Peer influence in Hysterectomies: A Report Using NFHS5 data

Sravan Kumar Davuluri

Objective of the Report

This report studies whether there is any peer influence on a woman's decision to undergo a hysterectomy using a subsample of National Family Health Survey 5 (NFHS-5) data. Specifically, it examines whether the likelihood of a woman having a hysterectomy is influenced by the hysterectomy status of other women in her Primary Sampling Unit (PSU), who are defined as her peers. The analysis suggests that there is a significant and positive association between a woman's hysterectomy status and the average hysterectomy prevalence among her peers. This tells us peer influence on a woman's hysterectomy decision.

Introduction

A hysterectomy is a surgical procedure in which a woman's uterus is removed. Women who undergo this procedure will no longer menstruate and cannot become pregnant. While there are several medical reasons for undergoing a hysterectomy, this report focuses on understanding the peer influence on decisions on hysterectomy.

The important hypothesis of this study is that a woman is more likely to undergo a hysterectomy if the average rate of hysterectomy among her peers is higher. Here, peers are defined as other women in the same Primary Sampling Unit (PSU). The underlying behavioral assumption is that a higher prevalence of hysterectomy in a woman's social environment reflects a community where the procedure is more socially accepted, less stigmatized, and where there may be weaker or no social norms that discourage hysterectomy. In this way peer group hysterectomy status influences the behavioural attitude of the woman towards her own decision regarding hysterectomy.

The hysterectomy status of a woman is identified in the NFHS-5 data using responses to the question "Had your uterus been removed?", captured under the variable "s253". If a woman has undergone a hysterectomy, the response is recorded as "Yes"; otherwise, it is "No". However, the dataset also contains responses such as "Don't know" and missing values, which raise important data handling challenges.

There are two common approaches to deal with such responses:

1. Treating “Don’t know” and missing responses as “No”, and

2. Dropping all “Don’t know” and missing responses from the analysis.

Each of these have their own limitations. The first approach risks underreporting hysterectomy cases if some women who underwent the procedure choose not to respond. The second approach assumes that missing values are random, but if the non-responses are systematically connected to certain demographic or social group characteristics of women, this could introduce bias into the analysis. These concerns are discussed in more detail in a later section of the report.

Despite these limitations, this report uses these two approaches for further analysis, while having awareness of their potential implications.

Question 1: **Find the average rate of hysterectomy at the Primary Sampling Unit (PSU). Plot the histogram and find the variability of this average over all PSU.**

The average rate of hysterectomy at the PSU level is defined as the proportion of women who reported having undergone hysterectomy among only those who responded to the question (strict definition).

Figure 1

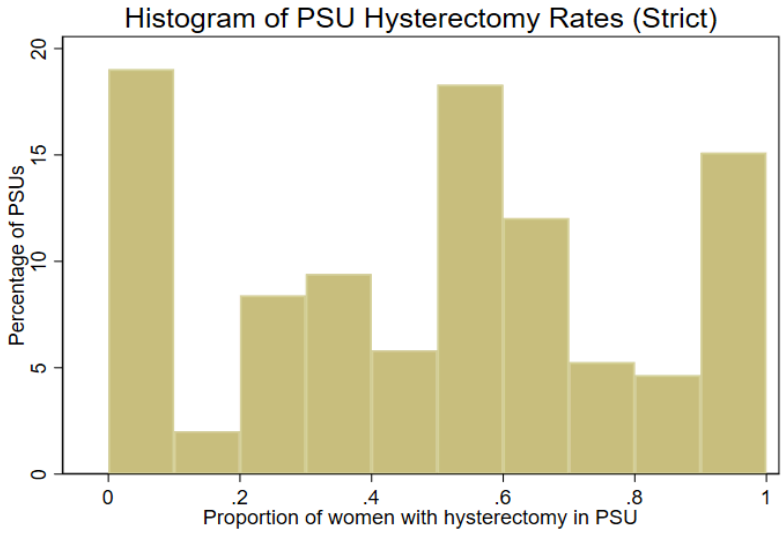


Figure 1 is the histogram based on the strict definition of hysterectomy. It shows the distribution of PSUs by the proportion of women who reported having undergone

hysterectomy. From this we can observe that both ends, that is 0% hysterectomy and 100% hysterectomy having higher frequency of PSUs. The middle values also have higher frequencies. This indicates that in many PSUs, either hysterectomy is highly prevalent or almost absent, rather than evenly spread.

Table 1: Descriptive Statistics of Hysterectomy Across PSUs

Statistic	Value
Mean Hysterectomy	0.476
Variance	0.107
Standard Deviation	0.327

Table 1 shows the mean hysterectomy across all PSU's and the variance of proportions across PSUs. It is showing a higher variance in proportions across PSUs. The figure 1 and Table 1 together indicate the higher variability and concentration of hysterectomy suggesting the possibility of peer influence. Further investigation will be done in the next questions.

Question 2: Define the peers of a woman as the other women in her PSU. Find the average hysterectomy of each woman's peers. Regress the woman's hysterectomy on the average of her peers.

In the previous question, I defined a strict definition of hysterectomy. Here I will define the loose definition of hysterectomy, calculate the average hysterectomy of each woman's peers based on both strict and loose definition of hysterectomy and run logit models based on two definitions.

The average rate of hysterectomy according to the loose definition of hysterectomy refers to the proportion of women underwent hysterectomy among total sampled women, here missing values and "don't know" responses we are treating as women not underwent hysterectomy.

The peer group refers to other women in the PSU other than the woman we are referring to. If there are n women in a PSU, then the peer group for each woman in the PSU is n-1 women. The peer average refers to the average number of hysterectomies that happened among the peers of a woman excluding her own status of hysterectomy. This gives the measure of the average hysterectomy of each woman's peers.

Now, if we regress the hysterectomy status of a woman on the peer average and if the coefficient becomes positive and statistically significant, we can infer that on average, the hysterectomy of a woman is influenced by the hysterectomy of the peers of that woman. To find this influence, I ran the following logit model:

$Hysterectomy_i = a + b.Peer\ Average_i + u_i$

Where,

$Hysterectomy_i$: Refers to whether the woman underwent hysterectomy or not,

$Peer\ Average_i$: Refers to peer average of the woman,

u_i: Refers to stochastic disturbance terms.

I ran two logit models based on strict and loose definitions of hysterectomy. For strict definition, I dropped missing values from hysterectomy status, peer average also calculated after dropping missing values. So a woman with a missing value of hysterectomy status doesn't influence the observed woman's peer group or peer average in this case. The results for the regressions are presented below in table 2 and table 3.

Regression Results

Table 2: Logit Regression of Hysterectomy (Loose Definition) on Peer Average

Dependent Variable	**Hysterectomy (Loose Definition)**
Peer Average	**3.458*** **
	(0.210)
Constant	**-2.680*** **
	(0.0284)
Observations	**38,493**

Robust standard errors in parentheses

***** p<0.01, ** p<0.05, * p<0.1**

Table 2 presents the results of the logit regression where the dependent variable is the hysterectomy status of a woman (as defined by the loose definition, i.e., treating missing or "don't know" responses as "no") and the explanatory variable is the peer average hysterectomy rate within her PSU (excluding her own status).

The regression results show that on average as the peer average increases by one unit, the log odds of a woman undergoing hysterectomy increase by 3.458 units.

To interpret this result more intuitively, marginal effects were calculated. These suggest that on average a one-unit increase in peer average hysterectomy is associated with an increase of 0.273 in the probability that a woman undergoes hysterectomy.

Table 3: Logit Regression of Hysterectomy (Strict Definition) on Peer Average

Dependent Variable	**Hysterectomy (Strict Definition)**
Peer Average1	**1.090*****
	(0.0782)
Constant	**-0.614*****
	(0.0448)
Observations	**6,845**

Robust standard errors in parentheses

***** p<0.01, ** p<0.05, * p<0.1**

Table 3 presents the logit regression results where the dependent variable is the hysterectomy status of a woman (as defined by the strict definition) and the explanatory variable is the peer average hysterectomy rate within her PSU (excluding her own status).

The regression results show that on average as the peer average increases by one unit, the log odds of a woman undergoing hysterectomy increase by 1.090 units.

Marginal effects suggest that on average a one-unit increase in peer average hysterectomy is associated with an increase of 0.264 in the probability that a woman undergoes hysterectomy.

These two regression results indicate a strong and statistically significant peer effect, suggesting that women are more likely to undergo hysterectomy if other women in their PSU have also undergone the procedure.

Question 3: Focus on PSUs where more than one woman has had a hysterectomy. Do women in a PSU tend to get hysterectomies done one after another? Find the average lag in years between two hysterectomies in a PSU.

This question explores whether hysterectomy decisions among women in the same PSU are clustered in time. If they are clustered in time, it suggests that a woman's hysterectomy affects the decision of the other woman's hysterectomy.

To examine this, I restricted the analysis to only those women who reported that they underwent hysterectomy. From this, I further limited the data to only those PSUs where at least two women have had a hysterectomy, this is because the time lag cannot be calculated with only one observation.

To estimate the average lag in years between hysterectomy procedures within a PSU, I use the variable **"**s254**"** in the NFHS-5 data, which reports the number of years ago the woman had her hysterectomy. Then I calculated the average lag using the following procedure:

1. Sort the women within each PSU in ascending order based on the value of "s254", it means the most recent hysterectomy comes first and the earliest comes last.

2. Calculating forward differences:

    - This gives the time gap between the 2nd most recent and the most recent hysterectomy, the 3rd and the 2nd, and so on. This gives lag values for all women in a PSU other than the last woman, that is the earliest one.

3. Calculating backward differences:

    - This gives the gap between the earliest hysterectomy and the second earliest, the second earliest and the third earliest and so on. This gives lag

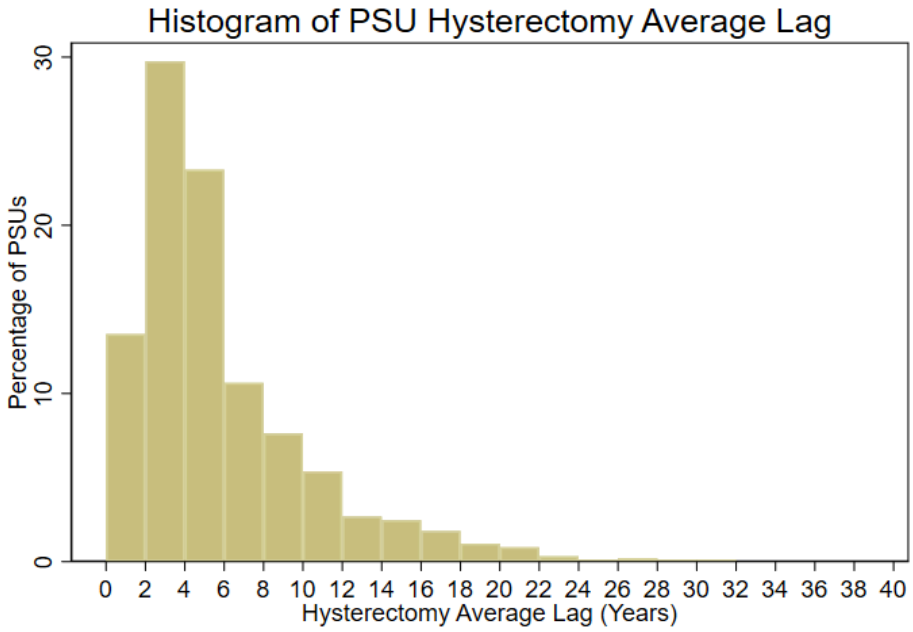
values for all women except the first woman, that is the most recent one.

4. By combining both forward and backward differences, I calculated time lags between hysterectomies for all women.

5. Finally, I compute the average lag in years between hysterectomies at each PSU.

I used the above logic in my STATA code to calculate average lag at each PSU. This method captures whether hysterectomies in a PSU occur in close succession or are spread out over longer periods. If hysterectomies are happening at closer succession, we can infer that there is influence of one woman hysterectomy on the other one. The average lag at PSU is presented below in figure 2 as a histogram.

Figure 2



The above histogram shows that PSUs are clustered at lower lags suggesting hysterectomies happening at closer succession. From this we can infer that the woman who had hysterectomy influenced the other woman's decision of having hysterectomy or not. And women are having hysterectomies one after the other.

Question 4: Compare PSUs where only one woman got a hysterectomy with those where two women get hysterectomies.

- **For these two groups, find the first woman who got hysterectomy and first compare the time she got her hysterectomy done. Are there any differences?**
- **Compare the other women and see if there are differences in terms of their age or education?**

This question aims to explore two key dimensions:

1. Timing of the first hysterectomy case:
   Is there a time difference between PSUs where only **o**ne woman has undergone a hysterectomy versus PSUs where two women underwent hysterectomy?
   If women in the PSUs where only one woman underwent hysterectomy tend to have had the procedure more recently than the first woman in the PSUs where two women underwent hysterectomy, then it might suggest that peer influence may not yet take place in single woman PSUs and it may influence in near future and more hysterectomies may follow in that PSUs in near future.

2. Demographic differences between women in PSUs where two women underwent hysterectomy:
   Are there any differences in age or education leve**l** between the two women who had hysterectomies in the same PSU? This can help us understand in what way the first woman's hysterectomy influences the second woman's decision of hysterectomy. If the first woman had hysterectomy at a younger age and faced any issues, it may help the second woman to delay hysterectomy. The woman influencing the other woman may have common characteristics, they may belong to the same age group and same education level.

Methodology:

1. **Time Comparison**:

   - For each PSU with only one hysterectomy,note the time the woman underwent hysterectomy with the help of variable s254.

   - For each PSU with two hysterectomies, identify the first woman who underwent hysterectomy with the help of the s254 variable and note down the value of s254.

   - Find the average time across two kinds of PSUs and compare time. To find whether the difference is statistically significant do a t-test.

2. **Demographic Comparison**:

    - Take the first and second woman in each PSU based on the time of hysterectomy. The first woman is the one who had hysterectomy the earliest.

    - Find their age, education level, age at which they had hysterectomy and do t-test to find whether the differences are statistically significant.

I used the above logic to do the coding in STATA and the results are presented below**.**

Table 4: Time difference between hysterectomies for first woman in group 1 and group 2

Mean Time Group 1	**Mean Time Group 2**	**Difference**	**Statistical Significance**
8.751	**13.268**	**4.516**	**Yes**

The above table suggests that there is a time difference between hysterectomy of first woman in group 1 (i.e one hysterectomy PSU) and first woman in group 2 (i.e two woman PSU), and the difference is statistically significant. It also suggests that on average group 1 first woman hysterectomy is more recent than first woman hysterectomy in group 2. It is supporting our hypothesis that one hysterectomy PSUs may get more hysterectomies in future.

Table5: Differences in characteristics of First and second woman in group 2

	Mean First Woman	**Mean Second Woman**	**Difference**	**Statistical Significance**
Age of Woman at Survey Time	43	41	2	Yes
Years of Education	2.4	3.4	1	Yes
Age at Hysterectomy	30	35	5	Yes

Table 5 investigates whether there are differences in age, education level and age at hysterectomy for first women and second woman at PSUs where two women had hysterectomy. The table suggests on average the age at hysterectomy of the first woman is lower than the second woman and it is statistically significant. There are differences in education also, but if we bin it into education levels, both fall under the same level on average, so we can conclude that there are no educational differences. There is an age difference also, but the difference is just two years on average, and on average the first woman tends to be older than the second woman.

From this we can conclude that there is similarity in education level from first woman and second woman, their age is very less and first woman tends to be older. On average, the first woman tends to have hysterectomy at a younger age. These factors explain the influence of the first woman on the second woman's decision of hysterectomy. Second woman tends to delay hysterectomies by observing the first woman; the similarities between them in education level and age make them have influence on each other.

Conclusion

This report investigates whether there is peer influence on a woman’s decision regarding hysterectomy. The analysis suggests that there is peer influence and it is significant.

Limitations

As I discussed earlier in the report, the missing values are not random. When I ran a logit model on missing as dependent variable and age, education level, religion and caste as independent variables. The age, education level and being a Muslim shows statistical significance. Younger women and Muslims are tend to not respond to the question on hysterectomy. So, these results contain potential bias.

Appendix (STATA Codes)

```
clear

* Setting up working directory

cd "C:\Users\APF\Desktop\IIMA test"

* My version of STATA by default only supports 5000 variables, so increasing it to 120000

set maxvar 120000

* Loading the NFHS-5 Data Set

use "C:\Users\APF\Desktop\IIMA test\NFHS5.dta"

br



* Answering Question 1

* Looking for which variable corresponds to hysterectomy

* I found that "s253" corresponds to question on hysterectomy

lookfor hysterectomy

tab s253

tab s253, nolabel

* Looking for which variable corresponds to PSU identifier, it is "v021"

lookfor Primary Sampling Unit

tab v021

* Creating variable for strict definition of hysterectomy

gen hysterectomy1 = .

replace hysterectomy1 =1 if s253==1

replace hysterectomy1 =0 if s253==0
```

```
* Creating variable for Loose definition of Hysterectomy

gen hysterectomy2 = 0

replace hysterectomy2 =1 if s253==1

tab hysterectomy2

save "NFHS5.dta", replace

* Finding mean hysterectomy by PSU unit

collapse (mean) hysterectomy1, by(v021)

rename hysterectomy1 hysterectomy_rate_strict

br

* Creating a histogram

hist hysterectomy_rate_strict, width(0.1) percent ///

    title("Histogram of PSU Hysterectomy Rates (Strict)") ///

    xtitle("Proportion of women with hysterectomy in PSU") ///

    ytitle("Percentage of PSUs") ///

    scheme(s1color)

* Saving histogram into my laptop

graph export "C:\Users\APF\Desktop\IIMA test\hysterectomy_strict.png", as(png)
name("Graph")

* Summarizing to find mean and variability across PSUs

summarize hysterectomy_rate_strict, detail




* Answering Question 2

* Modelling for all women sample
```

```
use "nfhs5.dta", clear

* Counting number of women per PSU

bysort v021: gen women_count = _N

* Generating total number of hysterectomies per PSU

bysort v021 (hysterectomy2): gen sum_hyst = sum(hysterectomy2)

bysort v021: replace sum_hyst = sum_hyst[_N]

*  Calculating number of peers undergone hysterectomy by removing woman's own value

gen peer_sum = sum_hyst - hysterectomy2

* Calculating peer average (excluding self)

gen peer_avg = peer_sum / (women_count - 1)


* Handling cases where PSU only has 1 woman (to avoid division by 0)

replace peer_avg = . if women_count == 1

br

* Regression Robust standard errors

logit hysterectomy2 peer_avg, robust

outreg2 using "logit_model_loose_definition.doc", replace ctitle("Logit Regression for loose definition of Hysterectomy")

* Calculating marginal effects

margins, dydx(peer_avg)

* For women who give response for the question on hysterectomy as either yes or no

drop if missing(hysterectomy1)

bysort v021: gen women_count1 = _N

* Generating total number of hysterectomies per PSU

bysort v021 (hysterectomy1): gen sum_hyst1 = sum(hysterectomy1)
```

```
bysort v021: replace sum_hyst1 = sum_hyst1[_N]

* Calculating number of peers undergone hysterectomy by removing woman's own value

gen peer_sum1 = sum_hyst1 - hysterectomy1

* Calculating peer average (excluding self)

gen peer_avg1 = peer_sum1 / (women_count1 - 1)

* Handling cases where PSU only has 1 woman (to avoid division by 0)

replace peer_avg1 = . if women_count1 == 1

* Regression logit model

logit hysterectomy1 peer_avg1, robust

outreg2 using "logit_model_strict_definition.doc", replace ctitle("Logit Regression for strict definition of Hysterectomy")

* calculating marginal effects

margins, dydx(peer_avg1)




*Answering Question 3

* Keeping only women who had hysterectomy

keep if hysterectomy1 == 1

* Counting number of hysterectomies per PSU

bysort v021: gen number_hysterectomy = _N

* Keeping only PSUs with more than one hysterectomy

keep if number_hysterectomy > 1

br

lookfor hysterectomy
```

```
* Sorting data by PSU and year of hysterectomy

bysort v021 (s254): gen hysterectomy_lag = .

* Calculating the lag for women in the same PSU based on years ago hysterectomy performed

bysort v021 (s254): replace hysterectomy_lag = s254[_n+1] - s254[_n] if _n < _N

bysort v021 (s254): replace hysterectomy_lag = s254[_n] - s254[_n-1] if _n > 1

br

* Calculating and storing mean lag by PSU

collapse (mean) hysterectomy_lag, by(v021)

br

* Histogram for average age

hist hysterectomy_lag if hysterectomy_lag >= 0 & hysterectomy_lag <= 40, ///

    width(2) percent start(0) ///

    xlab(0(2)40) ///

    title("Histogram of PSU Hysterectomy Average Lag") ///

    xtitle("Hysterectomy Average Lag (Years)") ///

    ytitle("Percentage of PSUs") ///

    scheme(s1color)

* Exporting histogram

graph export "C:\Users\APF\Desktop\IIMA test\Average lag between hysterectomies.png", as(png) name("Graph")

* Answering Question 4
```

```
use "nfhs5.dta", clear

* Keeping only women who had hysterectomy

keep if hysterectomy1 == 1

* Finding and assigning number of hysterectomies for PSU

bysort v021 (s254): gen hyst_count = _N

* Creating a group for PSUs with only one hysterectomy

gen group_one = (hyst_count == 1)

* Creating a group for PSUs with exactly two hysterectomies

gen group_two = (hyst_count == 2)

keep if hyst_count<=2

br

* Assigning number 1 for first hysterectomy and number 2 for second hysterectomy
based on years ago hysterectomy performed

bysort v021 (s254): gen hyst_order = _N - _n + 1 if _N == 2

* Testing whether there is difference between time of hysterectomy in btween group 1
and group 2

gen compare_group = .

replace compare_group = 1 if group_one == 1

replace compare_group = 2 if hyst_order == 1

ttest s254 if inlist(compare_group, 1, 2), by(compare_group)

* Testing whether there is difference between first women and second women in age

ttest v012, by(hyst_order)

* Testing whether there is education diferrence between first women and second women

lookfor education

ttest v133, by(hyst_order)
```

```
* Testing whether there is difference in age at hysterectomy between first and second women

gen age_at_hysterectomy = v012-s254

ttest age_at_hysterectomy, by(hyst_order)



* Additional excercise for some self clarification

* Testing whether missing value for question on hysterectomy is randon

use "nfhs5.dta", clear

gen miss = 1

replace miss =0 if !missing(hysterectomy1)

lookfor religion

lookfor caste

logit miss v012 v133 i.v130 i.s116, robust
```