I begin by merging the "G START" date variable with data in the "adherence" sheet and creating variables of interest for the analysis. These are:

- *time_trend:* this variable captures the within-panel time trend and controls for the change in the dependent variable over time. The values are a linear trend from the 1 to the length of an individual panel (1,…,n_i).
- *switch:* using the merged "G START" date we create a variable called *"switch"*. All values after the switch date are labeled as 1 indicating use of "plan G" and prior to the switch as 0, indicating use of "plan I".

id	full_adherence	adherence_ratio	total_days	adhere_days	missing_days	g_date	srv_date	switch	time_trend
Id_1	1	1	31	31	0	1-Dec-16	1-Jan-16	0	1
Id_1	1	1	30	30	0	1-Dec-16	1-Feb-16	0	2
Id_1	1	1	31	31	0	1-Dec-16	1-Mar-16	0	3
Id_1	1	1	31	31	0	1-Dec-16	1-Apr-16	0	4
Id_1	1	1	30	30	0	1-Dec-16	1-May-16	0	5
Id_1	1	1	31	31	0	1-Dec-16	1-Jun-16	0	6
Id_1	1	1	30	30	0	1-Dec-16	1-Jul-16	0	7
Id_1	1	1	31	31	0	1-Dec-16	1-Aug-16	0	8
Id_1	0	0.68	31	21	10	1-Dec-16	1-Sep-16	0	9
Id_1	0	0.66	29	19	10	1-Dec-16	1-Oct-16	0	10
Id_1	1	0.9	31	28	3	1-Dec-16	1-Nov-16	0	11
Id_1	1	1	30	30	0	1-Dec-16	1-Dec-16	1	12
Id_1	1	1	31	31	0	1-Dec-16	1-Jan-17	1	13
Id_1	0	0.63	30	19	11	1-Dec-16	1-Feb-17	1	14
Id_1	1	1	31	31	0	1-Dec-16	1-Mar-17	1	15
Id_1	0	0.65	31	20	11	1-Dec-16	1-Apr-17	1	16
Id_1	1	0.87	30	26	4	1-Dec-16	1-May-17	1	17
Id_1	1	1	31	31	0	1-Dec-16	1-Jun-17	1	18
Id_1	1	1	30	30	0	1-Dec-16	1-Jul-17	1	19
Id_1	1	1	31	31	0	1-Dec-16	1-Aug-17	1	20
Id_1	1	1	31	31	0	1-Dec-16	1-Sep-17	1	21
Id_1	1	1	28	28	0	1-Dec-16	1-Oct-17	1	22
Id_1	1	1	31	31	0	1-Dec-16	1-Nov-17	1	23
Id_1	1	1	15	15	0	1-Dec-16	1-Dec-17	1	24
Id_2	1	1	10	10	0	1-Aug-16	1-Aug-15	0	1

<u>Single panel dataset sample:</u>

In the above example, the “G START” date was December 2016 (ignore the “1” on the date, it signifies the first day of the month, but is irrelevant to the analysis), hence the “switch_point” becomes 1 thereafter. The since there are 24 time periods in this panel, the time trend goes from 1 to 24.

Empirical model

The model in mathematical form can be written as:

$$Dependent_{it} = \beta_0 + \beta_1 * time + \beta_2 * switch + \beta_3 * time * switch + \epsilon_{it}$$

Where i represents an individual, t represents time within a panel. The coefficient of $Time$ explains the effect in the pre-switch period and the coefficient of $switch$ explains the effect at the moment of switch (to plan G). Our main variable of interest is **the interaction term because it captures the mean dependent variable change in the post-switch period.**

We run three specifications with different dependent variables to thoroughly examine the effect of the switch. Note that while “full_adherence” is essentially derived from “adherence_ratio”, the two variables are not identical. They both signify separate outcomes. For instance, it is entirely possible that the switch to “Plan G” does not improve full_adherence, but does improve the adherence_ratio and vice versa. Thus, I test them both, along with “adhere_days”.

- Model 1: I use the logistic regression for the first specification with “full_adherence” as a dependent variable, since it is categorical.
- Model 2: “adherence_ratio” as dependent, which is a continuous variable and thus a linear regression model is computed.
- Model 3: “adhere_days” as dependent, which continuous variable and thus a linear regression model is again computed.

Results

Variables/dependent	**Model 1: “full_adherence”**	**Model 2: “adherence_ratio”**	**Model 3: “adhere_days”**
Constant (β_0)	1.614***	0.901***	22.26***
	(0.047)	(0.005)	(0.200)
Time trend (β_1)	-0.091***	-0.016***	0.123***
	(0.006)	(0.000)	(0.029)
Switch point (β_2)	-0.757**	-0.103***	6.319***
	(0.102)	(0.013)	(0.469)
Time*Switch (β_3)	0.110***	0.018***	-0.368***
	(0.008)	(0.001)	(0.038)

a: standard errors in parenthesis.
***,**, *: 1%, 5% and 10% statistical significance respectively.

Interpretation of β coefficients	
β_0	Model 1 (logit): Not meaningful for our analysis. Model 2 (linear model – adherence ratio): Average monthly adherence ratio at baseline is 0.90.

	<u>Model 3 (linear model – adherence days):</u> Average monthly adherence days at baseline is 22.26
β_1	<u>Model 1 (logit):</u> The estimated odds of "full_adherence" increase by $\exp(-0.091) = \mathbf{0.91}$ **each month** all else being equal in the pre-switch period. This depicts, a negative relationship between full adherence and plan I. <u>Model 2 (linear model: adherence ratio):</u> Mean monthly change in the adherence ratio is **-0.016** in the pre-switch period. <u>Model 3 (linear model: adherence days):</u> Mean monthly change in adherence days is **0.12**.
β_2	<u>Model 1 (logit):</u> The estimated odds of "full_adherence" increase by $\exp(-0.757) = \mathbf{0.46}$ **each month** all else being equal **at the switch point.** <u>Model 2 (linear model: adherence ratio) :</u> There is a decline in the adherence ratio **(-0.103**) at the switch point. <u>Model 3 (linear model: adherence days) :</u> There is an increase in adherence by 6.31 days at the switch point.
β_3	Model 1 (logit): The estimated odds of "full_adherence" increase by $\exp(0.110) = \mathbf{1.116}$ **each month** all else being equal **in the post-switch period.** A positive relationship between full_adherence and plan G <u>Model 2 (linear model: adherence ratio) :</u> Similar to model 1, there is an increase in the mean monthly change in adherence ratio by 0.018 in the **post switch period.** <u>Model 3 (linear model: adherence days):</u> There is a decline in mean adherence by less than a day (-0.368) in the **post switch point**.

<u>Summary of results</u>

Our results show that both full adherence and adherence ratio marginally decline month-to-month when patients use Plan I. Even at the switch point, both dependents decline, which is slightly surprising, but from a behavioral perspective this might be explained by sort of learning behavior . **However, in the post-switch period with plan G, both full adherence and adherence ratio increases.** While adherence days marginally declines, quantitatively the effect is negligible and recall that adhere_days only gives a partial picture because it does not incorporate total_days. In conclusion, I was able to discover statistically significant evidence exists that Plan G is more effective than Plan I, at least for the dataset provided.