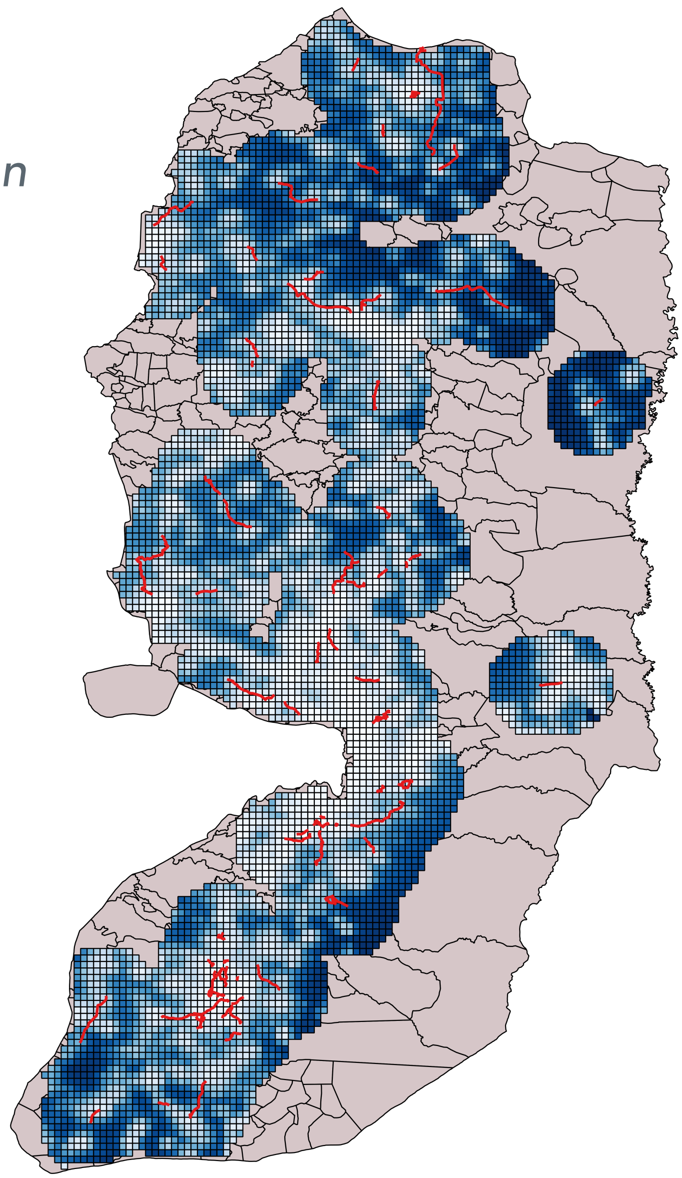
# FINAL REPORT: Evaluation of the Infrastructure Needs Program II



*A geospatial impact evaluation of the effect of INP II road improvements on economic development*

**Ariel BenYishay,**
**Rachel Trichler,**
**Dan Runfola,**
**Seth Goodman**

AIDDATA
A Research Lab at William & Mary

USAID
FROM THE AMERICAN PEOPLE

## Author Information

**Ariel BenYishay**
AidData, William & Mary
abenyishay@aiddata.org

**Rachel Trichler**
AidData, William & Mary
rtrichler@aiddata.org

**Dan Runfola**
AidData, William & Mary
drunfola@aiddata.org

**Seth Goodman**
AidData, William & Mary
sgoodman@aiddata.org



## Citation

BenYishay, A., Trichler, R., Runfola, D., & Goodman, S. (2018). *Final Report: Evaluation of the Infrastructure Needs Program II.* Williamsburg, VA: AidData at William & Mary.

## Acknowledgments

This evaluation was funded by USAID/West Bank and Gaza through a buy-in to a cooperative agreement (AID-OAA-A-12-00096) between USAID's Global Development Lab and AidData at the College of William and Mary under the Higher Education Solutions Network (HESN) Program. The authors would like to acknowledge the research assistance of John Lakwijk and Carey Glenn at AidData and the contributions of Tayseer Edeas, Reem Jafari, and their colleagues at USAID/West Bank and Gaza.

## Executive Summary

AidData and USAID/West Bank and Gaza have collaborated to conduct an evaluation of the Infrastructure Needs Program (INP) II, which funded the construction of water infrastructure, road networks, and schools with the ultimate goal of improving the Palestinian economy. We conduct a geospatial impact evaluation, which utilizes program data provided by the Mission on the timing and location of road improvements and remotely-sensed nighttime lights outcome data to determine the program's impact on economic development.

We employ a quasi-experimental panel framework with monthly and cell-level fixed effects to rigorously estimate program impacts due to completion of road improvements ("treatment"). To do so, we construct a dataset with monthly outcome measures between April 2012 (11 months prior to the first road improvements) and December 2016 (five months after the last road improvements) for 750m square grid cells within 5km of an improved road segment. We compare post-treatment nighttime lights in each cell to counterfactual outcomes obtained from that cell's own preceding nighttime light levels and trends, as well as the outcomes of cells near not-yet-improved road segments. The variation in the timing of road improvements across cells and the inclusion of time and cell-level fixed effects at fine geographic levels address concerns about confounding and omitted variables.

We find a substantive and statistically significant increase in nighttime lights due to INP II. We also find that proximity to multiple improved road segments further increases nighttime lights values and that cells with higher baseline luminosity experience relatively higher gains. These findings seem to point to the importance of improving roads in areas with multiple potential access points to the larger road network and in more urban or densely populated areas. It is worthwhile to both further examine the positive treatment effects and qualitatively understand the larger context of the INP II program in which these road improvements occurred.

## 1. Introduction and Background

*Program Overview*

The Infrastructure Needs Program (INP) II funded by USAID West Bank and Gaza was part of a larger effort to support the Palestinian Authority's High-Impact Micro-Infrastructure Initiative (HIMII). One of several infrastructure projects funded by USAID, it began in 2010 with the goal of developing the water and wastewater infrastructure, road networks, and schools of the region. The project was created with the purpose of improving the access of Palestinians to basic infrastructure needs, in hopes of ultimately improving the Palestinian economy.

While INP II included a number of interventions, this evaluation focuses only on the road improvements funded by the project. The road network component targeted small sections of road (ranging from 1-15 km) scattered throughout the West Bank with the objective of improving connectivity to main roads and the larger road network. The intervention itself varied by the needs of each road segment, and included both new construction and rehabilitation of existing road, such as milling and overlay and adding or improving retaining structures, water drainage, shoulders, intersections, and road signs. This evaluation does not distinguish among various types of improvement, but rather considers any type of improvement as one broad intervention.

*Geospatial Impact Evaluation*

The collaboration between AidData and USAID/West Bank and Gaza began as a result of an expression of interest in geospatial impact evaluation methods. Geospatial Impact Evaluations (GIE) utilize spatial program and outcome data to examine the change in an outcome over time that is attributable to a specific project. It requires data on the timing of project implementation across locations and outcome measures before, during, and after the project, and uses statistical techniques to create a counterfactual, or what would have happened in the absence of the project.

It became clear during our discussions with the Mission that the road improvement component of INP II would lend itself well to a GIE. The Mission had collected spatial data on the locations of road segments and the timing of improvement, which would serve as the "treatment" in the impact evaluation. The road segments were spread throughout the West Bank (rather than clustered in a small area) and the timing of improvements varied within the 4 years of the project (i.e. improvements were made throughout the project period, rather than simultaneously for all roads) – two features that are useful for this type of evaluation.

The Mission was interested in determining the impact of road improvements on economic development. We examined several potential sources of outcome data, looking for data at high spatial resolution (municipality or smaller) and that would include pre-project measures as well as frequent measures throughout the life of the project. Satellite data on nighttime lights, used as a proxy for economic development (see further discussion below), provided a relevant outcome measure at high resolution with monthly measurements before, during, and after the project period for all areas in the West Bank.

The Evaluation Design section details the process by which we merged these sources of existing data to conduct an impact evaluation of INP II.

*Nighttime Lights and Economic Development*

We use remotely-sensed nighttime lights as a proxy outcome for economic development. Many researchers have demonstrated that night lights indicate economic activity (Sutton and Costanza 2002, Ebener et al. 2005, Doll, Muller, and Morley 2006, Sutton, Elvidge, and Ghosh 2007, and Ghosh et al. 2010), and more recently it has been shown to be a strong proxy for GDP and an accurate measure of short-term and long-term economic growth at the country level (Henderson, Storeygard, and Weil 2012). Important to this work, higher levels of luminosity correlate strongly with greater subnational economic activity as well (Michalopoulos and Papaioannou 2014). The availability of nighttime lights data at higher resolution and with greater frequency than traditional forms of household economic data make it possible to conduct this impact evaluation of INP II as the program nears completion.

## 2. Evaluation Design

A Geospatial Impact Evaluation aims to determine a program's impact on an outcome over time, merging multiple sources of spatial data to do so. We employ a quasi-experimental panel framework to assess economic productivity before and after INP II road improvements (or "treatment") for areas within 5km of the improved road segments. The unit of analysis for this evaluation is a 750m x 750m grid cell, which is the unit of measurement for the nighttime lights satellite data. The evaluation period begins in April 2012 and ends in December 2016.

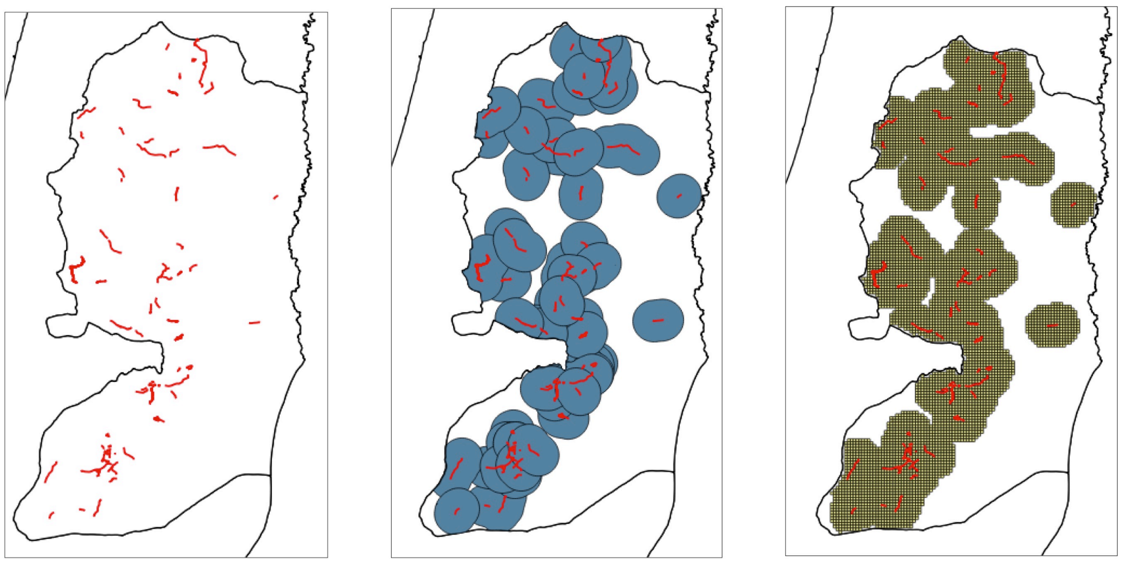
In this section, we first describe the identification of the evaluation sample and the sources of geospatial data used for the analysis. We then explain the quasi-experimental panel methodology in greater detail.

### Data

*Sample*

USAID/West Bank and Gaza provided a shapefile of 59 road segments to be improved through INP II. Of these, improvements to 30 segments were completed by December 2016 and improvements to the remaining 29 segments were completed after 2016. We draw a 5km buffer around each road segment to identify a "catchment area" with the potential to receive economic benefit from the road improvements, and include the 750m square grid cells (from the nighttime lights satellite data) that fall within this buffer (See Figure 1). In some cases, portions of the 5km buffer and its enclosed cells would fall beyond the administrative boundaries of the West Bank. We do not include any grid cells that fall outside of the West Bank's administrative boundaries.

**Figure 1: Creating the Evaluation Sample**



Identify road segments

Create 5km buffer around road segments (within admin boundaries)

Unit of analysis = 750m square grid cells within buffer

Our sample includes 6,729 unique grid cells. Using this sample of grid cells, we construct a dataset that includes monthly treatment and outcome data for each cell from April 2012 through December 2016, the evaluation period. Figure 1 demonstrates that many of the 5km buffers overlap. As such, about half (53%) of the grid cells fall within the buffer of two or more road segments. We address the implications of this on our treatment measure in the next section.

*Treatment Data*

In this impact evaluation, "treatment" occurs when the planned improvements to a road segment through INP II are completed. USAID/West Bank and Gaza provided information on the month and year of completion for improvements to the 59 road segments included in the project. Only 29 of these segments actually had completion dates, as the other 30 segments were completed after December 2016 and outside of the evaluation period. The earliest treatment date is March 2013 and the latest is July 2016.

We consider a grid cell to be treated in the month during which improvements were completed for its associated road segment. Prior to treatment, a cell is part of the control or untreated group. Our sample does include grid cells that only fall within the 5km buffer for one of the 30 road segments scheduled for improvements after December 2016. These cells comprise about 30% of our sample and, as they are never treated, always remain in the control group. The remaining 70% of the cells in our sample move from the control to the treatment group at the time of treatment. This differs from other impact evaluation methods, in which an observation is always part of the control group or always part of the treatment group. We discuss the panel methodology that allows for dynamic treatment and control groups in greater detail below.

It is also worth noting that about half of the grid cells in our sample fall within the 5km buffer for multiple road segments. It is easy to imagine that proximity to two or three improved roads, instead of just one, might have an additional impact on economic development in area. To explore the effects of exposure to multiple treatments, we construct multiple treatment measures in our regression models and we do so chronologically – i.e. Treatment 1 marks the earliest date of improvements, Treatment 2 marks the second earliest (and that a cell is doubly treated), and Treatment 3 marks the third earliest (and that a cell is triply treated). The number of cells that fall into 4 or more buffers is sufficiently small that we do not include any treatment effects beyond Treatment 3. The maximum number of buffers for one individual cell is nine.

*Outcome Data*

We use remotely-sensed satellite imagery of nighttime lights as a proxy outcome for economic development. Nightly satellite images collected by the National Oceanic and Atmospheric Administration (NOAA) Visible Infrared Imaging Radiometer Suite (VIIRS) measure nighttime light activity starting in April 2012 for 750m square pixels. The raw nightly data measure radiance or brightness in a given pixel and are aggregated into monthly composites that are also processed to address background noise and exclude exceptional instances of brightness (e.g. fires). Table 1 provides summary statistics for the nighttime lights values at baseline (April 2012) and endline (December 2016). The unit of measure is not particularly intuitive and the values are best understood in relation to each other (e.g. the median value in relation to the minimum and maximum values, the amount of change experienced by the average grid cell during the evaluation period, etc.).

While earlier nighttime lights products provide annual measures, the monthly intervals of this outcome data lend greater precision as we begin to measure the impact of INP II in the exact month of treatment. The frequency of collection also provides multiple pre-project measures to establish baseline trends for each individual grid cell and multiple post-project measures to detect delayed treatment effects. In our sample, there exists at least 11 months of pre-treatment outcome measures and 5 months of post-treatment outcome measures.

We also utilize another remotely-sensed nighttime lights measure as part of a robustness check. Nightly satellite images collected through the Defense Meteorological Satellite Program (DMSP) provide yearly composite measures from 1km square grid cells between 1992 and 2013.

## Methodology

Our primary estimation of treatment impact employs a panel structure with monthly treatment and outcome data for each of the 6,729 grid cells in our sample. The panel dataset includes monthly data beginning in April 2012 and ending in December 2016. We opt for grid cells as the unit of analysis to improve the precision of our outcome measures and validity of our estimates.

Our panel dataset includes 57 monthly measures (April 2012 – December 2016) of nighttime lights for each of 6,729 cells, or 383,553 observations total. The dataset also identifies the month in which a cell is treated (or when improvements are completed for its associated road segment) and considers a cell treated for every month thereafter. The timing of treatment differs for each of the 29 road segments and thus varies across grid cells. Panel methods essentially align the timing of treatment and aim to identify a pattern of change in nighttime lights for all cells relative to each cell's treatment, e.g. one month after treatment, five months after treatment, etc. Figure 3 provides an example of this process using illustrative data.

The variation in the actual calendar date of treatment across cells helps to address concerns about confounding variables, or other fixed factors specific to each grid cell that may also correlate with changes in nighttime lights. In a panel model at fine geographic scale and with varied timing of treatment, these alternative explanations are very unlikely to begin affecting each cell at the exact same time as INP II-funded improvements are completed. For example, if an employment program is rolled out throughout the West Bank in May 2014, it would likely benefit communities near roads treated late in INP II as well as those treated earlier. The threat to our causal attribution is thus only from other programs or factors that differentially affected newly improved road corridors with the very same timing as INP II. This threat appears quite small given the dispersion of INP II-improved road segments over both time and space. We are thus confident that our empirical approach produces casual estimates that can be attributed to INP II.

Our models also include month and cell-level fixed effects. Month fixed effects control for general temporal trends throughout the evaluation period (e.g. the average nighttime lights values increase between 2012 and 2016, lights become cheaper, or minimum wage increases), as well as changes in nighttime lights experienced by all cells in a specific month (e.g. nighttime light usage is more common in certain months of the year, a massive power outage occurs in a specific month, or a month includes more unpaid holidays which lowers average wages). Cell-level fixed effects control for the time-invariant characteristics of each cell, many of which might also impact nighttime lights values (e.g. proximity to a town or city, population density, proximity to water, slope, or elevation). Controlling for the time-invariant features of each cell is a key advantage of a panel model and allows each cell to serve as its own counterfactual (rather than trying to match similar control and treatment cells with limited data). Applying month and cell-level fixed effects at high levels of spatial resolution helps to control for potential confounds and omitted variables that would otherwise bias our results.

Using the panel framework with fixed effects, we estimate the following equation:

$$Nighttime\ Lights_{imt} = \propto\ + \beta\ Road\ Improvements_{imt} + D_i + D_t +\ \epsilon_{imt}$$

Where $Road\ Improvements_{it}$ indicates whether cell *i* in municipality *m* has been treated by month $t$, $D_i$ is a vector of grid-cell fixed effects, and $D_t$ is month fixed effects. We estimate treatment effects via ordinary least squares. We also use two-way clustering of standard errors by municipality and month.

## 3. Findings

### Main Treatment Effect

Table 2 presents the main model of treatment effects of road improvements on nighttime lights. All models include cell-level fixed effects, which control for all time-invariant features of each individual cell. The pre-treatment measures of nighttime lights included in the panel dataset establish baseline values and trends. We do not include any additional covariates, as time-varying data is limited at such high resolution, and the panel methods with time and cell-level fixed effects help to address concerns about confounding factors (or other variables that might be causing the observed changes in nighttime lights attributed to this program).

Columns 1 and 2 both estimate the effect of a cell's earliest treatment (or the earliest improvements made to a road segment within 5 km), with the only difference being the measurement of time trends. Column 1 includes a linear time effect, which measures the average effect of the passage of one month on nighttime lights, while Column 2 instead includes a more nuanced monthly fixed effect that accounts for the effects of each month individually. Columns 3-5 also include month fixed effects. Column 2 shows a significant and substantive treatment effect. The coefficient indicates that treated cells experienced an increase of 0.341 for nighttime lights measures during the evaluation period.

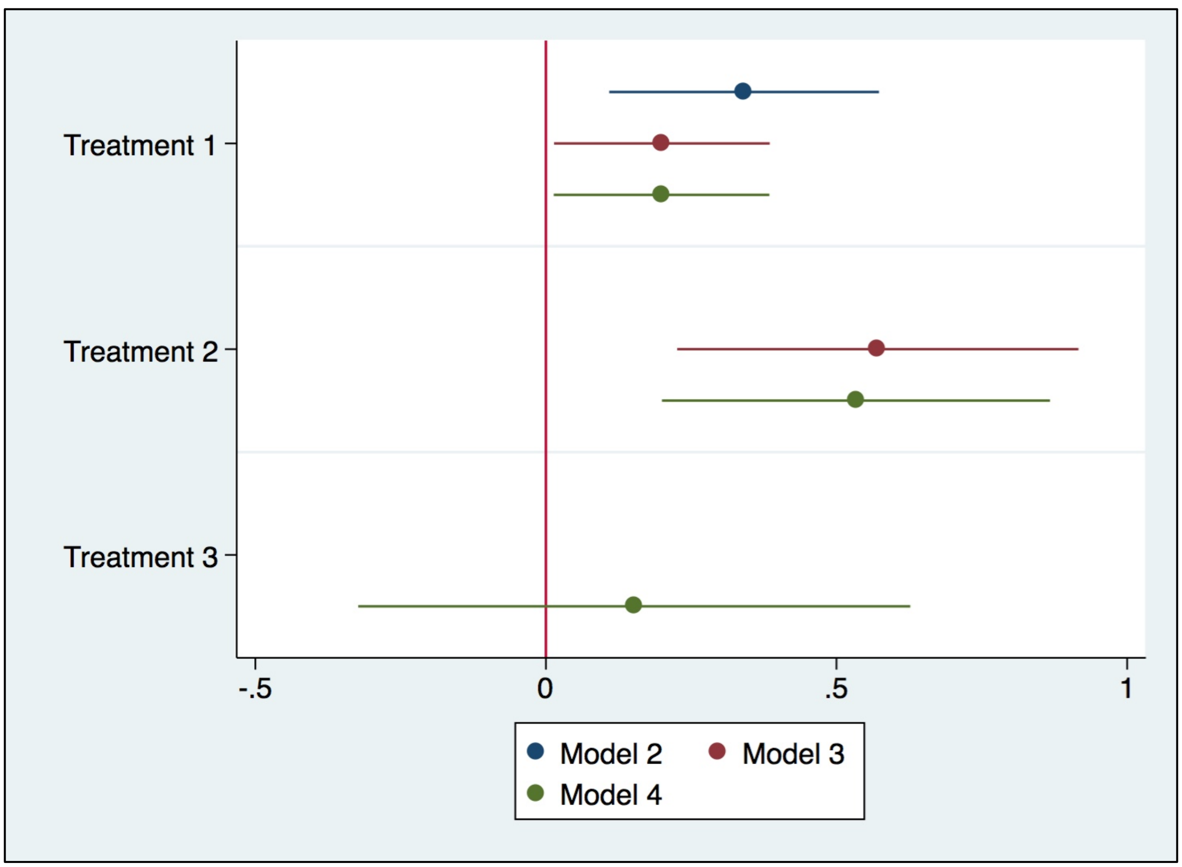
The meaning of a numeric change in nighttime lights is not particularly intuitive, but the median lights values and the average amount of change experienced by all cells during the program shed some light. The median lights value at endline is 3.3 (see Table 1), so a 0.341 increase is roughly 10% of the median value. From baseline to endline, the median value increased from 2.6 to 3.4 (see Table 1). Given this change of 0.8 experienced by the median cell during the evaluation period, a 0.341 increase over that same period is substantive (about 40% of the change for the median cell). To be clear, we cannot attribute 40% of the change over the time period to the road improvements made through INP II. However, a comparison of the treatment effect and the total increase for the median cell during the same time indicates that our estimate of treatment impact is not only statistically significant, but substantively significant as well.

### Secondary Treatment Effects

We next examine the impact of exposure to multiple treatments. About half of the cells in our sample fall within the 5km buffer of multiple road segments and are treated at least twice. In Column 3 of Table 2, we add in Treatment 2, which turns on for the second earliest set of road improvements for a cell. Interestingly, when we add a second treatment, Treatment 1 maintains a statistically significant effect, though the magnitude of that effect lessens (from 0.341 to 0.2) compared to our earlier model. Treatment 2 demonstrates a statistically significant effect that is even higher than the main treatment effect in our initial model, with a coefficient of 0.571. Returning to the context provided by the median cell, for the median nighttime lights value of 3.3 at endline (see Table 1), a 0.571 increase is equal to 17% of the total lights value (compared to 10% of the total value when only Treatment 1 is included). The Treatment 1 and Treatment 2 estimates hold when a third treatment is added in Column 4, though Treatment 3 does not cause any statistically significant change in nighttime lights values for treated cells.

Figure 2 visually displays the treatment coefficients from our main models in Columns 2-4 of Table 2. The effect of Treatment 1 decreases when we add a second treatment (Models 3 and 4), Treatment 2 demonstrates the largest effect on nighttime lights, and Treatment 3 does not cause any statistically significant change in the outcome for treated cells (shown by the line representing the 95% confidence interval crossing 0).

**Figure 2: Graph of Treatment Coefficients for Main Models**



The identification of what might be driving the larger increase in nighttime lights that results from a second treatment (as compared to and in addition to the first treatment) is beyond the scope of this analysis, but certainly worthy of pursuit. It is possible that the later improvements made to roads were better in some way. A more likely explanation is that a second treatment indicates greater connectivity to multiple roads. Due to the networked nature of roads, it seems possible that the marginal effect of a second connection point could be greater than the effect of the initial connection and would lead to greater economic opportunity. Not only are two connections to a road network better than one, later improvements would benefit from a better and more established road network to connect to (as a result of earlier improvements).

Using the road network explanation of the importance of the second treatment above, we would expect to see an additional effect from Treatment 3 in Column 4 of Table 2. However, we do not. About 30% of our sample of cells fall within the buffers of 3 or more road segments and the standard error on the coefficient is still quite small, indicating that we do have enough statistical power to detect any treatment effect that exists for our sample. It is possible that these cells have reached a saturation point, in which any impact on economic activity occurred from the first or second treatment and the third treatment does not provide any added benefit (at least as measured by nighttime lights) for a cell of that size. Further examination of the effects of exposure to multiple improved road segments would provide useful insight for this evaluation and the siting and implementation decisions of future road projects.

**Heterogeneous Effects: Nighttime Lights at Baseline**

We consider whether results vary by the baseline levels of nighttime lights in each cell – in other words, do we see a differential effect for cells that were darker at baseline (in less populated areas) versus those that were brighter at baseline (in more populated areas)? Nighttime lights at baseline serve as the best available measure of population density. In most settings, there is a high correlation between nighttime lights and population density. While other sources of population data do exist for our study area, they are available much less frequently (yearly, at best), and at much coarser resolution (often only available by governorate). This exploration of heterogeneous effects also addresses whether the INP II improvements shifted economic activity from initially less active to more active areas.

We first categorize nighttime lights values at baseline by quartiles. The darkest 25% have nighttime lights values ranging from 0.42 to 1.27, while the brightest 25% have nighttime lights values ranging from 6.56 to 148.099 (see Table 1). These numeric values are a measure of brightness that is not particularly intuitive and are best understood in relation to each other. While there is a large gap between the smallest value, 0.42, and the largest, 148.099, the middle half of the data (from 25% to 75%) only spreads across 5 points, indicating that small changes in brightness are meaningful. The range in the top quartile (6.56 – 148.099) reveals the existence of a small number of very bright outlier cells in our sample.

Column 5 of Table 2 presents the impact of the earliest treatment (Treatment 1) on cells with differing levels of baseline luminosity. The impacts are statistically significant for all quartiles. In cells that were darkest at baseline, we observe a coefficient of -0.31, or an actual reduction in nighttime lights values for treated cells. For cells that were semi-dark, the impact on nighttime lights is still negative but smaller than the darkest cells, and for cells that were semi-bright, we see a positive effect of 0.38. For those cells that were brightest at baseline, we observe a large increase in nighttime lights (1.11) during the evaluation period. It should be noted that while the baseline values in this brightest quartile are of course grater than the median value, the magnitude of the change (1.11) is similar percentage-wise to that of Treatment 2 (about 17% for the cells with baseline luminosity of 6.56). We consider a 17% increase in brightness to be a substantive change.

The differential impacts of INP II are interesting – the cells that are brightest at baseline drive the main effect we see in Column 2 of Table 2. The cells that are darkest at baseline actually grow darker. Importantly, we do not see evidence that the main treatment effect merely reflects a shift in nighttime lights values from darker to brighter areas (suggesting a shift of the same level of economic activity from one area to another) but in fact indicates an overall increase in nighttime lights and economic activity within the study area as a whole.

**Heterogeneous Effects: 2km Buffers**

Table 3 presents the treatment effects for a sample of cells that limited to 2km buffers around the improved road segments (rather than 5km). The smaller buffers help to determine if the estimates of treatment impacts hold for the areas closest to the roads. This is of particular interest in this setting due to concerns that increases in settlement populations (located farther from the road segments) might drive the observed increase in nighttime lights, rather than the road improvements.

Our main results generally hold when we use this smaller sample of 82,814 cells that fall within 2km of improved road segments. While the statistical significance of the main treatment effect disappears (Columns 1 and 2 of Table 3), the inclusion of the second treatment in Column 3 identifies an increase of 0.9 in nighttime lights values for treated cells during the study period. This is a substantial effect. The effect of a third treatment in Column 4 is actually a decrease in nighttime lights values. This result is a bit puzzling, although it is worth noting that the magnitude of the decrease is smaller than the increase from the second treatment, which still indicates an overall gain in nighttime lights values for cells with three treatments.

Heterogeneous effects due to baseline values of nighttime lights also demonstrate a similar pattern to the differential effects found in the main model. In Column 5 of Table 2, we observe a decrease in luminosity for the darkest cells at baseline. Semi-bright cells experience virtually no change, while lights values in the brightest cells increase by about 1. As in our main model, we conclude that while the darkest cells get darker, indicating some shift of economic activity within the study area, we also find evidence of an overall increase in economic activity in areas close to the improved roads.

In both this model and the main model, nighttime lights at baseline serve as the best available measure of population density. The results of the interaction between treatment and nighttime lights indicate that treatment effects are positive and strong in the most populated areas – a pattern demonstrated when limiting the study to a narrow 2km buffer around roads and at a larger 5km distance as well.

**Robustness Checks**

We conduct several robustness checks to help validate the results of our main models. We first consider whether our results vary based on the size and placement of our unit of analysis. We draw the unit of analysis in our main models from the grid cells corresponding to the VIIRS data, the size and placement of which are determined by the way in which the satellite data is collected. We would not expect the grid cells to align with roads, political boundaries, or any other divisions in a way that would bias our results. As a check, we aggregate outcome and treatment values to a 1km square grid cell and re-run our models using this sample of cells. As we would expect, Table 4 demonstrates results at the 1km grid cell that are consistent with those of the smaller grid cells in our main model.

We also assess whether our results are driven by a trend that takes place over time regardless of our treatment. We conduct a placebo test in which we arbitrarily shift the timing of treatment earlier for all cells – e.g. if a cell was treated in September 2013, we construct a dataset that instead indicates treatment in March 2013. We would not expect to see a treatment effect in these models as no treatment actually takes place at that time.

We construct two different placebo tests. In the first, we construct a treatment date that is six months earlier than the actual treatment. We use the same VIIRS monthly nighttime lights outcome data. Table 5 displays the results. We find no statistically significant effect in Column 1, which includes a linear time trend. When we replace that with month fixed effects we actually do observe an effect, though one that is negative and barely statistically significant. This indicates that in the six months prior to the actual date of treatment, treated cells may actually experience a slight decrease in nighttime lights values. If anything, these findings strengthen the results demonstrated in our main model, as the observed increase in nighttime lights reversed the trend in those cells in the six months prior to treatment.

For the second placebo test, we construct a treatment date that is 5 years earlier than the actual treatment date. As the VIIRS nighttime lights satellite data only begins in April 2012, we utilize an alternate satellite outcome measure in these models. The Defense Meteorological Satellite Program (DMSP) provides yearly aggregates of nighttime lights between 1992 and 2013 at a 1km square grid cell unit. As the DMSP data is only available yearly, the timing of treatment is also assigned yearly – e.g. if a cell was actually treated in September 2013, the placebo treatment is 2008, and this would be true for any cell treated in any month of 2013. As we would expect, we do not find evidence of a treatment effect for this second placebo test.

## 4. Conclusions and Next Steps

In this evaluation, we do find evidence that the road improvements made through INP II increase nighttime lights, and thus economic development, for areas within 5km of the improved road segments. We find that the level of brightness at baseline differentially affects the change in nighttime lights, and these results hold when we vary the size of the buffer and the size of the grid cell. Darker areas at baseline get darker as a result of program treatment, while areas that are brighter at baseline get brighter. The increases to nighttime lights in these brighter cells more than offset the decreases in the darker cells, suggesting an overall increase in economic activity across the study area as a whole, rather than a mere shift of activity from one area to another.

Despite the overall gains, the differential effects by level of brightness at baseline are of interest and deserve additional exploration. It is important to note that nighttime lights measures the production of economic activity and these findings suggest that production has increased overall, but decreased in some areas and increased in others. However, we cannot conclude that *welfare* for those living in the darker areas at baseline has necessarily decreased in the darker areas, as our models do not account for the movement of people. It is certainly possible that the road improvements increased mobility and allowed those living in darker areas to obtain economic opportunities elsewhere that maintained or increased their welfare levels. The opposite is also possible, in which both economic production and welfare levels decreased in the areas that were darkest at baseline. The distinction is important, though the data and design of this evaluation are limited in shedding greater light on these heterogeneous impacts. Additional exploration through household surveys or other on-the-ground estimates of economic wellbeing can provide further understanding of the effects on beneficiaries to contribute to future planning and program decisions.

Overall, the positive impacts we measure on economic activity due to the road improvements made through INP II are encouraging. INP II is a component of a broader initiative improving access to infrastructure, and this evaluation indicates that gains from this investment are likely to be substantive and important. In an effort to better learn from this analysis, it is worthwhile to both further examine the treatment effects and qualitatively understand the larger context and implementation of the INP II program. While the latter is probably best done by Mission staff, there exist a number of ways to build on this analysis to accomplish the former. One clear next step is to re-run the analysis with the information on road improvements that were completed after December 2016. These improvements were not included in our analysis, but given the availability of nighttime lights satellite data beyond 2016 and the ability to reapply the same design and even code for the analysis, this is relatively low-hanging fruit that would serve as another robustness check to validate the results or identify greater nuance within the program impacts.

Building on the foundation of this evaluation, new analysis could include incorporating data on the larger road network to better understand how the INP II improvements connected our sample of cells with roads that connect to other areas of the West Bank (including considerations of number of connection points, travel time to major roads, etc). It would also be instructive to connect the satellite measures of economic activity with on-the-ground estimates of wellbeing through census or household survey data, in part to better understand the heterogeneous effects we observe due to baseline luminosity. Both of these options require additional data that may not be available, at all or for the needed time periods, or new data collection. Rigorous evaluations of ongoing or future infrastructure projects (beyond roads) also offer opportunities to contribute to learning and targeting of resources to improve the economic development in the region.

## References


Doll, C. N., Muller, J., & Morley, J. G. (2006). "Mapping regional economic activity from night-time light satellite imagery." *Ecological Economics* 57, (1), 75-92. doi:10.1016/j.ecolecon.2005.03.007.

Ebener, S., Murray, C., Tandon, A., & Elvidge, C. C. (2005). "From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery." *International Journal of Health Geographics,* 4(5). doi:10.1186/1476-072X-4-5

Ghosh, T., Anderson, S., Elvidge, C., & Sutton, P. (2013). "Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being." *Sustainability, 5*(12), 4988-5019. doi:10.3390/su5124988

Henderson, J. V., Storeygard, A., & Weil, D. (2012). "Measuring Economic Growth from Outer Space," *American Economic Review*, American Economic Association, 102(2), 994-1028. doi:10.1257/aer.102.2.994.

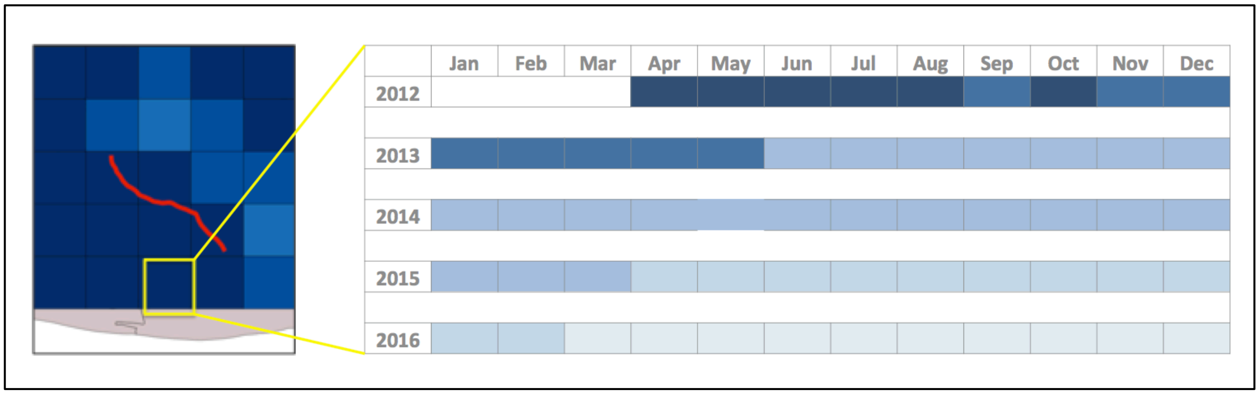
Michalopoulos, S., & Papaioannou, E. (2014). "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, Oxford University Press, vol. 129(1), 151-213.

Sutton, P. C., & Costanza, R. (2002). "Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation." *Ecological Economics,* 41(3), 509-527. doi:10.1016/s0921-8009(02)00097-6

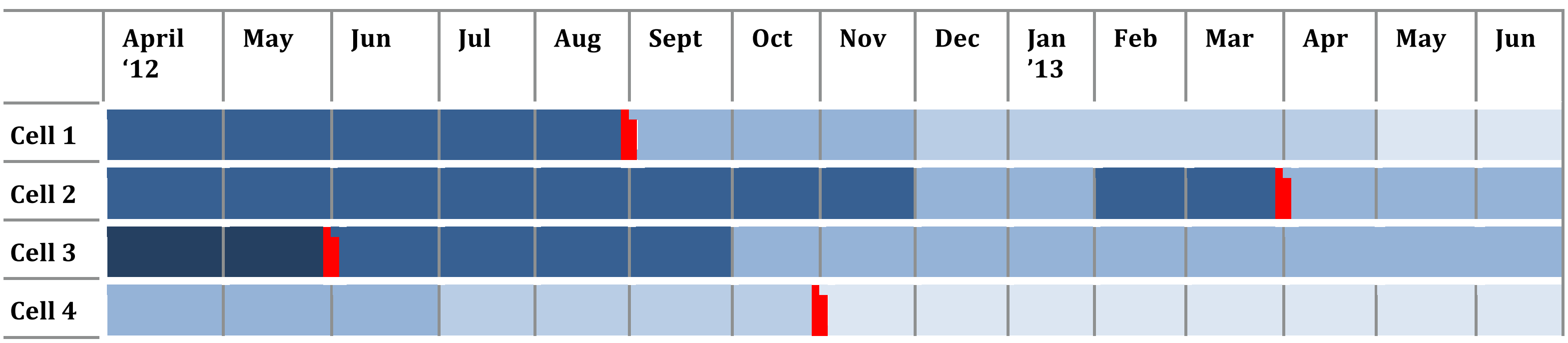
Sutton, P. C., Elvidge, C. V., & Ghosh, T. (2007). "Estimation of Gross Domestic Product at Sub-National Scales using Nighttime Satellite Imagery." *International Journal of Ecological Economics & Statistics,* 8(S07), 5-21.

**Figure 3: Construction of Panel Dataset (Using Illustrative Data)**

1. Collect monthly measure of nighttime lights data for each 750m square grid cell in our sample, beginning in April 2012 and ending in December 2016 (57 months).




2. Compile monthly nighttime light measurements for all cells and identify month and year of treatment (when improvements were made to associated road segments through INP II), marked by red line.




3. Align relative timing of treatment to examine patterns of change in nighttime lights across all cells and relative to each cell's timing of treatment (i.e. one month after treatment, five months after treatment, etc.).

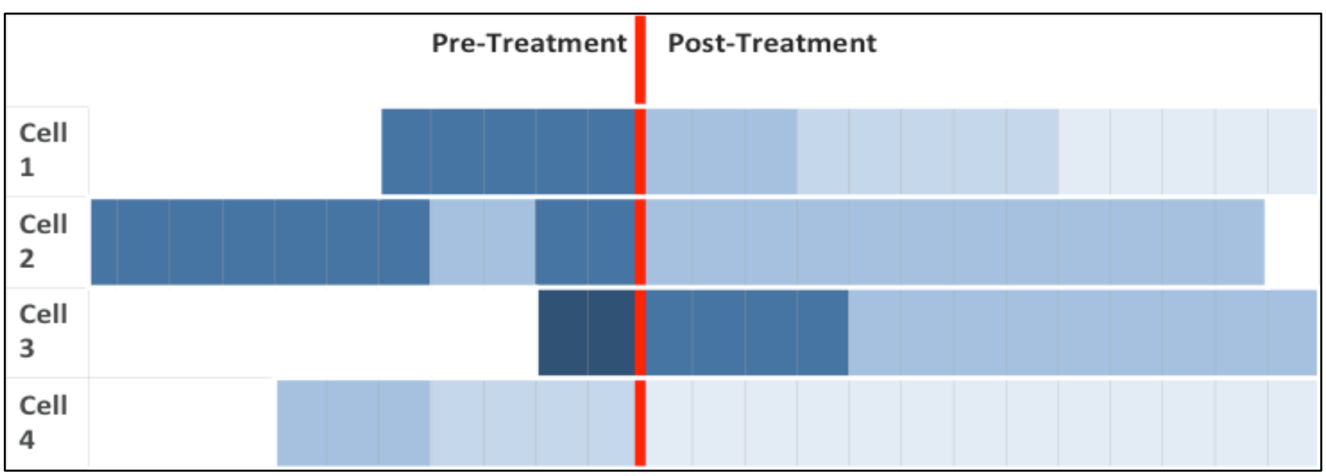


**Table 1: Summary Statistics**

| Statistic | N | Mean | St. Dev. | Min | Median | Max |
|---|---|---|---|---|---|---|
| NTL: Baseline | 6,729 | 6.529 | 10.581 | 0.422 | 2.632 | 148.099 |
| NTL: Endline | 6,729 | 7.418 | 11.199 | 0.503 | 3.358 | 155.636 |
| 1st Treatment Length | 4,598 | 31.6 | 9.527 | 6 | 34 | 46 |
| Distance to 1st Treat | 4,598 | 3,369.01 | 1,454.75 | 53.963 | 3,587.17 | 5,860.26 |
| Distance to 2nd Treat | 3,297 | 3,388.04 | 1,456.46 | 51.871 | 3,598.21 | 5,834.54 |
| Distance to 3rd Treat | 1,978 | 3,259.60 | 1,477.99 | 73.392 | 3,421.70 | 5,847.58 |

*Notes:* Table 1 displays descriptive statistics for the VIIRS cell-month panel dataset used to estimate the main model for 750 m square grid cells (see Table 2 for main model regression results). The first two rows summarize nighttime lights values at baseline (April 2012) and endline (December 2016). The nighttime lights values are a measure of radiance or brightness, which are not intuitive and best understood in relation to each other. First Treatment Length refers to the duration of a cell's earliest treatment (or the earliest road within 5 km to receive improvements). The distance measures reflect the average distance of a grid cell to the improved road segments (first being earliest).

**Table 2: VIIRS Cell-Month Panel Model Results (Main Model)**

| VARIABLES | (1) Model 1 | (2) Model 2 | (3) Model 3 | (4) Model 4 | (5) Model 5 |
|---|---|---|---|---|---|
| Treatment 1 | 0.321** | 0.341*** | 0.200** | 0.199** | |
| | (0.129) | (0.116) | (0.0926) | (0.0926) | |
| Treatment 2 | | | 0.571*** | 0.533*** | |
| | | | (0.172) | (0.167) | |
| Treatment 3 | | | | 0.152 | |
| | | | | (0.237) | |
| Treatment 1 x NTL Baseline (1st quartile; Darkest) | | | | | -0.326*** |
| | | | | | (0.0983) |
| Treatment 1 x NTL Baseline (2nd quartile) | | | | | 0.146*** |
| | | | | | (0.0397) |
| Treatment 1 x NTL Baseline (3rd quartile) | | | | | 0.414*** |
| | | | | | (0.0832) |
| Treatment 1 x NTL Baseline (4th quartile; Brightest) | | | | | 1.584*** |
| | | | | | (0.404) |
| Month | 0.00259*** | | | | |
| | (0.000411) | | | | |
| Observations | 383,525 | 383,525 | 383,525 | 383,525 | 383,525 |
| R-squared | 0.973 | 0.974 | 0.974 | 0.974 | 0.976 |
| Grid cell FEs | Y | Y | Y | Y | Y |
| Month FEs | N | Y | Y | Y | Y |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 3: VIIRS Cell-Month Panel Model Results, 2km Buffer**

| VARIABLES | (1)<br>Model 1 | (2)<br>Model 2 | (3)<br>Model 3 | (4)<br>Model 4 | (5)<br>Model 5 |
|---|---|---|---|---|---|
| Treatment 1 | 0.276 | 0.298 | 0.243 | 0.255 | |
| | (0.213) | (0.204) | (0.193) | (0.195) | |
| Treatment 2 | | | 0.900** | 0.931** | |
| | | | (0.445) | (0.432) | |
| Treatment 3 | | | | -0.763** | |
| | | | | (0.356) | |
| Treatment 1 x NTL Baseline | | | | | -0.481*** |
| ($1^{st}$ quartile; Darkest) | | | | | (0.143) |
| Treatment 1 x NTL Baseline | | | | | 0.195*** |
| ($2^{nd}$ quartile) | | | | | (0.0572) |
| Treatment 1 x NTL Baseline | | | | | 0.468*** |
| ($3^{rd}$ quartile) | | | | | (0.108) |
| Treatment 1 x NTL Baseline | | | | | 1.471*** |
| ($4^{th}$ quartile; Brightest) | | | | | (0.548) |
| Month | 0.00344*** | | | | |
| | (0.000534) | | | | |
| Observations | 82,814 | 82,814 | 82,814 | 82,814 | 82,814 |
| R-squared | 0.969 | 0.972 | 0.972 | 0.972 | 0.974 |
| Grid cell FEs | Y | Y | Y | Y | Y |
| Month FEs | N | Y | Y | Y | Y |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 4: VIIRS Cell-Month Panel Model Results, using 1km Grid Cells**

| VARIABLES | (1)<br>Model 1 | (2)<br>Model 2 | (3)<br>Model 3 | (4)<br>Model 4 | (5)<br>Model 5 |
|---|---|---|---|---|---|
| Treatment 1 | 0.317** | 0.336*** | 0.196** | 0.195** | |
| | (0.125) | (0.112) | (0.0882) | (0.0882) | |
| Treatment 2 | | | 0.547*** | 0.498*** | |
| | | | (0.168) | (0.161) | |
| Treatment 3 | | | | 0.194 | |
| | | | | (0.238) | |
| Treatment 1 x NTL Baseline | | | | | -0.286*** |
| (1st quartile; Darkest) | | | | | (0.0977) |
| Treatment 1 x NTL Baseline | | | | | 0.256*** |
| (2nd quartile) | | | | | (0.0563) |
| Treatment 1 x NTL Baseline | | | | | 0.691*** |
| (3rd quartile) | | | | | (0.128) |
| Treatment 1 x NTL Baseline | | | | | 1.326*** |
| (4th quartile; Brightest) | | | | | (0.401) |
| Month | 0.00250*** | | | | |
| | (0.000409) | | | | |
| Observations | 218,196 | 218,196 | 218,196 | 218,196 | 218,196 |
| R-squared | 0.975 | 0.976 | 0.976 | 0.976 | 0.977 |
| Grid cell FEs | Y | Y | Y | Y | Y |
| Month FEs | N | Y | Y | Y | Y |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 5: VIIRS Placebo Treatment Panel Model Results (6 months)**

| VARIABLES | (1)<br>Model 1 | (2)<br>Model 2 |
|---|---|---|
| Treatment 1 Placebo | -0.00892 | -0.180* |
| | (0.102) | (0.0967) |
| Month | 0.00319*** | |
| | (0.000457) | |
| Observations | 383,525 | 383,525 |
| R-squared | 0.972 | 0.974 |
| Grid cell FEs | Y | Y |
| Month FEs | N | Y |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 6: DMSP Placebo Treatment Panel Model Results (5 years)**

| VARIABLES | (1)<br>Model 1 | (2)<br>Model 2 |
|---|---|---|
| Treatment 1 Placebo | 0.250 | -0.737 |
| | (2.204) | (0.834) |
| Year | -0.506 | |
| | (0.696) | |
| Observations | 33,645 | 33,645 |
| R-squared | 0.893 | 0.923 |
| Grid cell FEs | Y | Y |
| Month FEs | N | Y |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1