned pregnancy than the general population. Thus, this population likely also considers abortion at a higher rate. It would be beneficial for them to have knowledge regarding the topic and the politics surrounding at a rate more closely matched with the rate at which it is affecting them.

Table 1: Transition matrix for Sim. 1 Markov Chain with absorbing pregnancy state and random switching after each step

	Withdraw	FAM	IUD	Condom	Injection	Pill	Implant	Pregnant
Withdraw	0.062	0.016	0.125	0.187	0.055	0.320	0.016	0.220
FAM	0.061	0.015	0.122	0.182	0.053	0.312	0.015	0.240
IUD	0.079	0.02	0.158	0.238	0.069	0.406	0.02	0.010
Condom	0.068	0.017	0.136	0.204	0.060	0.349	0.017	0.150
Injection	0.075	0.019	0.15	0.226	0.066	0.385	0.019	0.060
Pill	0.073	0.018	0.146	0.218	0.064	0.373	0.018	0.090
Implant	0.079	0.020	0.158	0.238	0.069	0.406	0.020	0.010
Pregnant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

C. Modelling Birth Control Efficacy with Markov Chains

The first simulation (Sim. 1) assumed pregnancy to be an absorbing state (meaning that once an individual becomes pregnant she is removed from the population). It was also assumed that after each step in the Markov Chain a woman would randomly switch her birth control. The probability of switching to any given birth control method (including the method that the woman was using at the current time step) was proportional to the percentage of Americans who used that particular method of birth control. Those probabilities were calculated from statistics provided by the Guttmacher Institute [3]. The probabilities were scaled to account for the probability of being put in the absorbing pregnant state and to preserve the stochastic nature of the rows. The transition matrix for Sim. 1 is given by Table 1.

Under these assumptions, after five time steps (five years), the probability of having an unplanned pregnancy ranged from 0.354 to 0.504. Within 30 years, approximately the average length of fertility for a woman, the probability of unplanned pregnancy was over 0.95 in all rows. Thus, under these assumptions, most sexually active women would experience an unplanned pregnancy in their lifetimes. Table 2 summarizes the probability of having an unplanned pregnancy for each birth control method from all 4 simulations.

To more closely simulate reality, we ran an additional trial (Sim. 2) that biased the random switching greatly to keeping the birth control method from the previous time step. For all methods, it was assumed that if the woman did not become pregnant, the probability to keeping the same birth control method was 0.99. The remaining probabilities were maintained to reflect their relative popularity, but additionally scaled to keep each row sum equal to one.

In this simulation, after five time steps, the pregnancy probabilities varied greatly with the lowest being at 0.057 for the implant (with the IUD at 0.058) and the highest being at 0.741 for the family planning method. After 30 time steps, the probabilities of pregnancy ranged from 0.374 (implant) to 0.994 (family planning).

To account for the conservative rounding of birth control failure, we did an additional simulation (Sim. 3) in which all assumptions were the same as the previous simulation, but the efficacies of the pill and implant were both assumed to be 0.999.

Table 2: Summary of Markov Chain intermediate phases for n = 5 and 30 for all simulated environment

	Withdraw	FAM	IUD	Condom	Injection	Pill	Implant
Sim. 1, n = 5	0.491	0.504	0.354	0.446	0.387	0.406	0.354
Sim. 2, n = 5	0.705	0.741	0.058	0.553	0.269	0.377	0.057
Sim. 3, n = 5	0.705	0.741	0.015	0.552	0.269	0.377	0.014
Sim 4, n = 5	0.182	0.465	0.006	0.095	0.006	0.007	0.006
Sim. 1, n = 30	0.965	0.966	0.955	0.962	0.958	0.959	0.955
Sim. 2, n = 30	0.993	0.994	0.391	0.981	0.851	0.930	0.374
Sim. 3, n = 30	0.992	0.993	0.206	0.979	0.848	0.926	0.184
Sim 4, n = 30	0.649	0.928	.0625	0.419	0.060	0.074	0.058

Finally, to demonstrate the effects of proper use of birth control on unplanned pregnancies, we ran a fourth simulation (Sim. 4) that assumed perfect use for all of the birth control methods. Methods with efficacies that were reported as greater than 99% were assumed to be 99.9%. The probabilities of switching methods after each time step were calculated in the same fashion as in simulations 1 and 2. After five time steps, the probability of pregnancy ranged between 0.006 (IUD and implant) and 0.465 (family planning). After 30 time steps, the probabilities ranged between 0.058 (implant) and 0.928 (family planning).

The differences in the resultant probabilities of unplanned pregnancy in Sim. 4 as opposed to Sim. 3 highlight the importance of both proper use of birth control and the importance of selecting an effective method. Thus, if a woman wishes to prevent pregnancy, it is imperative for her to select a more effective method like the implant or IUD because, over time, the probability of having an unplanned pregnancy will vary greatly from method to method. Furthermore, the woman must use the selected method properly as this too can greatly impact efficacy as summarized in Table 2. For example, if condoms are used perfectly, the probability of unplanned pregnancy after 30 years is 41.9% which, although high, is significantly less than average use which, after 30 years, is 97.9%.

It is worth noting that the more effective methods of birth control (less than 1% chance of pregnancy after 30 years) such as IUD, injection, pill, and implant, are more difficult to obtain, cost relatively more, and require an appointment with a doctor to obtain which could potentially make them more difficult to access. However, methods like the IUD and implant do not see a large decrease in failure rates between proper use and average use which should make them more desirable.

The consequences of these results are that certain methods of birth control if used consistently are virtually guaranteed to result in many unplanned births given a reasonable amount of time ($\leq$ 30 years). Even for more effective methods, improper use can dramatically increase the probability of having an unwanted conception. Thus, for a woman who wishes to avoid pregnancy, it is very important for her to be using an effective method of birth control, but also to be using it correctly.

D. *Differences Between Unplanned Pregnancy vs Planned Pregnancy Populations*

Significant differences in the answers to nearly every question were found while comparing the differences in responses to the PRAMS questionnaire between the unintended and intended pregnancy populations. As the sample size is large, there are several statistically significant differences that are not large in magnitude. As such, the questions which had the largest absolute valued differences in the proportions of each population which responded a particular way were examined more closely.

It was found that there are four main categories in which the unintended and intended populations differ: contraception use, economic, health/care, and personal relationships.

1) Contraception Use

While only 6% of those intending to get pregnant were using birth control, 49% of those who labeled their pregnancy as unintended were using some form of birth control. Among unintended pregnancies, condoms, withdrawal, and Calendar rhythm methods are cited at much higher rates than the general population. This is in line with the results from the Markov Chain simulations.

Of those who were not using birth control but still had an intended pregnancy, there may not be enough drive to actively prevent pregnancy as 31% did not necessarily mind getting pregnant. Another common reason given for lack of contraception use that occurs more prominently in the unintended group (p-value$\approx$0) is the belief that the respondent cannot become pregnant. 32% of the respondents with unintended pregnancies believed this to be the case compare to only 20% of intended pregnancies.

2) Economics

In the respondents who termed their pregnancy as unintended, the rates of usage of Medicaid for delivery and being enrolled in the WIC program when surveyed were significantly higher. The usage of the programs are indicators of low income. As these are federally funded programs, this serves as evidence that the economic burden on the country for an unintended pregnancy is greater than for an intended pregnancy.

In addition, when a pregnancy was unintended, 31% of the mothers reported they had difficulty paying bills in the 12 months before delivery compared to 17% of mothers with intended pregnancies. Unfortunately, due to the nature of the data, we could not look for direct relationships between income bracket and these variables.

However, because other data sets had economic groupings, financial ties between planned and unplanned pregnancies were investigated further. In order to unify disjoint data sets, an initial insight, however obvious, had to be made into unplanned pregnancies. Using the PRAMS data, we were able to identify a strong, negative linear correlation of -0.940 between income bracket and proportion of women who indicated they had an unplanned pregnancy. This means that the proportion of women reporting unplanned pregnancies decreases with increasing income levels. Because there were only two responses and no missing values, the correlation for planned pregnancies was the same magnitude, but with the opposite sign (0.940). To clarify, "proportion" refers to the percentage of women from a given income bracket to answer the question with the given answer. Figure 6 visualizes the data from which these correlations were calculated.

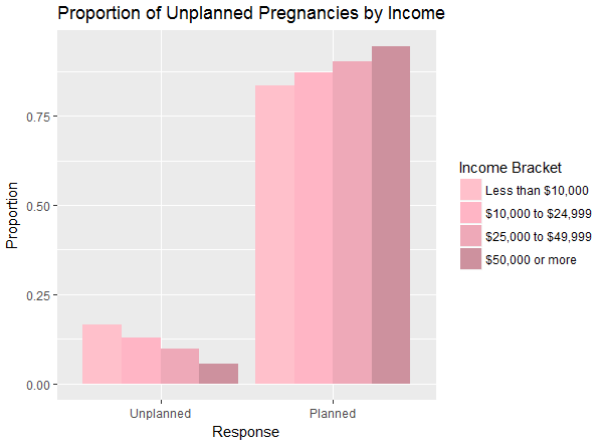

Figure 6: Bar plot of the proportion of planned and unplanned pregnancies by income bracket from the PRAMS data set

With this linear relationship established, we then looked at the March 2017 Kaiser Health Tracking Poll data to extract which questions had answers that were correlated with income. The data set was subset to only look at female respondents. This survey had more income brackets than the PRAMS data, but because we were looking at linear relationships between income and answers, these additional brackets were kept. A total of four questions were identified. Of those four, three were knowledge based.

The first question was opinion based and only looked at a subset of the data. The data was subset only to individuals who supported a law requiring all private healthcare plans to cover the cost of maternity care. That subset of individuals was then asked whether or not they still supported the requirement for private health plans to cover maternity care even if it meant that some people had to pay for benefits they did not use. The correlation between income and the proportion of women from a given income bracket who now opposed this requirement was -0.841. This correlation is strongly negative and indicates that as income increases, the proportion of women opposing such a requirement decreases. This finding is counterintuitive as lower income women are at greater risk of having an unplanned pregnancy and may need those benefits which they now oppose. Because of missing values, the correlation of income and proportion of respondents is not the same magnitude for those who still support the requirement, but they are still similar (0.866). As expected, the correlation is in the opposite direction which indicates a very similar trend. A log-likelihood ratio test confirmed that there were statistically significant differences between the income brackets and the proportions of people indicating they were in support of such a law.

The first of the knowledge-based questions asked the respondent if she believed that there was currently a ban on federal Medicaid funds being used for abortions. The correct answer to this question is "Yes, there is currently such a ban." Among people answering correctly, there was a 0.707 correlation between proportion of respondents from income category and income, meaning that, proportionally, more people from higher income brackets are getting this question wrong. However, the correlation is not as strong among people who answered "No, there is currently no such ban," only -0.290. This large difference is due to higher numbers of people responding that they did not know the answer to the question. The weaker correlation here suggests there is a weaker linear relationship between believing in the wrong answer and income; however, the relationship is still present. Because there were different ways people could fall into the "unknown" categories and because they generally had very few people in them, these "unknown" categories were not analyzed for any of the four questions.

The final two questions dealt with Planned Parenthood services. The first asked respondents whether or not Planned Parenthood provided contraception/birth control. The proportions of people incorrectly answering "No, Planned Parenthood does not provide contraception/birth control" were generally lower than for all the other questions (ranging between 0% and 12.1% among the different income levels). However, despite higher proportions of people answering this question correctly, the responses were still strongly correlated with income. The correlation between the proportion of women answering "Yes, Planned Parenthood does provide contraception/birth control" and income bracket was 0.835, meaning greater proportions of women from higher income brackets were answering the question correctly. Similarly, the correlation was -0.757 among women answering the question incorrectly. Interestingly, when the entire survey was analyzed (not just subset to women), the correlations were a little stronger for both answers, 0.896 and -0.902 respectively. A log likelihood ratio test confirmed statistically significant differences between the groups.

The last question asked respondents whether Planned Parenthood provided abortions. The responses to this question, summarized by Figure 7, varied greatly. We cannot be sure of the cause of these differences, but the strength of the correlations indicates that there is a strong linear relationship between income and the proportion of women to answer this question correctly as well as incorrectly. The correlation between proportion of women responding correctly ("Yes, Planned Parenthood does provide abortions") to the question and income was 0.927 which indicated a very strong linear relationship between income and knowing that Planned Parenthood provides abortions, meaning that higher proportions of women from higher income brackets are answering correctly. Similarly, the correlation among women answering incorrectly was -0.893 which indicates a very similar trend. Again, a log likelihood ratio test confirmed statistically significant differences between the groups.

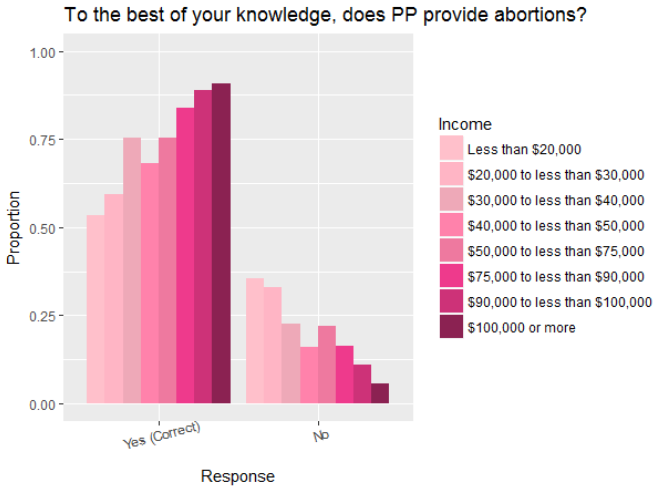

Figure 7: Bar plot of the proportion of women responding either "Yes" or "No" when asked if Planned Parenthood provides abortions grouped by income bracket

To summarize, there exist apparent linear relationships between income bracket and unplanned pregnancy and responses to certain questions to the March 2017 Kaiser Health Tracking Poll. The proportion of women answering knowledge-based questions that relate to pregnancy and contraceptive services is shown to be lower when compared to women from higher income brackets. Seeing as women from lower income brackets are more greatly affected by unplanned pregnancy, this sort of information would be of greater value to them, but for one reason or another, they do not seem to have access to it.

To further analyze the economic impact on the occurrence of unintended pregnancy, the Kaiser 2017 data was used to investigate how women from different income brackets felt about the Affordable Care Act (ACA) and its provisions providing family planning services and birth control to women at no cost. Income tested to be significant (p-value = 0.0057) as a factor determining how a female might answer questions regarding the favorability of the ACA. Surprisingly, women whose income was less than $50,000 per year were roughly 54% more likely to say that the woman herself should pay for birth control.

Two other factors closely related to income, ethnicity and educational level, were also analyzed. Educational level was found to be a significant factor in a woman's response (p-value = 0.015), but women without a college degree were 66% more likely to view the ACA as unfavorable (Odds Ratio = 1.66). This is surprising because less educated women are much more likely to experience an unintended pregnancy. Also, ethnicity was found to be not statistically significant (p-value = 0.223) as a factor influencing a woman's view toward the ACA. Interestingly, black women experience the highest rate of unintended pregnancies when compared to other ethnicities. It seems that the birth control provisions of the ACA would be attractive to both groups of women.

Other questions from Kaiser 2017 were found to be of interest in the study. One of the questions was whether or not congress should repeal the ACA. The only significant factor was educational level (p-value = 0). It was found that women without any college education were 183% more likely to favor a repeal of the ACA. The second question deals with who should pay for a woman's birth control: the government, private insurance companies, or the woman herself. Two factors were found to be of interest. Income was one of the factors that tested to be significant (p-value = 0.0057). Specifically, women whose income was less than $50,000 per year were 54% more likely to say that the woman herself should pay for birth control. The second factor was ethnicity (p-value = 0.002). Black and Hispanic women were 20% more likely to say that a woman herself should pay for birth control. All of these results are surprising. Because of the negative financial impact of an unintended pregnancy, women would benefit greatly from the family planning and birth control provisions of the ACA.

3) Health/Care

When trying to become pregnant, many women will begin taking vitamins to ensure a healthy system in which to cultivate a baby. For unintended births, this habit of taking vitamins is not established, potentially impacting the health of the child. Without having the intent to become pregnant, these women would not have had conversations with their doctors to learn fundamental facts that could help ensure a healthy pregnancy. For instance, only 67% of women with unintended pregnancies had exposure to information that folic acid could prevent birth defects. 83% of women with intended pregnancies had heard this fact.

In the unintended pregnancies many women entered into prenatal care later than desired, or in some cases not at all (56%). If Medicaid is needed to pay for health care during pregnancy, which was found to often be the case for unintended pregnancies, there is a two to four-week period to gain approval on qualification after submitting the required documentation.

4) Personal Relationships

Partner related stressors were twice as likely to be reported in the women with unintended pregnancies (42%) versus intended pregnancies (21%). These women were more likely to argue with their partner more than usual in the 12 months before delivery (34%). This is an added source of stress which has a trickle-down impact on a child.

V. Conclusions

Our Markov Chains have shown that the popular birth control methods cannot fully protect all women from unintended pregnancies. However, some birth control methods are substantially more effective than others. Even with perfect usage, methods like withdrawal, family planning, and even condoms have a very high probability of resulting in pregnancy within 30 years of use. Thus, it is not surprising that nearly half of women who had an unplanned pregnancy reported using birth control at the time of conception. Thus, the issue is not an unwillingness to use birth control, but rather that these women are using less effective methods of birth control or are using them incorrectly.

We have also detected several income related relationships. Lower income women account for the majority of unplanned births, but generally are less likely to favor policies that could offer them assistance in the event that they have an unplanned

pregnancy. This is counter-intuitive especially when considering the fact that they are more likely to experience unplanned pregnancy and are less likely to have the financial capacity to care for themselves as well as a new baby.

Thus, this study concludes that efforts to prevent unintended pregnancy should be targeted at lower income women who do not wish to become pregnant and should center around providing them with effective birth control as well as educating them on proper usage. Additionally, methods like the IUD and implant do not see much difference in efficacy between perfect usage and average usage, so advocating for such methods would likely result in significantly lower rates of unplanned pregnancies.

VI. Future Work

There were many questions about unintended pregnancies that could not pursued because of the nature of the PRAMS data set. Obtaining the person level data of PRAMS from the CDC or beginning a process to collect new data at the person level would open the door for further analysis that was unable to be completed.

Our results demonstrating education to be a large factor in unintended pregnancies open doors for future research as well. There are two primary areas that can be identified. One is how educational material is disseminated to women at high risk for unintended pregnancy. For instance, what would be the optimal routes of getting educational material to as many women as possible? Or what information is being given to these women? There seems to be a chasm between what information is being taught and what factors influence a woman's decision to not use birth control. Providing the right information in the most efficient way possible to the most women possible could aid in reducing the numbers of unplanned pregnancy.

Finally, our results demonstrate that there may be some improper use of birth control and lack of knowledge in selecting the right birth control that would produce optimal results. Determining how to educate and assist women in selecting a birth control method that would provide optimal outcomes for a woman given her individual circumstances is a priority

VII. Acknowledgements

This work would not have been possible without the help and support of Dr. Amy Langville, Abhishek Mehta, Brittany Box, and the entire Tresata team. We would also like to acknowledge the Spring 2018 Operations Research Class at the College of Charleston for their thoughtful input and advice throughout the course of this project.

VIII. Data Sources and References

[1] Centers for Disease Control and Prevention. (2018). Births to Unmarried Women by Age Group: United States [Data File]. Retrieved from https://www.healthdata.gov/dataset/nchs-births-unmarried-women-age-group-united-states

[2] Planned Parenthood. All About Birth Control Methods. Retrieved from https://www.plannedparenthood.org/learn/birth-control

[3] Guttmacher Institute. Contraceptive use in the United States. Retrieved from https://www.guttmacher.org/fact-sheet/contraceptive-use-united-states

[4] What is PRAMS? (2017, August 30). Retrieved April 04, 2018, from https://www.cdc.gov/prams/index.htm

[5] Centers for Disease Control and Prevention. Unintended Pregnancy Protection. Retrieved from https://www.cdc.gov/reproductivehealth/unintendedpregnancy/index.htm

[6] Chae, S., Desai, S., Crowell, M. & Sedgh, G. (2017). Reasons why women have induced abortions: a synthesis of findings from 14 countries. Contraception, 96, 233-241. https://doi.org/10.1016/j.contraception.2017.06.014

[7] Kost, K. (2015). Unintended Pregnancy Rates at the State Level: Estimates for 2010 and Trends Since 2002. Retrieved from https://www.guttmacher.org/sites/default/files/report_pdf/stateup10.pdf

[8] United States Census Bureau. (2016). [Graph illustration heatmap of poverty in the United States 2015]. Poverty in the United States. Retrieved from https://www.census.gov/library/visualizations/2016/comm/cb16-158_poverty_map.html

[9] Guttmacher Institute. (2016). Unintended Pregnancy in the United States. Retrieved from https://www.guttmacher.org/fact-sheet/unintended-pregnancy-united-states.

[10] Institute of Medicine (US) Committee on Unintended Pregnancy; Brown SS, Eisenberg L, editors. The Best Intentions: Unintended Pregnancy and the Well-Being of Children and Families. 3, Consequences of Unintended Pregnancy. Washington (DC): National Academies Press (US); 1995. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK232137/

[11] Moore KA. (1992). Bivariate analyses: Age at first birth and well-being at ages 23, 27, and 35. In The Consequences of Early Childbearing in the 1980s, Final Report to National Institute of Child Health and Human Development, Contract No. N01-HD-9-219.

[12] Hayes C, ed., editor. (1987). Risking the Future. Vol. 1. Washington DC: National Academy Press.

[13] Center on Children and Families at BROOKINGS. (2011). The High Cost of Unintended Pregnancy. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/07_unintended_pregnancy_thomas_monea.pdf

[14] Kaiser Family Foundation. (2011). Kaiser Health Tracking Poll: ACA, Replacement Plans, Women's Health. Retrieved from https://www.kff.org/health-costs/poll-finding/kaiser-health-tracking-poll-aca-replacement-plans-womens-health/