Stabilization in Afghanistan: Trends in Violence, Attitudes, Well-being and Program Activity

Radha Iyengar, Jacob N. Shapiro, Benjamin Crisman, Manu Singh, and James Mao

RAND Labor & Population

WR-1192

June 2017

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the RAND Labor and Population Unit.

For more information on this publication, visit www.rand.org/pubs/working_papers/WR1192.html

Published by the RAND Corporation, Santa Monica, Calif.

Stabilization in Afghanistan: Trends in Violence, Attitudes, Well-being and Program Activity

Radha Iyengar,[1] Jacob N. Shapiro,[2] Benjamin Crisman,[3] Manu Singh,[4] and James Mao[5]

April 11, 2017

This research was conducted by the Empirical Studies of Conflict Project (ESOC) at Princeton University. The authors and ESOC are grateful for the support provided by the United States Agency for International Development (USAID) and the United States Institute of Peace (USIP). USIP convened and chaired several Advisory Board meetings on this project, and the inputs and guidance provided by independent Advisory Board members is gratefully acknowledged. USAID helped organize interviews and provided unprecedented access to internal data. Funding was provided through an interagency research agreement from the Office of Afghanistan and Pakistan Affairs at USAID and USIP. All errors are our own.

[1] Senior Economist, RAND Corporation; Visiting Research Scholar, Princeton University

[2] Professor of Politics and International Affairs, Princeton University. Co-Director, Empirical Studies of Conflict Project (ESOC)

[3] Research Specialist, Princeton University

[4] Research Specialist, Princeton University

[5] Research Specialist, Empirical Studies of Conflict Project (ESOC) and Graduate Student at Princeton University

Table of Contents

Acronyms

ACSOR	Afghan Center for Socio-Economic and Opinion Research
AGE	Anti-Government Elements Afghan Info
AI	
AIMS	Aid Information Management Systems
AISCS	Afghanistan Infrastructure and Security Cartography System
ALLI	Alternative Licit Livelihoods Initiatives
ANDP	Afghanistan National Development Program
ANQAR	Afghanistan Nationwide Quarterly Assessment Research
ANSO	Afghanistan NGO Safety Office
ANVIL	Name of survey (not an acronym)
ASI	Afghanistan Stability Initiative
BINNA	Name of survey (not an acronym)
CBSG	Community Based Stabilization Grants
CCI	Community Cohesion Initiative
CDP	Community Development Program
CERP	Commander's Emergency Response Program
CIDA	Canadian International Development Agency
COIN	Counterinsurgency
COM	Community Cohesion

CSO	Central Statistics Office
DFID	Department for International Development
DHS	Demographic and Health Surveys
DOD	Department of Defense
DTEM	Digital Terrain Elevation Map
ECO	Economic Well-being
ESOC	Empirical Studies of Conflict Project
FOB/COP	Forward Operating Base/Combat Outpost
FOGHORN	Name of survey (not an acronym)
GIRoA	Government of the Islamic Republic of Afghanistan
GIS	Geographic Information Systems
GOV	Support for Government
HEA	Health

ISAF	International Security Assistance Force
ISVG	Institute for the Study of Violent Groups
LGCD	Local Governance and Community Development Project
MICS	Multiple Indicator Cluster Survey
MISTI	Measuring the Impact of Stabilization Initiatives Project
MRRD	Ministry for Rural Rehabilitation and Development
NATO	North Atlantic Treaty Organization
NOAA	National Oceanic and Atmospheric Administration
NRVA	National Risk and Vulnerability Assessment
NSP	Afghanistan National Solidarity Program
NTMA	NATO Training Mission in Afghanistan
OAPA	USAID Office of Afghanistan and Pakistan Affairs
OTI	USAID Office of Transition Initiatives
PAP	Pre-Analysis Plan
PCA	Principal Component Analysis
SIGACTS	Significant Activities (e.g. violent events)
SIKA	Stability in Key Areas Project
STAY	Skills Training for Afghan Youth
UNDSS	United Nations Department for Safety and Security
USAID	United States Agency for International Development
USG	United States Government
USIP	United States Institute for Peace
VIO	Violence

Executive Summary

After billions of dollars and decades of investment in Afghanistan, Iraq and elsewhere, practitioners, policymakers and academics are still uncertain about which programs best foster stability in fragile and conflict-affected areas. This lack of clarity leaves decision makers without the information they need in order to select and support successful counterinsurgency, counterterrorism, and economic development goals. The political, social, and economic consequences of failing to learn from previous efforts are serious. Moreover, ongoing conflicts in Syria, West Africa and other parts of the world beg the question: what worked?

A large number of external performance and impact evaluations of stabilization projects implemented by the US government (USG), US military, other donor governments, and international organizations in Afghanistan have been conducted during the past decade. As have multiple external analyses and reports on various aspects of stabilization and local-governance programming, including the work of other international donors. These include several carefully designed, quantitative evaluations of several programs: a small subset of USAID stabilization programs were evaluated in the Measuring Impacts of Stabilization Initiatives (MISTI) Project, the World Bank's National Solidarity Program (NSP) was evaluated using a randomized controlled trial, and the US military's Commander's Emergency Response Program (CERP) was subjected to a detailed independent evaluation by RAND and an academic paper published in a leading political science journal. In addition to these evaluations, there was detailed and ongoing collection of attitudinal data on stabilization and governance programs as well as on and perceptions of security and governance in Afghanistan, including, but not limited to: the Asia Foundation's Survey of the Afghan People, ACSOR and D3's Afghan Futures Survey, and Afghanistan Nationwide Quarterly Assessment Research (ANQAR).

To better understand the relationship between stabilization programming and trends in key outcomes – including: security, popular support for the government, popular support for anti-government elements (AGE), community cohesion and resilience, health of the Afghan people, economic well-being of the Afghan people, and conflict – the research team compiled and analyzed data from multiple sources. These sources include the MISTI evaluation data, the World Bank's NSP evaluation data and the U.S. military's CERP program activity, as well as administrative data, such as spending and location, on USAID's Office of Transition Initiatives (OTI) programs in Afghanistan. We also collected data on the types and location of USAID projects in Afghanistan from 2010 to 2015, data from 16 quarterly surveys in Afghanistan spanning 2008 - 2016, and satellite imagery data to measure population density and economic activity.

The analysis was centered on six research questions that were identified before conducting the

research review.[6] These questions were:

> **R1**: What did stabilization projects achieve in terms of key outcomes: security; popular support for the government; popular support for anti-government elements (AGE); community cohesion and resilience; health of the Afghan people; economic well-being of the Afghan people; and conflict events?
>
> **R2**: Over what time horizon is these effects apparent and how quickly do any gains or losses fade?
>
> **R3**. How does the presence of the military impact the outcomes of stabilization projects?
>
> **R4**. What types of synergies and confounding factors exist between stabilization programs by different actors (other parts of the USG, other countries, Afghan government, international organizations like the World Bank, etc.)?
>
> **R5**. Are impacts of stabilization programs amplified or reduced when considering specific aspects (size, contract type, etc.) or sectors (agriculture, infrastructure, skills, etc.) of projects?
>
> **R6**. What commonalities exist when looking across a number of successful or unsuccessful stabilization projects between different actors and different sectors?

This report summarizes the quantitative trends generally as well as specific trends related to these questions. However, because of data limitations, we were unable to develop robust analysis related to questions R5 and R6 and thus we do not present findings related to those research questions. To better contextualize the findings to questions R1-R4, we also present information from two other sources. First, we conducted a research review to identify common themes and broad lessons from research spanning academic, government, and policy literatures (See Iyengar, Shapiro, and Heagarty, 2017 for additional details). Second, we interviewed 13 program officers, contract officers, and program staff. The interviews were very useful for providing an "on the ground" perspective and highlighting some key themes evident in the data. The participants provide a complementary perspective on issues identified in the research review and quantitative analysis.

Table E.1 summarizes our findings on the relationship between six outcomes: levels of violence (VIO), support for the Afghan government (GOV), support for anti-government elements (AGE), community cohesion (COM), public health (HEA), and economic activity (ECO). For these outcomes, we look at correlations with stabilization program location and spending from five

[6] The questions listed below were adapted slightly after the initial research design due to data limitations and feasibility constraints.

sources: Afghan Information (AI) (a database maintained by USAID for management purposes), OTI records (OTI), Local Governance and Community Development program records (LGCD), project-level data on the Commander's Emergency Response Program spending (CERP), and World Bank data on the NSP program (NSP).

Significant correlations between programs and desired outcomes are shown in green. So, reductions in violence, increases in support for Afghan government, decreases in support for AGE, increases in community cohesion, and increases in health and economic well-being are all illustrated in green. Significant correlations between programs and negative outcomes are shown in red. This means that increases in violence, decreases in support for Afghan government, increases in support for AGE, decreases in community cohesion, and decreases in health and economic well-being are all shown in red.

To summarize how large the estimated effects are, we show effects that are 0.1 or 0.2 standard deviations in size—what we would identify as substantial in magnitude—in bold. Overall, stabilization programming was associated with some improvements in perceived access to healthcare and economic activity, despite more resources being spent in more violent areas. This varies substantially by program type, however, with LGCD having a robust relationship to perceptions of improved health access and CERP having a small but consistent relationship with economic activities. However a number of programs had the opposite relationship with CERP being negatively associated with community cohesion, for example. However, there is almost no evidence of effectively reducing violence. The only program that is negatively correlated with violence is NSP and this is likely due to site selection in relatively low-violence areas rather than impact on violence direct. In all other cases, programs are positively correlated with violence suggesting (as was confirmed by the program documents and the interviews) that these programs specifically chose to work in some of the most violent and unstable places in Afghanistan. There is limited evidence that any substantial gains persisted

E.1 Summary of Key Findings

Research Question	Other Variables for Analysis	Estimated Program Outcome Relationship
R1. Short-term relationship between stabilization programs and key outcomes	N/A	VIO: AI,***OTI, LGCD,*** NSP, ***CERP*** GOV: ***AI,*** LGCD, NSP AGE : ***AI***, LGCD, OTI COM : AI, ***CERP*** HEA: AI, ***LGCD,*** CERP ECO: AI, OTI, CERP
R2. Longer time-horizon measures of well-being	N/A	VIO: AI, LGCD, NSP GOV:OTI, LGCD AGE: NSP, AI,OTI, LGCD COM:AI HEA:CERP
R3. Relationship between military presence and outcomes	Total forces present during the time period	VIO :AI, ***OTI, NSP, CERP*** GOV :AI AGE, OTI HEA: ***LGCD, AI,*** CERP
R4. Synergies between USAID and other donor programs	Total other donor spending, Total other donor project count	VIO :AI, ***NSP*** AGE:AI, OTI HEA:***LGCD*** ECO: OTI

Our analysis resulted in several conclusions, which should be informative for those designing and implementing similar programs in the future. These include

- *Programs are likely to have small, but meaningful gains):* While none of the gains are large, and while we cannot establish causal relationships with the same credibility of impact evaluations conducted outside of conflict zones, increased spending on stabilization programming is associated with gains in perceived access to health services and support for the government. These statistically significant improvements are especially notable given that stabilization programming targeted more insecure areas, and suggest that the effort resulted in a small overall improvement in outcomes compared to a scenario where no stabilization programs were run.[7] These gains show that stabilization aid can be effective and must be more rigorously evaluated to enable more effective program designs in the future.

[7] We considered "small" improvements to be 0.1 standard deviations and larger gains include 0.2 standard deviation improvements or greater.

- *Detailed, accessible implementation data can allow real-time adaptation*: Although USAID collects data in a variety of ways during program design and implementation, often the implementation data are not being recorded with sufficient detail or accessible to the full-range of people involved in implementing the program to enable retrospective learning and adaptive management. Pulling the information together for this project required substantial coordination between USAID, USIP, the World Bank, and an academic institution. Despite that effort, data on many programs simply could not be feasibly shared because of the difficulty in combing through a wide range of information collected. The requirement for better, standardized, centralized record keeping needs to be built in to future contracts from day one if the USG hopes to learn from and improve upon its stabilization efforts. These data should be as comprehensive as possible; to include what was done where and when, but also the rationale for decision-making (such as why certain sites were chosen over others).
- Third, structuring evaluations appropriately in conflict zones requires long-term thinking and coordination. The MISTI evaluation was an unprecedented effort to measure stabilization impacts in one place, but it was not set up to learn about key design elements for stabilization programming and was not part of a family of similar efforts that could have probed whether what worked (or did not work) in southern Afghanistan worked (or did not work) elsewhere. In the future, processes to evaluate program design should be an explicitly identified goal within the broader scope of planned evaluations.
- Fourth, there is tremendous potential in using remote sensing data to track outcomes. Modern open-source tools for working with geo-spatial data can be applied to remote sensing data available from the National Oceanic and Atmospheric Administration (NOAA), as well as commercial providers to measure economic conditions and population welfare in even the toughest areas. Doing so requires sensitivity to the quirks of data collection and to cross-regional differences in the relationship between on-the-ground conditions and what can be seen from space, but the potential exists to measure changes at fine geo-temporal scales in any location on earth. That opens up tremendous opportunities for learning and policy feedback provided that detailed programmatic data is maintained. This technology should be leveraged to provide the most complete information possible to integrate into future evaluations.

1. Introduction

After billions of dollars and decades of investment in Afghanistan, Iraq and elsewhere, practitioners, policymakers and academics are still uncertain about which programs best foster stability in fragile and conflict-affected areas. This lack of clarity leaves decision makers without the information they need in order to select and support successful counterinsurgency, counterterrorism, and economic development goals. The political, social, and economic consequences of failing to learn from previous efforts are serious. Moreover, ongoing conflicts in Syria, West Africa and other parts of the world beg the question: what worked?

Afghanistan is among the most compelling case studies from which we can potentially identify effective programs and best practices for stabilization efforts in conflict-affected areas. With billions of dollars spent, massive military commitments by the US and a host of partners, and more than 15 years of ongoing engagement, the programs and projects that have been implemented in the country represent the full spectrum of stabilization activities. Afghanistan also provides an opportunity to examine one of the most concerted and well-resourced efforts to use "hearts and minds" projects to achieve greater security and improved social and economic well-being. Not surprisingly, given the relevance for current and future policy and the allocation of resources, there have been a range of empirical studies to better understand the impact of different programs on political, social, economic, and security outcomes.

A large number of external performance and impact evaluations of stabilization projects implemented by the US government (USG), US military, other donor governments, and international organizations in Afghanistan have been conducted during the past decade, as have multiple external analyses and reports on various aspects of stabilization and local-governance programming, including the work of other donors. These include carefully designed, quantitative evaluations of several programs: a small subset of USAID stabilization programs were evaluated in the Measuring Impacts of Stabilization Initiatives (MISTI) Project, the World Bank's National Solidarity Program (NSP) was evaluated using a randomized controlled trial, and the US military's Commander's Emergency Response Program (CERP) was subjected to a detailed independent evaluation by RAND and an academic paper published in the top political science journal. In addition to these evaluations, there was detailed and ongoing collection of attitudinal data on stabilization and governance programs as well as on perceptions of security and governance in Afghanistan, including, but not limited to: the Asia Foundation's Survey of the Afghan People, ACSOR and D3's Afghan Futures Survey, and Afghanistan Nationwide Quarterly Assessment Research (ANQAR).

These multiple evaluations each provide insight into the specific programs under study but raise questions as to the broader applicability of their results. Moreover, these studies have not been

placed in the context of the broader trends in Afghanistan in large part because the multiple sources of data have not been fully de-conflicted and connected, limiting the feasibility of assessing and analyzing across data types, programs, and time periods.

This study aims to link the wealth of knowledge in existing quantitative resources to better understand the impact of stabilization programming in Afghanistan. To do this, we compile several sources of data on different programs including detailed data from the MISTI, NSP, and CERP evaluations; detailed information on Office of Transition Initiatives (OTI) programs in Afghanistan; information on the types and location of all USAID projects in Afghanistan from 2010 to 2015; a range of information from different quarterly surveys in Afghanistan; and newly analyzed satellite imagery. Combining these data sources and analysis, we aim to better understand the extent to which stabilization projects were associated with changes in key outcomes including security, popular support for the government, popular support for anti-government elements, community cohesion and resilience, access to healthcare for the Afghan people, economic well-being of the Afghan people, and conflict events.

We cannot estimate the causal relationships between stabilization programming and any of the outcomes of interest. This is because, although many of our outcomes are likely affected by stabilization programs, it is also quite likely that the reverse is true—that some programs are affected by the outcomes. For example, some locations may have experienced stabilization independent of the influence of stabilization programming, but donors and implementers would be more likely to replicate projects active in that location in other areas due to their perceived correlation with success. We cannot address this reverse causation or the potential for omitted variable bias (when failing to account for or measure the influence of one variable leads to misattribution of causation to other variables) in our analysis; thus, we cannot establish what would have happened in the counterfactual scenario (what would have happened in the absence of the programs), which is necessary to estimate causal impact (the difference between outcomes in the presence and absence of programs).

Due the difficulty in measuring the impact of specific programs and projects, the research team decided to focus instead on understanding the broader context in which the assessments were conducted. Conditional correlations between stabilization programming and outcomes are quite informative because they extract a specific relationship between variables, while holding a wide range of other factors fixed. It is also helpful to examine the correlation between outcomes and both *levels* of and *changes* in stabilization activity. For instance, while we may see that high levels of violence and high levels of project activity are correlated because stabilization efforts are focused in "tougher" locations, we may also see that changes in violence and changes in project activity are negatively correlated. This would suggest that while more projects exist in violent areas, changes in project activity are associated with reductions in violence (such a sign

flip is observed for small scale aid projects in Iraq but not larger scale ones, suggesting that small-scale aid spending was more effective in reducing violence in that setting).

These approaches allow us to consider the broad, systematic relationship between the outcomes, project activity, and a host of relevant covariates (e.g., ethnicity, agricultural activity, amount of infrastructure). Our focus is on whether trends in outcomes are systematically associated with the effects of similar programs across time periods or locations but not on determining the direct causal effect of a specific program. Moreover, our data allows us to explore the types of synergies and confounding factors that exist among stabilization programs by different actors. We supplement this quantitative analysis with detailed information from 13 structured interviews with individuals involved in the design, implementation, and oversight of stabilization programs in Afghanistan to understand the commonalities that existed across a number of stabilization projects between different actors and different sectors.

2. Approach

The research team identified, collected, processed and then analyzed quantitative data sets to produce the standardized panel data set used for subsequent analysis. The panel data set includes information on outcomes of interest, project activity and a host of geographic, sociodemographic and temporal variables for each of the 398 districts in Afghanistan every quarter from 2009 until 2016. [8] In the rest of this section, we introduce the sources from which we collected our data and the measures created from those sources, and include a brief summary of the approaches employed to construct and validate these measures.

2.1 Data Sources

The data used to produce the standardized panel data set comprise a variety of information types, including survey research, geospatial data, administrative records, and demographic surveys. The datasets described in Table 2.1 span a variety of types of data. Many of the sources of information come from administrative datasets that contained information on program spending, location, dates of activities, and sometimes additional details on the program intent or beneficiaries. Many of these datasets were collected for use by USAID or the implementing partner, rather than for systematic analysis for programmatic activity and thus had only a limited number of variables relevant for our analysis. In addition, several of the datasets are constructed from detailed surveys conducted in Afghanistan over the period of evaluation. When feasible, we cross-validate survey questions across separate related questions, time intervals, and with other measures (e.g., geospatial data). The data sets used include:

ANQAR – The Afghanistan Nationwide Quarterly Research (ANQAR) survey data aims to gain a broad understanding of the attitudes, behaviors, and issues that are important to the people of Afghanistan. The fieldwork was conducted by ACSOR for NATO/ISAF and the survey was conducted through face-to-face interviews. The sample was drawn using a multi-stage random stratification process using settlements as the primary sampling unit and stratifying by province and urban/rural status using population data released by the Central Statistics Office. The survey is nationally representative; respondents were interviewed in all 34 provinces. The survey respondents are 18 years of age or older and included both males and females. Broadly, the questionnaire consisted of management and quality control questions, demographic questions, and substantive questions on topics including security, government services, reconciliation, and elections. The surveys spanned from 2008 until 2015, but a substantial change in survey design occurred starting in Wave 11 (year 2011) with additional modifications made through Wave 15 (2012). While many questions remained the same or similar, changes in sampling and survey protocol may have affected responses. We correct for this potential confounding factor by

[8] We note that the district and province definitions in Afghanistan changed over time and were recorded in an unsystematic way in many data sources. For this study, we use the 398 districts defined by the Afghan Central Statistics Office, and, when possible, map the geolocations directly into that file to ensure we are comparing the same areas over time. However, in some data sources GPS coordinates were not available; in such cases, we relied on the recorded information on village or district.

including survey-wave fixed effects, which controls for unique factors in each survey waves. This estimation approach allows us to compare key attitudinal outcome measures, as well as social and demographic descriptive variables, over time.

Table 2.1: Data Sources for Analysis

Data Source	Type (Key variables)	Date(s)
Afghan Info	Administrative (project level data from USAID)	2011-2015
ANQAR	Quarterly Surveys (Support for Gov/AGE & Community Cohesion)	2008-2014
CERP	Administrative Project Data (project budget)	2009-2011
FOB/COP Locations	Geospatial (military presence) ORBAT- Order of Battle	2009-2016
MISTI	Survey (Support for Gov/AGE & Community Cohesion)	2011-2014
National Risk and Vulnerability Assessment Microdata	Survey (Health and Economic outcomes)	2003, 2005, 2007, 2012
NSP	Administrative (project level data)	2009
OTI Data	Administrative (project level data) for CCI programs	2012-2015
SIGACTS	Incident data (violence)	2002-2016

LGCD	Administrative (project level data)	2007-2009
NOAA Nightlights	Geospatial (Economic Activity)	Annual
WorldPop/Landscan	Population	Fixed

Afghan Info: Afghan Info was used to identify programmatic information on stabilization projects in Afghanistan between 2010-2015. Based on detailed interviews with USAID experts on this database, there are several key aspects of Afghan Info that limited its utility for research and analysis. First, it is not designed or intended to serve as a financial system of record and thus could not be used to account for all USAID's spending in Afghanistan at the sub-national level. Second, Afghan Info only began tracking USAID expenditures since 2011 and does not track all USAID accounts including humanitarian accounts or operational expenditures. It excludes Office of Transition Initiatives (OTI) programs, which we obtained separately. Third, while Afghan Info does track most Economic Support Funds (ESF), it does not record ESF Transfers to other agencies, including USDA, PAS, Treasury, and even other parts of USAID like OTI. The ESF program promotes the economic and political foreign policy interests of the United States by providing assistance to allies and countries in transition, frequently in a multi-donor context. Despite these limitations, Afghan Info represented a new, rich set of data on USAID activity and is included for comparison in subsequent analysis.[9]

CERP – The Department of Defense (DOD) created the Commander's Emergency Response Program (CERP) in fiscal year (FY) 2004 to help military commanders respond to urgent humanitarian relief and reconstruction requirements in Afghanistan. Since its inception, CERP funds have been used to implement projects in all 34 provinces with a significant portion of these funds used in the South and South West regional command areas. Projects included, but were not limited to, transportation, education, agriculture/irrigation, healthcare, water and sanitation, and economic, financial and management system improvements. Most CERP projects were relatively low cost and limited in time-duration, therefore this data was useful in studying small-scale projects that, on average, are estimated to cost less than $500,000 each. However, the project management systems tracking CERP projects for DOD do not contain comprehensive information about the costs of all projects undertaken, limiting the CERP data's usability for our broader purposes.

[9] We also considered using AidData, a portion of Aid Information Management System (AIMS). This data source contains 1,580 projects in Afghanistan from 93 donors, spanning 2001 – 2014. However, the disbursement data in this dataset is not reported by quarter, nor even by year – a project's total disbursements for its entire duration are reported and thus were not useful for subnational, quarterly analysis.

Forward Operating Bases (FOB)/Command Outpost (COP) Locations – The Order of Battle (ORBAT) dataset describes the location and area of responsibility of all international military forces units in Afghanistan. This dataset enables us to determine the length and intensity of military presence, and by district. The data also provides brigade and battalion information, and number of casualties sustained by the unit.

MISTI – The Measuring Impact of Stabilization Initiatives Project (MISTI) is one of the largest trend analysis and impact evaluations of stabilization interventions conducted by USAID. The project was created to determine whether the USAID projects studied caused a change in stability at the local level. Baseline data was collected from September to December, 2012. Four successor surveys were then completed biannually (the last wave ended in December 2014). The data allowed MISTI to evaluate projects by quantifying changes in the stability between survey waves in intervention villages compared to non-intervention equivalent villages. The key indices studied were government capacity, local governance, quality of life and community cohesion.

NRVA – The National Risk and Vulnerability Assessment Microdata (NRVA) survey data provides information required for monitoring development progress and formulating development policies. The survey is conducted by the Central Statistics Office (CSO) of Afghanistan and provides results that are representative at national and provincial levels. The survey covers a wide range of developmental themes such as demography, poverty, food security, labor forces, agriculture and livestock, education, health, household amenities, and challenges. The survey instrument consisted of paper questionnaires for households, male and female communities (shuras), and commodity prices in the nearest market places.

NSP – The National Solidarity Program (NSP) Administrative Database was created by the Government of Afghanistan to plan, manage and monitor its own development projects. The program primarily focuses on promoting rural development. The data set provides the location of projects, start and end dates, project type, project budget, and spending information.

OTI –This dataset includes information on projects implemented by the USAID Office of Transition Initiatives' (OTI) Community Cohesion Initiative (CCI) from mid-2012 until early 2015; it includes program location and spending data from a total of 1,781 unique projects across villages in 106 districts in 20 Afghan provinces. This program intended to improve the relationship between local populations and local and national governments as well as increase cohesion within and among communities.

SIGACTs – Our primary source of data on violence is a unified version of Significant Actions (SIGACTs).[10] This dataset includes approximately 500,000 geo-referenced incidents. For each

[10] This dataset was compiled by Andrew Shaver and Austin Wright of ESOC – publication introducing the data forthcoming. Their paper provides a detailed discussion of how data was unified across different reporting periods to

action, we have precise latitude and longitude, time and date, instigator and target, and a brief description. To cross-validate the SIGACTs measures we are also working on processing alternative violent events data from the Institute for the Study of Violent Groups (ISVG) and additional measures from the United Nations Department for Safety and Security (UNDSS). The violence measures that we construct from these data are used throughout our analyses both as a primary outcome variable and as a potential confounder.

LGCD –USAID's Local Governance and Community Development (LGCD) Program dataset includes program location and spending information on 3,038 unique projects across villages in 189 districts in 27 provinces of Afghanistan.[11] The nominal goals of this program were to "1) assist the Government of the Islamic Republic of Afghanistan (GIRoA) to extend its reach into unstable areas and engage at-risk populations; 2) create an environment that encourages local communities to take an active role in their own stability and development; and 3) address the underlying causes of instability and support for the insurgency."[12]

NOAA Nightlights – This was the primary data source for understanding local economic activities and comes from satellite imagery of nighttime lights provided by National Geophysical Data Center at the National Oceanic and Atmospheric Administration (NOAA). Initially, this data source was mainly used for identifying bright lights from the cities and gas flares. But more recent studies have shown that nighttime lights strongly correlate with overall economic activity and other welfare proxies at local and national levels.[13]

There is a significant advantage to using satellite imagery in areas such as Afghanistan, where ground-based micro data is quite difficult to collect. Moreover, using sensor data, we can develop a single, stable indicator to provide fine-grained local data that could be viewed at very small geographic units (e.g., town or village). We aggregate this precise data to district,

---

ensure consistency in the measure over time.

[11] Although the data includes 27 provinces, the report summarizing this information lists only 21 provinces. https://www.usaid.gov/node/51846

[12] USAID. "Fact Sheet: Local Governance and Community Development (LGCD). June 2011. https://www.usaid.gov/sites/default/files/documents/1871/Fact%20sheet%20LGCD%20FINAL%20June%202011.pdf. Accessed 22 February 2017.

[13] 1. Filho, C. D. S., Zullo Jr, J., & Elvidge, C. (2004). Brazil's 2001 energy crisis monitored from space. *International Journal of Remote Sensing*, *25*(12), 2475-2482.

2. Chand, T. K., Badarinath, K. V. S., Elvidge, C. D., & Tuttle, B. T. (2009). Spatial characterization of electrical power consumption patterns over India using temporal DMSP-OLS night-time satellite data. *International Journal of Remote Sensing*, *30*(3), 647-661.Ghosh et al. 2010

3. Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *The American Economic Review*, *102*(2), 994-1028.

4. Christopher N. H. Doll, Jan-Peter Muller, & Elvidge, C. (2000). Night-Time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions. Ambio, 29(3), 157-162.

5. Elvidge, Christopher D., et al. "Mapping city lights with nighttime data from the DMSP Operational Linescan System." Photogrammetric Engineering and Remote Sensing 63.6 (1997): 727-734.

provincial and national levels, as required by the analysis. The data is also cross-validated with reported economic indicators from surveys such as ANQAR, NRVA and DHS and is found to correlate at above 60% in all three cases (see appendix).

WorldPop/Landscan – LandScan data assembled with the help of geographic information system and remote sensing provides a good quality dataset for global population distribution. This data provides fine-grained (available at a resolution of one square kilometer) ambient population levels (the average population of that square kilometer over a 24-hour period). For the purposes of this report, the population data is aggregated at district and provincial levels.[14]

2.2 Variable Construction for Outcome Measures

Using existing research, including the outcomes in the MISTI and NSP evaluations, we identified several key outcomes of interest, which include: intensity of conflict, popular support for the Afghan government (both national and local entities), popular support for anti-government entities (AGE), community cohesion, health of the Afghan people, and economic well-being of the Afghan people. Table 2.2 shows the outcomes of interest, metrics, data sources, and specific variable for each of the outcomes of interest.[15]

Table 2.2. Outcomes of Interest, Data Sources, Metrics, and Time Horizon for Analysis

	Outcome of Interest	Data Source	Variable
Violence	Intensity of conflict/security	SIGACTS	Number of attacks per capita
Near-term attitudes	Popular support for the Afghan government	ANQAR	Index of survey questions on stated support for national and local government
	Popular support for anti-government elements (AGE)	ANQAR	Index of survey questions on support for general and specific AGEs
	Community cohesion and resilience	ANQAR	Index of survey questions on willingness to rely on community members for support
Longer term well-being measures	Health Access for Afghan people	ANQAR	Perception of health care access
	Economic well-being of the Afghan people	Nightlights + NRVA	Night lights as a measure of market activity household consumption, index of poverty measures

[14] We use the base year estimates of 2009 to do

[15] These outcomes and measures were specified in the PAP with the exception of the health measure. Because we were unable to acquire sufficient cross-sectional data over relevant time periods for DHS or NRVA, we use a measure taken from the ANQAR surveys which asks respondents to rate the quality of healthcare available in their area. This metric was chosen because of the length of availability in the ANQAR and cross-validated by comparing it to relevant measures from DHS and has the advantage of being available at higher frequencies.

In the rest of this section, we detail the steps and processing actions taken to build the panel data for analysis, and discuss the potential biases or data quality issues for each source of data.

2.2.1 Violence

As previously stated, SIGACTs is our primary source of data on violence.[16] In the SIGACT data set, each attack is coded as one observation. After aggregating the data into usable units of analysis (cumulative weekly/monthly/quarterly incidents), we generated analogous measures for types of violent acts, including 1) combat acts, 2) criminal acts, and 3) counter-insurgency operations. Once aggregated at the district level, these measures were divided by the population of each district to get a measure of violence per-capita.

There are a number of issues related to the collection of violent events data in Afghanistan, the first of which is that the nature of collection is related to the focus of the collecting agency. In this case, ISAF likely emphasizes violent events related to insurgency rather than overall criminal violence. Second, the focus on the types of actions changes over time; in this case, ISAF's emphasis on documenting smaller-scale attacks likely changed as its strategic priorities changed from counterinsurgency (COIN) operations to rebuilding campaigns. Third, data collection depends on awareness of incidents by the agency; in this case, the number of incidents may be correlated with the level of troop presence in an area. We address these concerns in a number of ways, including by controlling explicitly for the presence of ISAF troops in a given area over time.

2.2.2 Afghan Attitudinal Measures

Our three attitudinal outcomes (support for government, support for anti-government elements (AGE), and community cohesion) are derived from the 26 quarterly ANQAR surveys using multiple questions for each measure. Questions and response options varied as the survey went through these multiple waves, so we compiled data from questions that were exactly the same or sufficiently similar and coded response options to provide consistent measures over time. We took these multiple questions and then constructed indices relevant to each of the attitudinal outcomes of interest. We discuss the construction of these indices below (and in greater detail in the Data Appendix).

Support for Afghan Government (GOV): Fifteen potential Support for Government Indicators were highly correlated and present in greater than 80% of ANQAR rounds (broadly, these questions consider the performance of national-, provincial-, and district-level government officials, overall, on development, security, etc.). For waves when individual questions were missing, responses were imputed at the individual level using responses from related questions

[16] This dataset was compiled by Andrew Shaver and Austin Wright of ESOC – publication introducing the data forthcoming.

that were present. Then, we took the first principal component of these responses (a measure of the shared covariation among them) at the individual level and the mean at the district level.

Support for Anti-Government Elements (AGE): There are fewer questions that asked about support for AGE and only one of these is consistently present throughout the 26 rounds. This question asks, "In your opinion, if the Taliban were to return to power and govern Afghanistan, would this be a good thing or a bad thing for the country?" Responses to this question ("good" or "bad") were averaged at the district-quarter level to get a measure of support for AGE. Unsurprisingly, support for government and support for AGE are significantly and negatively correlated.

Community Cohesion Indicator (CC): Questions related to Community Cohesion were less frequently asked. For this reason, our community cohesion indicator only begins in the second quarter of 2011. Community Cohesion and resilience are fairly amorphous concepts, but based on a reading of the literature and interviews with stabilization program implementers, the goal of community cohesion is 1) to improve attitudes toward and performance of local governance, 2) to provide mechanisms for local dispute resolution, and 3) to maintain security in the community. We developed an indicator for each of these three components of community cohesion. First, we generated a Principal Component Analysis (PCA) index of support for local government (using several measures and the same method used to generate the Support for Government Indicator). Second, we gathered responses from questions asking, "If you had a dispute, who would you take it to?" (Responses indicating a willingness to take the dispute to either a local Shura/Jirga or a state court at the individual level demonstrate the availability of local conflict resolution mechanisms.) Third, our measure of community resilience was generated from a survey item asking whether or not respondents believed that the government would be capable of maintaining security in the future. These three sub-indicators were then aggregated using the same PCA into our main Community Cohesion Indicator. Each of the three sub-indicators is significantly and positively correlated. As with our other attitudinal measures, these measures were aggregated at the district-quarter level for analysis.

2.2.3 Economic Outcomes (ECO)

Measuring economic activities in dispersed or conflict-affected areas has always been problematic. Intensity of economic activity, density of interactions, and other important indicators are difficult to measure on a regular basis. For decades, researchers have tried to measure human activities (such as markets) using satellite photography generated via space flight programs. This relationship between measured activity and the imagery varied regionally and was typically attributed to societal and economic differences. However, this type of remote sensing measurement was not widespread due to expense and logistic concerns. This has changed in the last few years, as the cost of measuring intensified nighttime lights observed through satellite imagery has decreased dramatically. In particular, the Defense Meteorological

Satellite Program Operational Linescan System (DMSP OLS) has the capability to detect low levels of visible near infrared radiance at nighttime, making it possible to detect cities, towns, human settlements and even ephemeral gas flares and fires.[17] Research has shown a strong and robust relationship among observed areas with greater lighting, population density, gross domestic product and electric power consumption in over 21 countries.[18] However, we caution that nightlights should be interpreted as the resultant of human interaction and density, which can combine the effects of economic activities, population density and even violent incidents and political activism.

Despite this caveat, nightlights as an indicator for economic activity are particularly useful in Afghanistan, with its remote and difficult-to-access areas. Based on research by Elvidge and colleagues at the National Oceanic and Atmospheric Administration (NOAA), detailed evidence now exists that validates nightlights as a measure of activity in areas with as few as 150 inhabitants.[19] Country or state *total lit area* has been shown to correlate highly with World Resources Institute (WRI) statistics such as Gross Domestic Product (GDP), urbanization, electric power consumption and more.[20] Using survey techniques alone it would be extremely difficult to track these populations. In the following sections, we estimate the relationship between other measures of economic activities and nightlights based on two sources of satellite imagery.

The first source is the DMSP, collected from 2008 to 2013, which contains the visible and thermal infrared data (day and night) to form yearly composite images. The composites contain lights from sites with persistent lightings.[21] The second source is the Visible Infrared Imaging Radiometer Suite (VIIRS) satellite, which replaced the DMSP in 2014. VIIRS is a more sophisticated technology able to report actual radiance values in Nano Watts/cm2/sr[22]. The

---

[17] Cinzano, P., Falchi, F., & Elvidge, C. D. (2001). The first world atlas of the artificial night sky brightness. Monthly Notices of the Royal Astronomical Society, 328(3), 689-707.

[18] Elvidge, Christopher D., et al. "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption." International Journal of Remote Sensing 18.6 (1997): 1373-1379. These results have been demonstrated in a wide range of countries, including relatively small economies such as Suriname, Guyana, and Grenada.

[19] Christopher N. H. Doll, Jan-Peter Muller, & Elvidge, C. (2000). Night-Time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions. Ambio, 29(3), 157-162. Retrieved from http://www.jstor.org/stable/4315020

[20] Elvidge, Christopher D., et al. "Mapping city lights with nighttime data from the DMSP Operational Linescan System." Photogrammetric Engineering and Remote Sensing 63.6 (1997): 727-734.

[21] Ephemeral events such as short term fires or flares are discarded. The background noise is replaced with 0 and data values of radiance range from 1-63. Additional details on specific processing decisions are included in the Data Appendix Section A.3.

[22] The two measuring systems report outputs on different incomparable scales. But instead of trying to recalibrate scales of one system to include another, we have avoided the calibration issue by using the two brightness measures as separate independent variables. Moreover in addition to the raw regression year and quarter fixed effects have been taken into account.

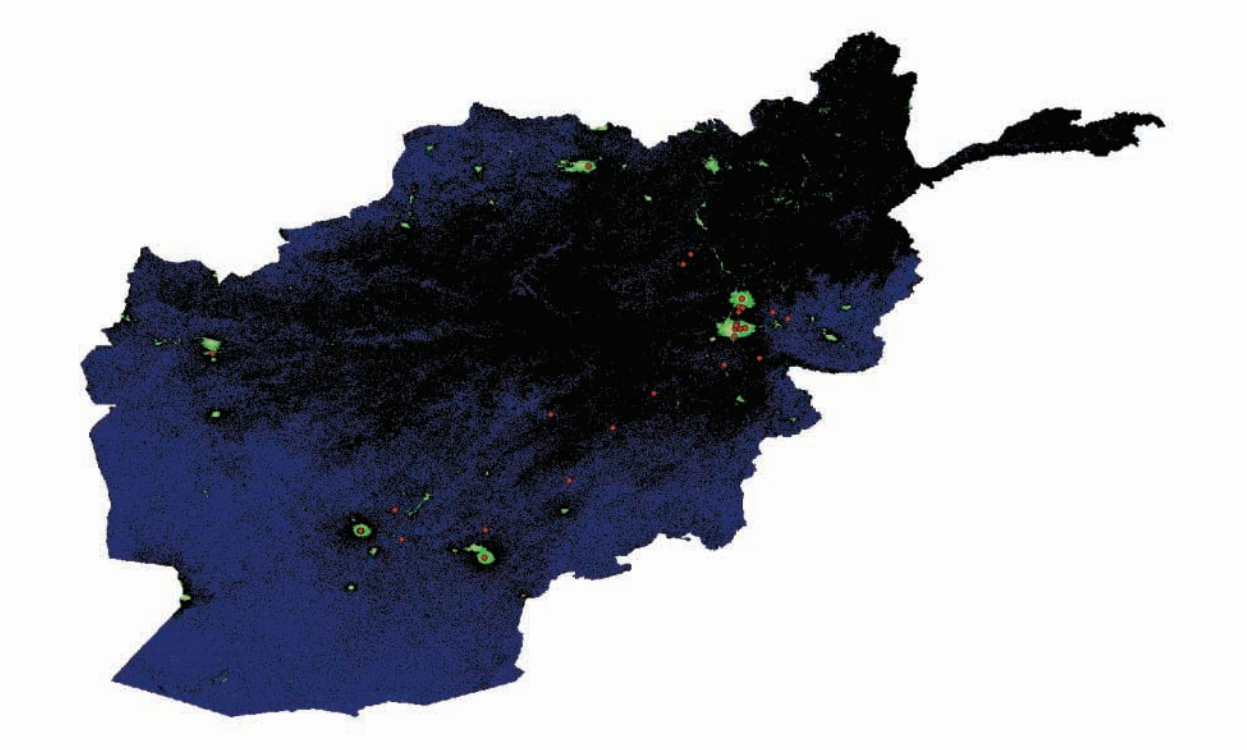
composite images thus comprise nighttime light in an area filtered to exclude lightning, lunar illumination, and cloud cover. Both the DMSP and VIIRS geo-referenced files are analyzed in a standardized projection space to facilitate comparability.[23] However, the images are not directly comparable, generating a break in the time series, but still permitting regional, within-year comparison.

We mapped the geolocated nightlight measures to a political boundary map so that each pixel is associated with a district and province. The values of these points within each district are aggregated to form a measure of overall nightlights present and normalized by population. This normalization is critical as areas with sparse population will tend to have low night illumination regardless of the local economy. However, areas with higher population and low nightlight illumination likely do have limited economic activity and changes in nightlights can be detected even in low population areas, allowing for measures of economic conditions over time.

In Afghanistan, a unique confounding factor was military presence. Nightlights in certain locations were potentially related to military and troop presence and density rather than economic or social activity of Afghans. To test this theory, we examined the correlations between violence levels, troop presence (ORBAT), and nightlights. As expected, there is a positive relationship between troop presence and violence as well as troop presence and nightlights. The map, below, shows the brightest lit areas in the country and troop presence in red. While there is some overlap, and thus we control for troop presence in a subset of the subsequent analysis, there are a number of bright areas with relatively low troop presence and military intensive areas without significant nightlights.

[23] The difference in the weekly and monthly correlations suggests that nightlights may be more correlated with overall economic activity across households rather than week-to-week variation in household income. Specifically, both are forced onto the WGS84 projection.

Figure 2.1: Afghanistan Nightlights with Troop Presence



To better understand how nightlights are related to economic conditions in Afghanistan, we cross-validate these measures with economic measures from several different surveys to establish the degree to which nightlights correlate with key measures, such as weekly household income and measures of household wealth/assets. We find a robust correlation of over 60% between nightlights and weekly income and household assets (for monthly income, the correlation is over 70%), confirming that our use of nightlights for measures of economic activity is reasonable. [24]

2.2.4 Health Access Outcomes

We initially sought to use the Demographic Health Survey (DHS) data to better understand health and well-being of the Afghan population. The DHS is a robust, internationally recognized, and regularly validated survey with a host of relevant health questions that can be tracked over time and across Afghanistan. However, the latest DHS wave results (2015) were not publicly available at the time of this analysis, and the previously available DHS-like survey, the Afghan Mortality Survey (DHS Special Survey, 2009) were not appropriately timed with the available program data to facilitate the broad panel analysis.[25] The next best source was the NRVA, which has a section on maternal and child health. The relevant questions from this NRVA section were

[24] Details on comparisons are presented in the Data Appendix Section A.3.

[25] Afghanistan Special, 2010 - Mortality Survey - and as AMS in the report released.

sufficiently correlated (over 50%) with corresponding DHS responses; however, there is insufficient temporal coverage—being only two years and prior to most stabilization projects in our analysis to identify a clear relationship between stabilization spending and health outcomes using these data.

Instead, we focus on a small number of health-related questions that were asked uniformly across all the waves of the ANQAR. Our health measure is constructed using the district average of the question "How would you rate the quality of healthcare available in your area?" We note that this question is related to perception of health services, not health outcomes, and primarily relates to the perceived quality (and availability) of these services. However, this question correlated at the 60 percent level with a host of DHS and NRVA questions and is the only regularly collected and available data on health outcomes.[26]

2.3 Variable Construction for Program Variables

The other half of the broad panel analysis relies on programmatic data, which contains information on the size, scope, and locations of specific programs. To measure program activity, it is critical to identify metrics that are consistently reported across a broad range of projects, objective in nature, and plausibly linked to the level of activity and delivery of goods/services on the ground. The most commonly used measure meeting these criteria is total project spending. We use this measure, but must acknowledge its significant shortcomings. Our first caveat is that project spending is not a perfect measure of project activity or intensity. In many areas of Afghanistan, variation in spending may be due to security-related operating costs instead of variations in programmatic activity. For example, higher project spending may occur in areas with higher violence simply because the cost of security provision is drastically increased over areas with lower threat levels.

In order to account for the discrepancy between dollars invested and effort made, we created a simple indicator variable on whether any *program* dollars were spent on projects in that district-quarter. We term this analysis "project presence" analysis to differentiate it from the project spending analysis. The difference between the correlation between spending and outcomes and the correlation between presence and outcomes is useful for diagnosing whether observed correlations are reliable indicators of a relationship between programmatic activity and the outcome or whether a separate variable is at work (e.g., the increased costs of operating in high-violence areas has misrepresented the activity level of stabilization programming and while the outcomes are not correlated with project spending, they are correlated with project presence).

Our intuition is that project presence may more accurately capture the direct effects of project activity, separate from the higher spending rates in violent areas. For example, if project

[26] Detailed presentation of this cross validation is presented in the Data Appendix Section A.4.

presence were observed to correlate with improved attitudes toward the Afghan government, we might still see that project spending was associated with negative attitudes toward the Afghan government. This could indicate that project presence improves attitudes but in high violence areas the effect is masked because that level of violence reduces support for the government overall. The project presence variable also helps correct for the initial selection bias in certain areas; for example, if projects in a certain area are associated with higher levels of violence, we may assume this was a deliberate decision to locate that project in violent areas. In addition, project presence helps capture the effects of stabilization spending without assuming projects with higher associated funds actually executed more activities. On the other hand, the spending measure, while imperfect, captures some information on how the size of a program is associated with the outcomes.

Another caveat regarding program data is that project spending may miss key aspects of project quality or structure that are relevant for efficacy and impact. More specific measures are not consistently measured across projects, however, so we are not able to include such measures in our broad panel analysis.

We combined usable data from AI with information on program activity from the NSP, OTI, CERP, and MISTI, where relevant.[27] Thus we use both project spending and project presence as our measure of program activity. For each of the datasets we aggregated cumulative spending in a district-quarter across all projects, visually inspecting resulting data for extreme outliers that are likely the result of data entry error. We also constructed a project presence indicator to measure the activities of programs in a given district-quarter.

2.4 Methodological Approach

After combining the outcome and programmatic data into a single dataset that measures each variable in every district-quarter from 2009 to 2016 (data permitting), we then conducted the "broad panel" analysis. The benefit of this analysis is two-fold. First, the analysis reviews and assesses the varying quality of specific variables—outcome measures, programmatic management, and descriptive sociodemographic data—to create consistent and replicable measures to be used for this and future analysis. Second, the analysis combines information from a range of survey, administrative, and geospatial data sources to enable a holistic view of the

[27] The PAP cited Afghan Info as the primary source of programmatic data. Based on a detailed review, including interviews with USAID officials familiar with the database, we identified a number of issues with data quality and fidelity that limited its utility for analysis. First, it is not the financial system of record for the Mission, nor was it ever intended to be and could not be used to account for USAID's spending in Afghanistan. Second, Afghan Info only began tracking USAID appropriations in 2011 and does not track all USAID accounts. Third, it did not include humanitarian accounts (including OTI). Fourth, while Afghan Info does track most Economic Support Funds (ESF), it does not track ESF transfers to other agencies, including USDA, PAS, Treasury, and even other parts of USAID like OTI. Moreover, much of the detailed, sub-national level spending information that would be required for this kind of measurement and evaluation is kept by implementing partners or contractors. For a more detailed discussion see Iyengar, Shapiro and Mao (2017).

range of outcomes and correlative factors that may affect the impact of stabilization programs.

We conducted the analysis in several steps. First we explored the broad trends nationally and regionally in the outcome variables and separately in the programmatic variables. These trends are critical in establishing a baseline of the degree of variation in each of the measures. We discussed the overall trend, percentage change, and volatility of the measures to understand the degree to which these measures identify improvements or decline over time as well as regional variation in outcomes. Second, we conducted a series of regression-based analyses to measure the relationship between the outcomes of interest and measured program activity. These estimates have the benefit of systematically and explicitly controlling for different sources of variation including time, location, military presence, geographic features, sociodemographic features (e.g., ethnicity), and even cross-correlation of some of the outcomes (e.g., violence impacting other outcomes of interest).

While there is important information to be gained from these regression-estimated conditional correlations, we note explicitly that we cannot claim to estimate the causal relationships between any of the outcomes of interest and project spending. Many of the outcomes and the measured level of project activity affect each other in a variety of ways that our analysis does not address. For instance, areas that are more violent may have higher levels of stabilization project activity in order to address some of the causes of the violence and instability. A correlation between the spending and violence may therefore appear positive, because the spending is actually intentionally higher in those more violent places—not because the spending *causes* higher levels of violence. We cannot, in other words, cleanly establish what would have happened in the counterfactual scenario in which the program did not happen.

To better contextualize these findings, we also present findings from two other sources. First, we conducted a research review to identify common themes and broad lessons from research spanning academic, government, and policy literatures.[28] Second, we interviewed program officers, contract officers, and program staff.[29] The interviews were very useful for providing an "on the ground" perspective for designing and implementing stabilization programs in Afghanistan. The participants provide a complementary perspective on issues identified in the research review and quantitative analysis.

2.5 Summary of Data and Approach

This chapter presented the data sources and approach to analysis. The five outcome variables of

[28] See Iyengar, Shapiro, and Hegarty, 2017 for additional details on methods and detailed findings.

[29] Interviews were conducted with a set of self-selected participants based on a list of 24 individuals identified by USAID. These individuals were selected based on their familiarity with a set of stabilization programs implemented in Afghanistan by USAID between 2009 and 2015. Based on the responses to our request, we conducted 13 interviews between October 27 and December 2, 2016.

interest used in subsequent analysis are: violence (VIO), support for the Afghan government (GOV), support for Anti-Government Elements (AGE), community cohesion (COM), health (HEA), and economic well-being (ECO). These outcomes vary in both the length of coverage and the frequency with which they are collected.

Figure 2.2: Outcome Coverage over Time

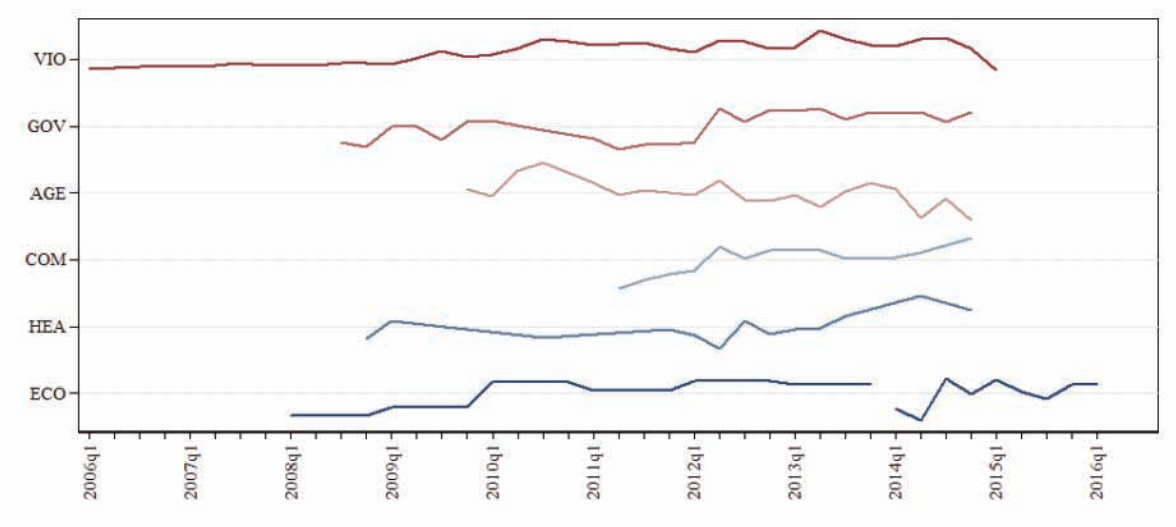
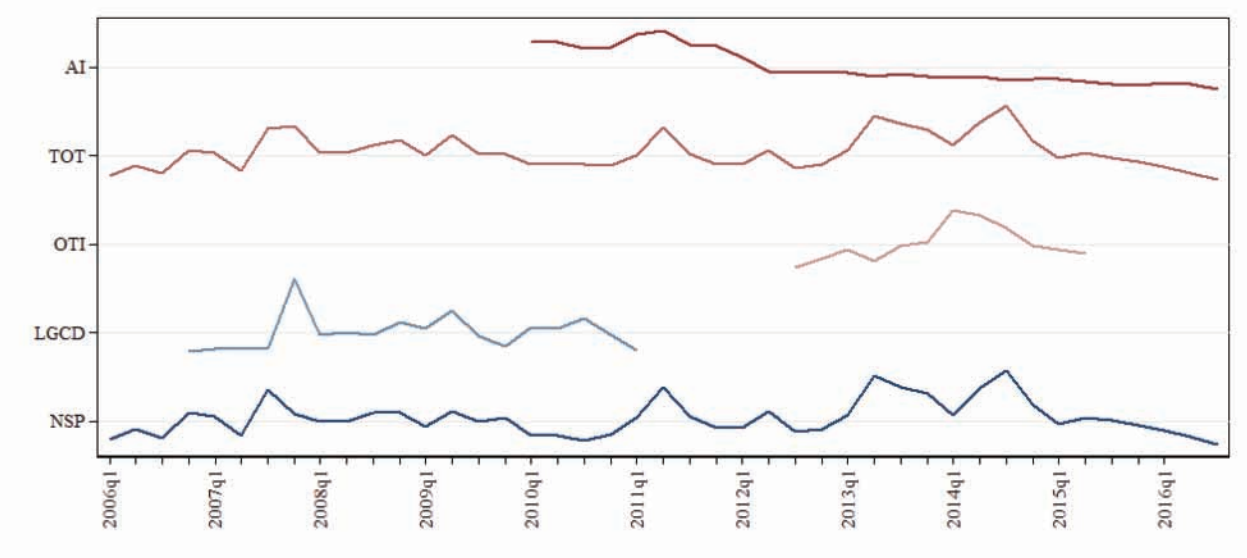


Figure 2.2 illustrates the coverage over time, meaning that it indicates whether that outcome variable was measured in any given quarter between 2006 and 2016. The longest time trends are available for violence and support for the Afghan government. As previously discussed, we a limited set of years for which we can measure community cohesion. This is due to the small number of consistent questions across survey waves. We also have a break in coverage for the economic measure (nightlights) from the improved detection capability of the VIIRS satellite. Figure 2.3 shows the coverage of program activity over time. While the individual programs themselves span a range of different periods, overall, we have information on programmatic activity for nearly all of the time period over which we observe the trends in outcomes.

Figure 2.3: **Treatment Coverage over Time**



Chapters 3 and 4 present the trends in these variables separately and then together, including a range of confounding and potentially explanatory variables. These approaches allow us to consider the systematic relationship between the outcomes, project activity, and a host of relevant covariates (e.g., ethnicity, agricultural activity, amount of infrastructure) in a holistic manner, but should not be considered credible estimates of a causal effect.

3. National and Regional Trends in Stabilization Program Activity for Violence, Attitudes, and Well-being in Afghanistan

This chapter presents trends in the outcomes over time and location for both the programmatic data and the outcome variables of interest. There are three objectives of this analysis: (1) to present the broad trends in outcomes within which all of the findings should be contextualized; (2) to illustrate the period of time and areas of Afghanistan that are under analysis in subsequent chapters; and (3) to highlight the overlap between the programmatic measures and outcome measures, which is relevant for the analysis presented in Chapter 4. We present the analysis for program activity and outcomes separately.

3.1 Nationwide Trends in Outcomes

As discussed in Chapter 2, there are five outcome variables of interest: violence (VIO), support for the Afghan government (GOV), support for insurgent and other anti-government elements (AGE), community cohesion (COM), health (HEA), and economic well-being (ECO). The data on each outcome variable vary in time coverage. Table 3.1 shows the summary measures of these variables, including mean and standard deviation. The first two rows show the number of violent incidents per district-quarter in raw form and then scaled by population (per 10,000 people). One clear point in both the unscaled and per capita (PC) measures is that violence varies significantly over time and location. In fact, more than a quarter of the district-quarter observations have no violent incidents. On the other hand, the top one percent of district-quarters has over 45 attacks per 10,000 people—12 times the average.

Table 3.1 also helps highlight that the unit of measure in the index measures (GOV, AGE, COM and HEA) are limited in their direct interpretation. That is, a score of "two" on "support for the Afghan government" is not directly meaningful. Rather, the index provides insight into the relative degree of support in an area compared to support over the entire nation and time period of coverage. The same is true of the economic measures using nightlights.

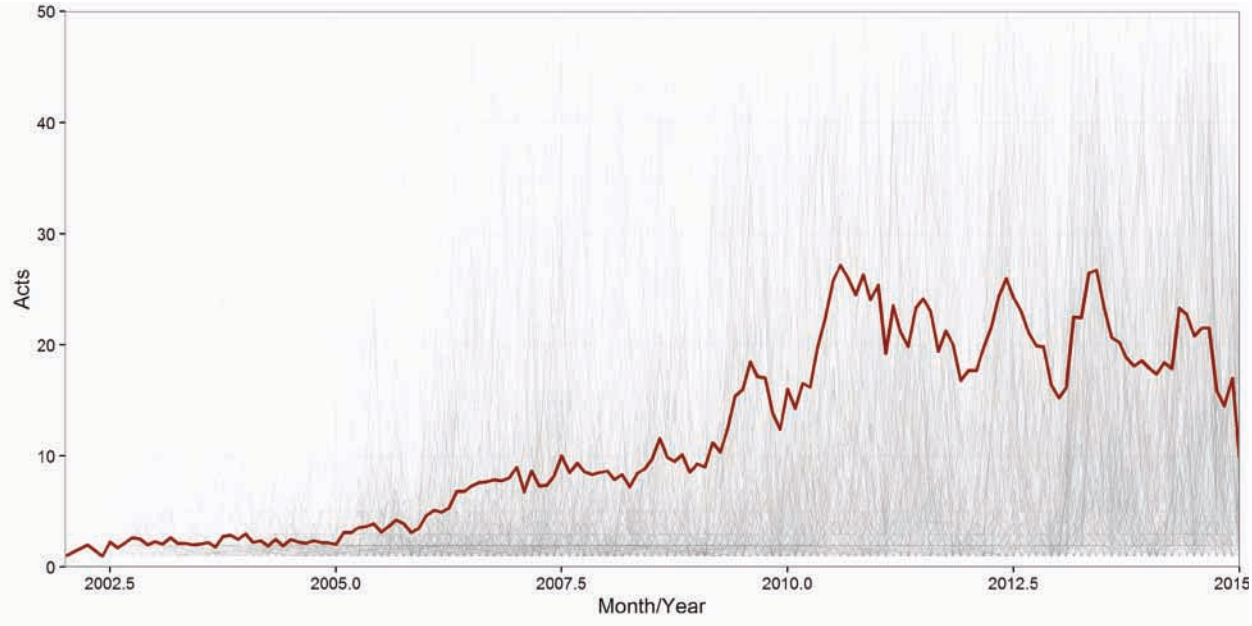
Table 3.1: Outcome Measures Summary Statistics

Variable	District-Quarters	Mean	Std. Dev.	Min	Max
VIO (PC)	21,306	3.67	9.99	0	220.79
VIO	21,306	23.37	90.08	0	3,086
GOV	6,711	0.48	1.57	-5.18	5.97
AGE	5,485	0.04	1.14	-1.09	3.92
COM	4,039	-0.02	0.72	-2.99	2.22
HEA	3,690	-0.15	1.16	-4.39	3.30
ECO (VIIRS)	3,618	0.03	0.15	-0.00	4.96
ECO (DMSP)	9,648	0.00	0.018	0.00	0.37

Focusing again on violence, Figure 3.1 shows the average number of events by region and a per capita measure of violent events. The thicker, maroon line represents the national average (unweighted by district) over the same period. Over time, we see a general increase in levels of violence; however, we see a marked drop toward the end of the period covered. This could be related to lower overall levels of violence and/or reduced coverage by ISAF forces during the planned drawdown and end of NATO combat operations. To validate this measure and determine if there was a potentially problematic observation bias, we compiled a related indicator from the ANQAR survey data. We aggregate district-quarter responses from the survey item "How is the security situation in your Mantaqa [*village*]?" When compared to the violence indicator generated from SIGACTs, we find that an increase in the level of violence is negatively correlated with self-reported measures of security suggesting that while there may be some undercounting of specific types of criminal incidents, overall SIGACTs remains a valid measure of violence.

As illustrated in Figure 3.1, the greatest number of violent incidents occurred in the year 2010, with the number of significant actions nearly double that of 2009 and more than double that of 2011. Violence levels varied by 58 percent each year on average, and by as much as 182 percent between 2004 and 2005. There were more than 147,000 enemy action events logged between 2004 and 2015. The next most frequent type of incident was explosive hazard, with more than 108,000 incidents, followed by friendly fire actions with approximately 68,000 incidents.

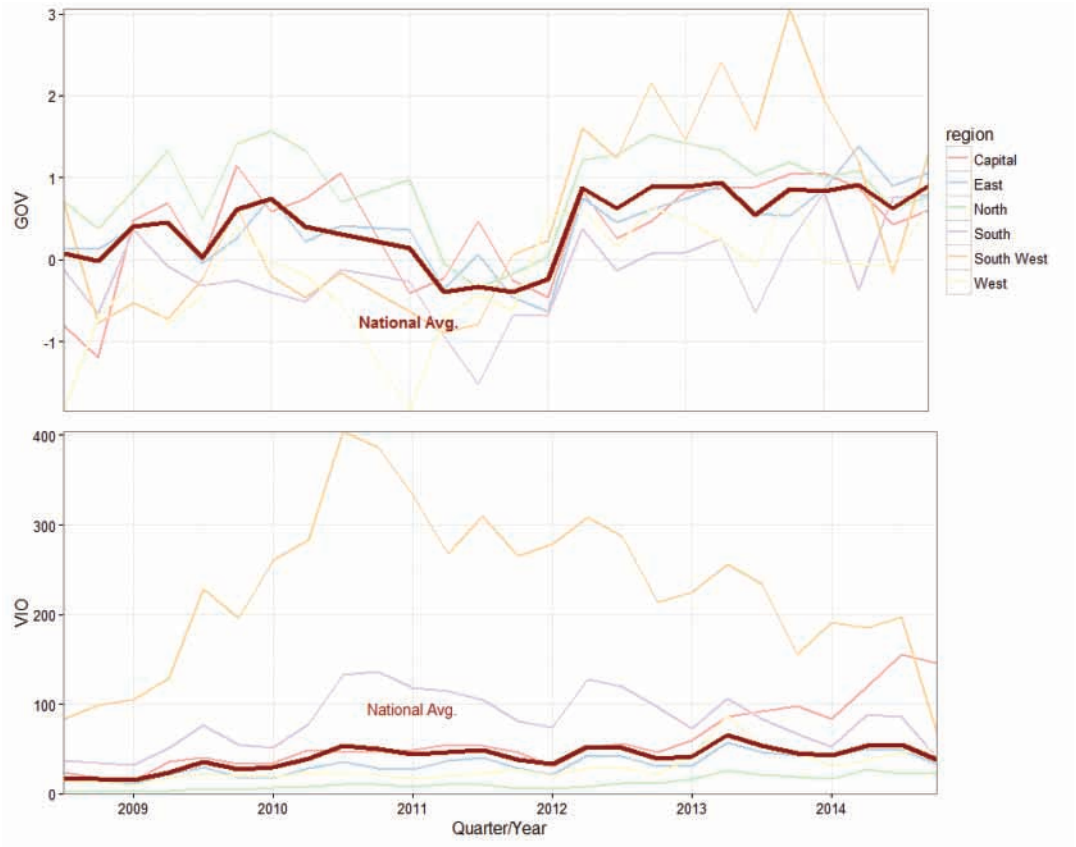
Figure 3.1: Trends in Violence over Time



To put this in context, consider the estimated causal effect of NSP, for instance, on violence. Evidence from a detailed analysis suggested no significant impact on violence level in NSP villages versus other comparison villages (Beath, Christia, and Enikolopov, 2015). This appears to be a consistent result when reviewing evidence on the impact of stabilization programs (see Iyengar, Hegarty, and Shapiro, 2017). Similarly, negative impacts—for example a potential rise in violence due to the targeting of stabilization programs (as indicated in some evaluations, see for example, MSI 2015)— is also extremely small relative to the aggregate number and frequency of attacks reported during the project time. This is not to suggest that these programs did not impact these or other measures but rather that their impact in the broader context of the conflict was likely negligible in the national context. In later analysis considering the relationship between stabilization and violence, it is important to understand that even if stabilization programs were to have an effect on violence (either increase or decrease), that effect is likely much smaller than average violence levels experienced by many people in Afghanistan over the same time period.

These broad trends in violence drive a number of other key indicators and thus, consistent with the broader literature, we treat security as not only an outcome of interest but also as a key factor that affects other outcomes of interest and broader trends. Figure 3.2 tracks the trends in support for government over the period of the survey and the number of violent events per capita (within district) over the same period from our SIGACTs data.

Figure 3.2: Trends in Support for Governance and Violence over Time



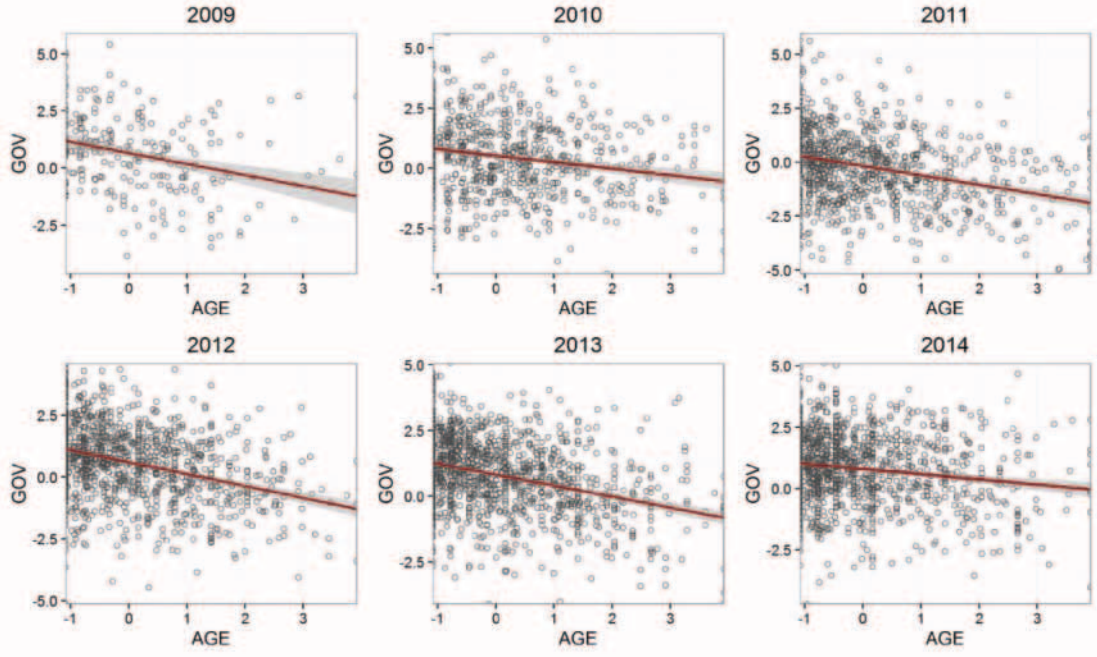
The general trends appear negatively correlated; a peak in average quarterly violent events in the second quarter of 2009 (2009.5 on this graph) is matched by a dip in support for the government.[30] The level of government support changes significantly over time, declining precipitously from its peak in 2010. While some of the decline in support occurred over the same time period as the increased violence seen in Figure 3.1, existing research on stabilization also suggests that corruption in the Afghan government was a key determinant in support for the Afghan government (Iyengar, Hegarty, and Shapiro, 2017). While this was certainly a factor in the implementation of a number of stabilization programs, absent widespread reform

[30] There is a small decline between 2011 and 2012 likely due to changes in the design of the survey from which this measure is derived. We address this with controls for survey wave in the analysis in Chapter 4.

in the Afghan government, it seems unlikely that any individual program would have a sizeable enough effect on attitudes to overcome this broader trend. Overall, however, government support appears to have increased slightly from 2009 until 2015, a trend which could be influenced by any number of factors and for which we cannot posit causality.

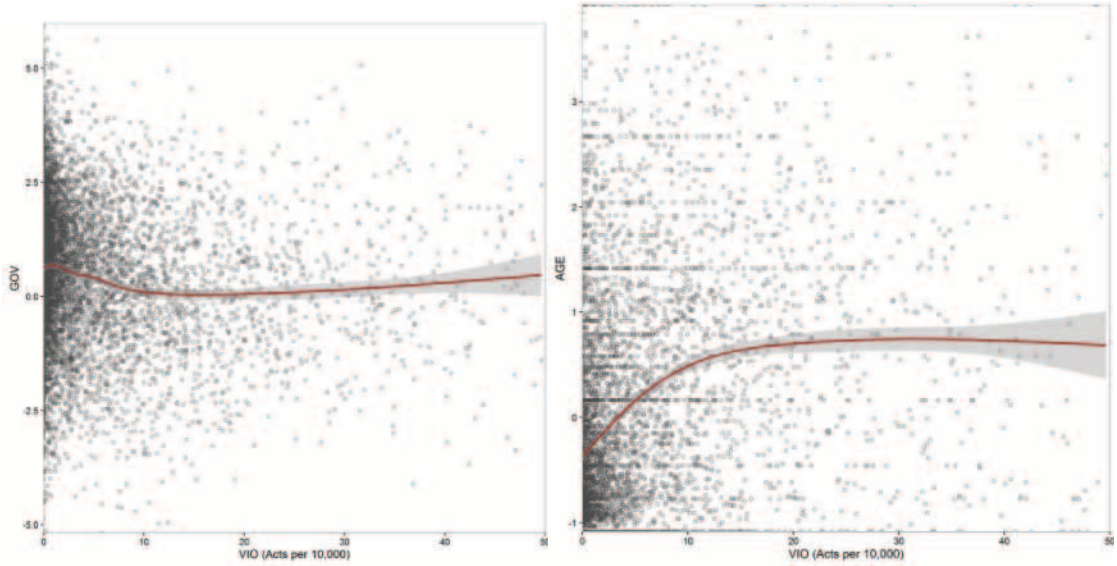
Not surprisingly, support for AGE is inversely correlated with support for the government. As shown in Figure 3.3, increased support for anti-government elements was associated with a decrease in support for the government. This relationship is consistent over time.

Figure 3.3: Support for AGE and Support for Government



Support for government and anti-government entities also varies in response to the level of violence. Figure 3.4 demonstrates the variation in each indicator as violence increases. An increase in violence is associated with a decrease in support for government.

Figure 3.4: Support for GOV and AGE over Violence



. In contrast, violence is positively and strongly correlated with support for AGE at low to mid-levels of violence. It is important to note here that these trends should not necessarily be interpreted as an increase in violence leading to increases in support for AGE. Programs that were specifically targeted at especially violent areas would in many cases also be in areas with higher than average reported support for AGE. This higher than average level of support for insurgents may be larger than any program-specific effects on attitudes in many evaluation designs and could potentially mask gains made by stabilization efforts. This also does not mean that such gains were made but rather even if gains were made, they would be difficult to detect.

We do observe modest but steady increases in economic activity as shown in Figure 3.5. The upper figure illustrates average radiance for the whole country and the lower figure excludes Kabul. The red reference line illustrates the 2008 national average to help illustrate how trends have changed since that baseline date. Overall, there appears to be a modest but steady increase.

Figure 3.5: Annual Economic Trends

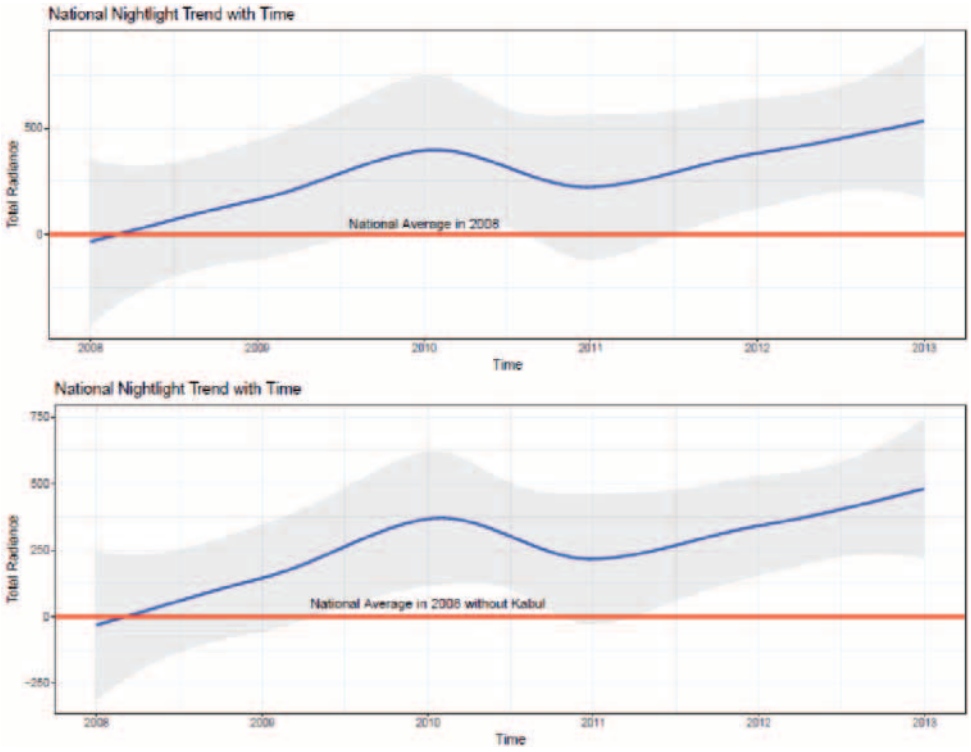
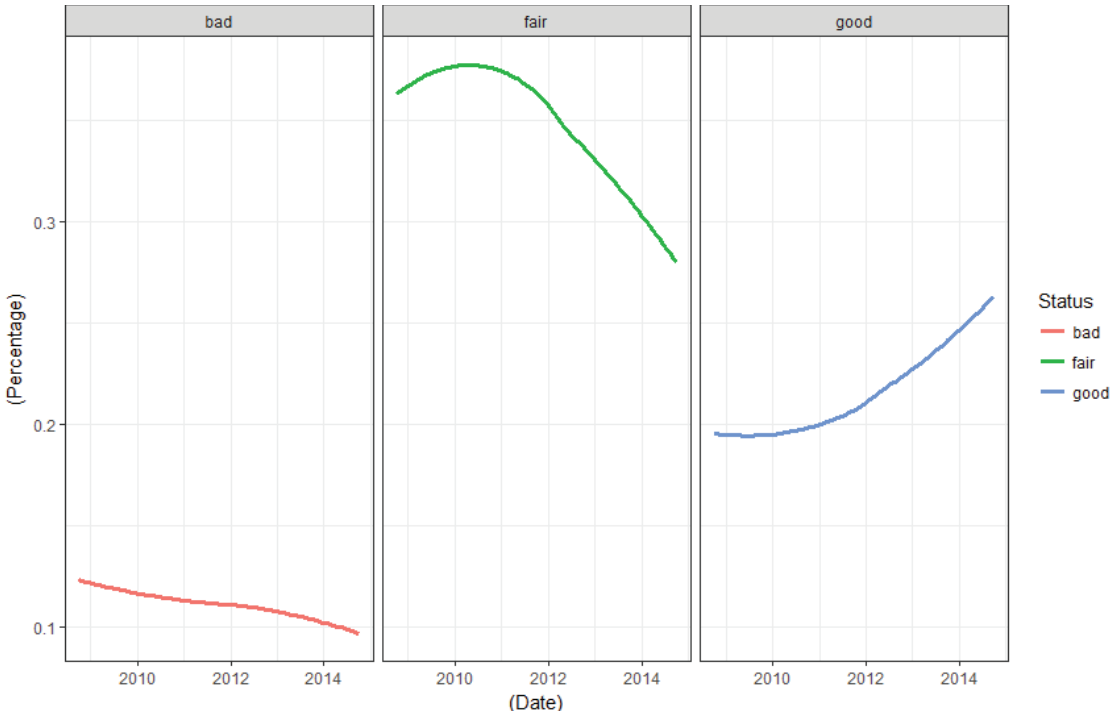


Figure 3.6 shows trends in the measure of healthcare quality over time. The leftmost graph shows the trend in individuals reporting bad or very bad healthcare access, the middle panel shows fair healthcare and the rightmost panel shows those reporting "good" or "very good" healthcare access. Overall, the percent of people reporting good or very good healthcare access increased by 5% (from just below 20 to 25 percent). This appears to occur at the same time those reporting very bad, bad and even fair access to health services has declined.

Figure 3.6 Perceptions of Quality of Healthcare Responses 2010-2015

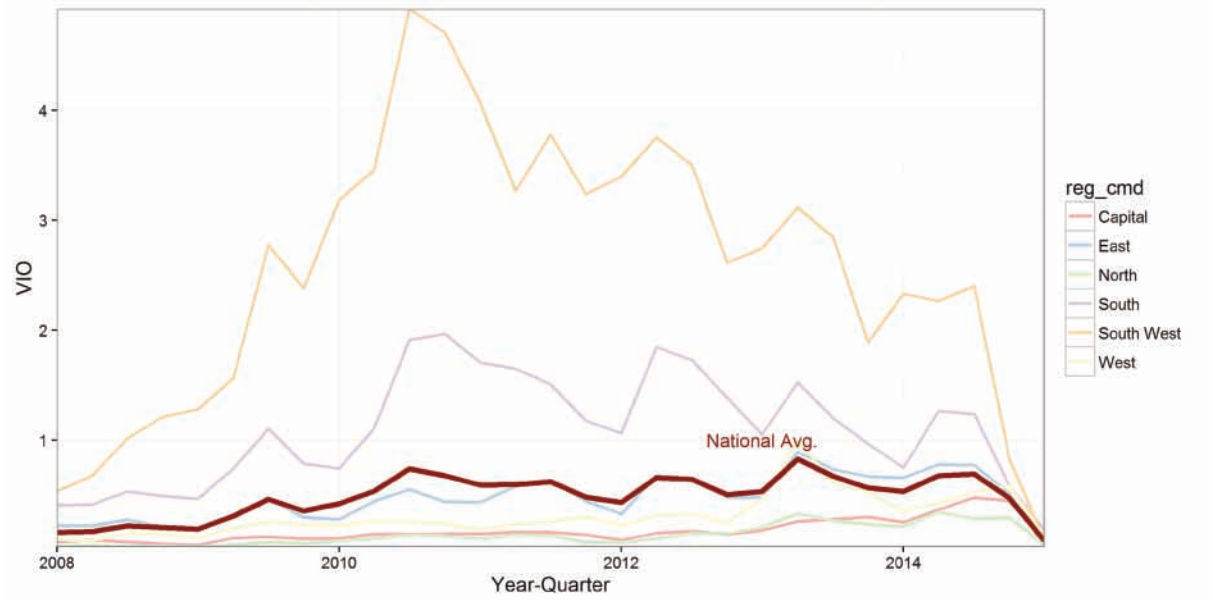


The trends also suggest small but statistically detectable increases in support for the Afghan government, healthcare access and economic activity. While none of the gains are very large, the steady increase over time may suggest overall improvement in outcomes at the aggregate, nationwide level.

3.2 Time Trends and Regional Variation in Program Activity

Although the national trends are informative, there is a tremendous amount of regional variation in Afghanistan and the differences between the regions change over time. For instance, in Figure 3.1, above, the per capita measure of violence shows an average violence level rising and falling over time, peaking in 2010 and 2011 and declining between 2011 and 2016. The regional variation is as shown in Figure 3.7. Much like the national trends, violence is generally increasing between 2008 and 2011 and decreasing thereafter. The level of SIGACTs is higher in the South and South West compared to other parts of the country, with more violence occurring in these areas at all points in time. Moreover, while violence (as measured by SIGACTs) declines in the South West, and to a lesser extent in the South and East, it increases in 2013 and 2014 in the Capital region.

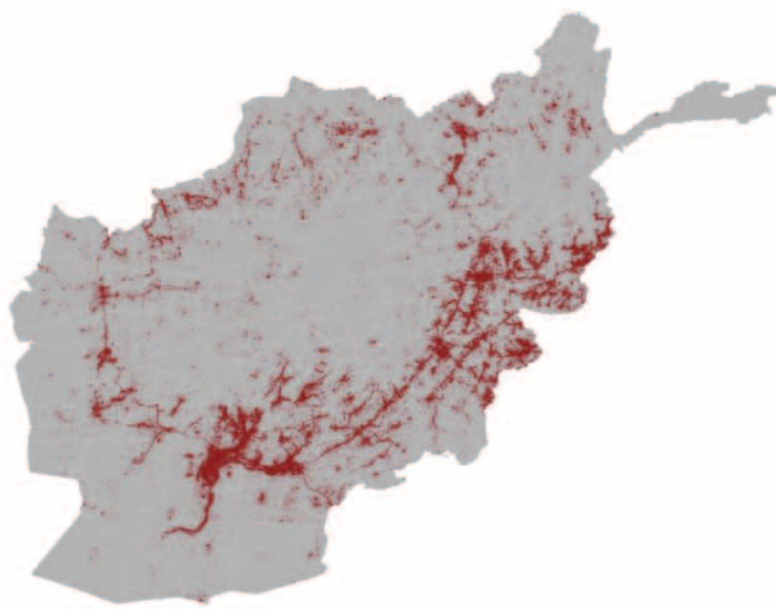
Figure 3.7: Violence by Region, 2008-2015



Some of this geographic variation is due to the concentration of incidents in areas that had more international forces operating in more populated areas or in areas with key travel routes. For example, Figure 3.8 highlights the geographical distribution of these incidents; each point in the figure represents an incident. The data suggest that the violent events are highly targeted –concentrating on specific road networks and cities– while the distribution across districts demonstrates that much of the conflict took place in the South and South west of Afghanistan, in Kandahar and Helmand Provinces in particular.[31] Some of the variation, however, in particular the sharp decline in violence in the South West that was not observed in the East, for example, cannot be explained by roads or population density. Additionally, we note that while violence declined overall between 2011 and 2015, there is an uptick in 2014 and 2015 in the Capital region.

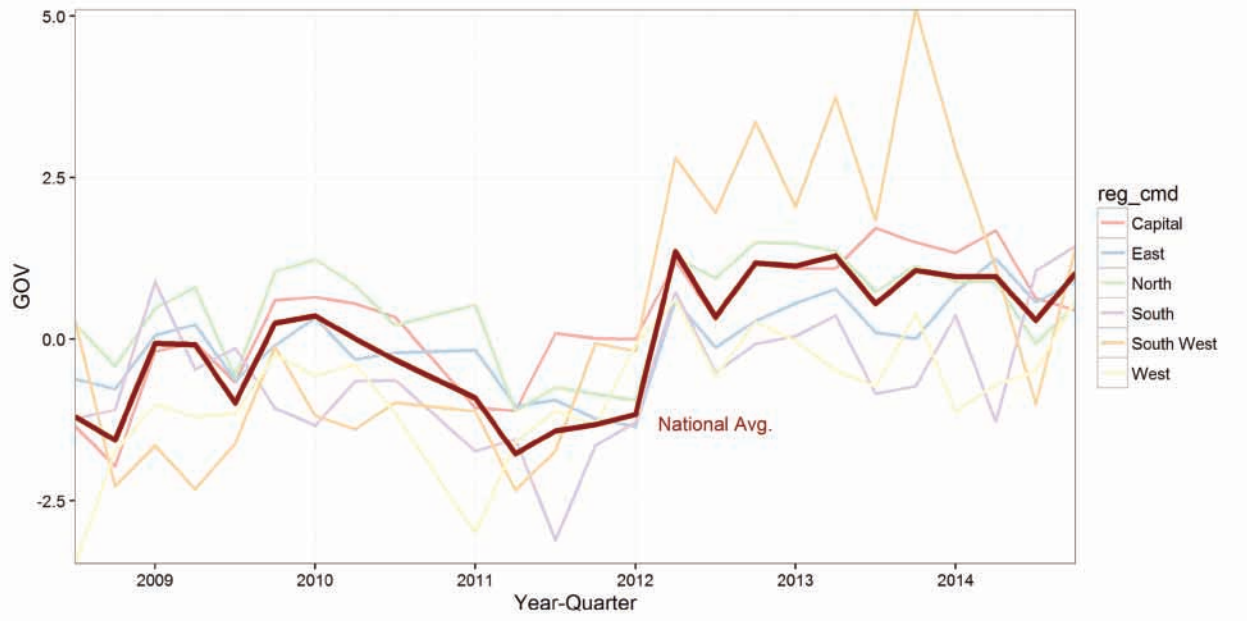
[31] We consider road density in some of the regressions presented in the results appendix.

Figure 3.8: Map of Spatial Distribution of Violent Events 2008-2015



In terms of support for the Afghan government, the estimates presented in Figure 3.2 suggest that it is increasing over time. However, there is also substantial regional variation in this measure. As illustrated in Figure 3.9, support for government appears to be highest in the South West after 2012, increasing over time despite the increases in violence over time. This regional variation is important in trying to understand how differences in trends might be associated with the broader differences in stabilization activity. In particular, the South West experienced a significant increase in program activity from 2010 to 2016 while other regions, such as the North and West did not. It is important to note that such regional variation in support for the Afghan government cannot be directly attributable to any individual program or activity, because of the difficulty in identifying an appropriate counterfactual with which to compare the trends. However, the broad differences in outcomes over time across regions provide potential insight into the aggregate relationship between concentrated programmatic and military activity in different geographic and sociodemographic environments.

Figure 3.9: Support for Government by Region



In contrast to the relatively substantial differences over time in support for the Afghan government, support for AGE appears fairly constant over time across region as shown in Figure 3.10. The notable exception to this trend is in the South West, which starts as one of the friendliest regions towards AGE and ends as one of the least. On the other hand, the South records consistently high support for AGE, relative to other regions.

Figure 3.10: Support for AGE by Region

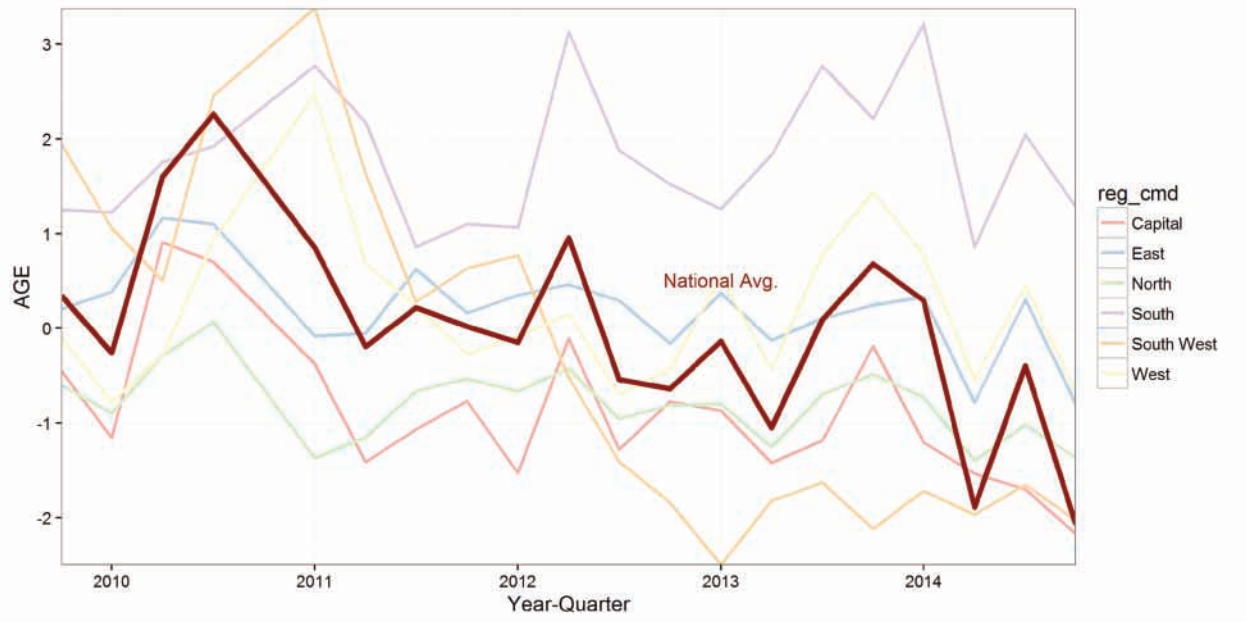
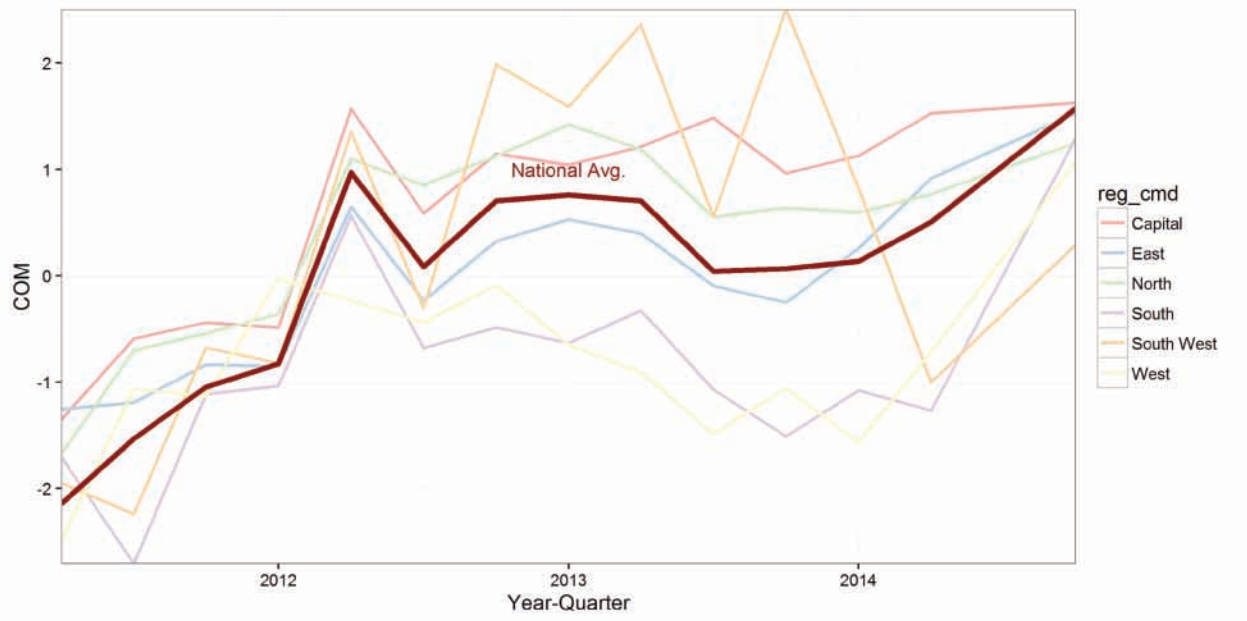


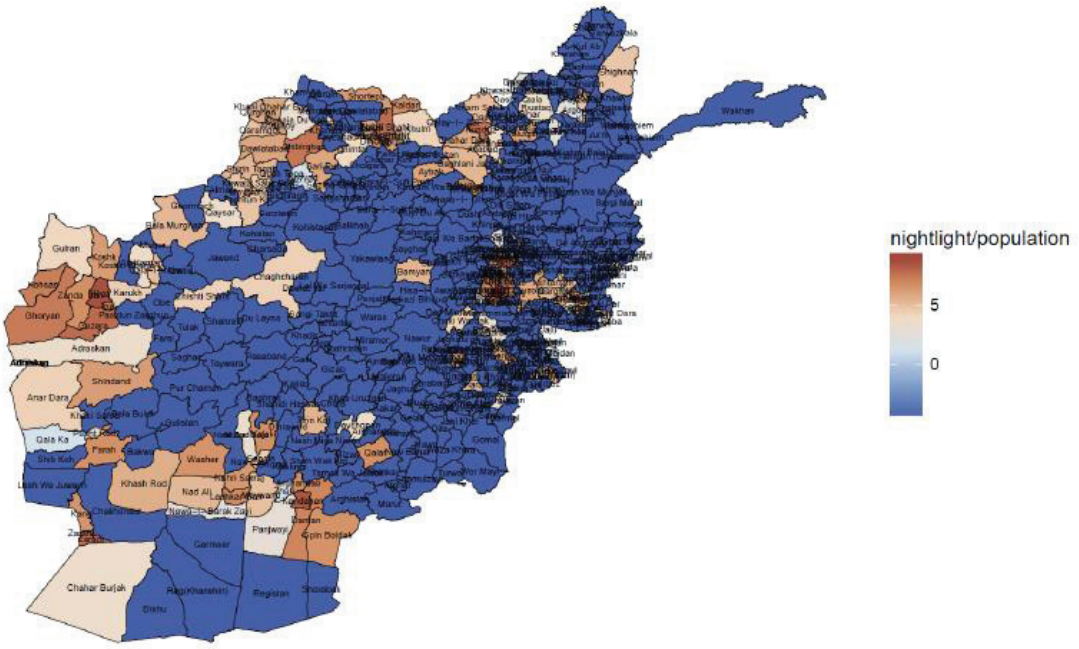
Figure 3.11: Community Cohesion by Region



We do not see this kind of progress in the South West on community cohesion. Figure 3.11 illustrates the community cohesion measure that was collected over a shorter time horizon than the other support measures. We observe positive trends in the cohesion over time across all regions with the notable exception of the South West region that saw some drop in cohesion toward the end of the sample.

There is also significant variation in economic activity (as measured by illumination) over region and time. Not surprisingly, there are a number of areas with low population and low illumination, as illustrated in blue in Figure 3.12. However, there are a number of areas that show relatively high levels of activity, as illustrated in the range of orange areas. In particular, the South West and West of Afghanistan, particularly in areas near the roads network, appear to have significantly greater economic activity than other regions of Afghanistan.

Figure 3.12: Geographic Variation in Nightlights by District in Afghanistan



Note: This figure shows the log of illumination per capita in each district in Afghanistan over 2010.

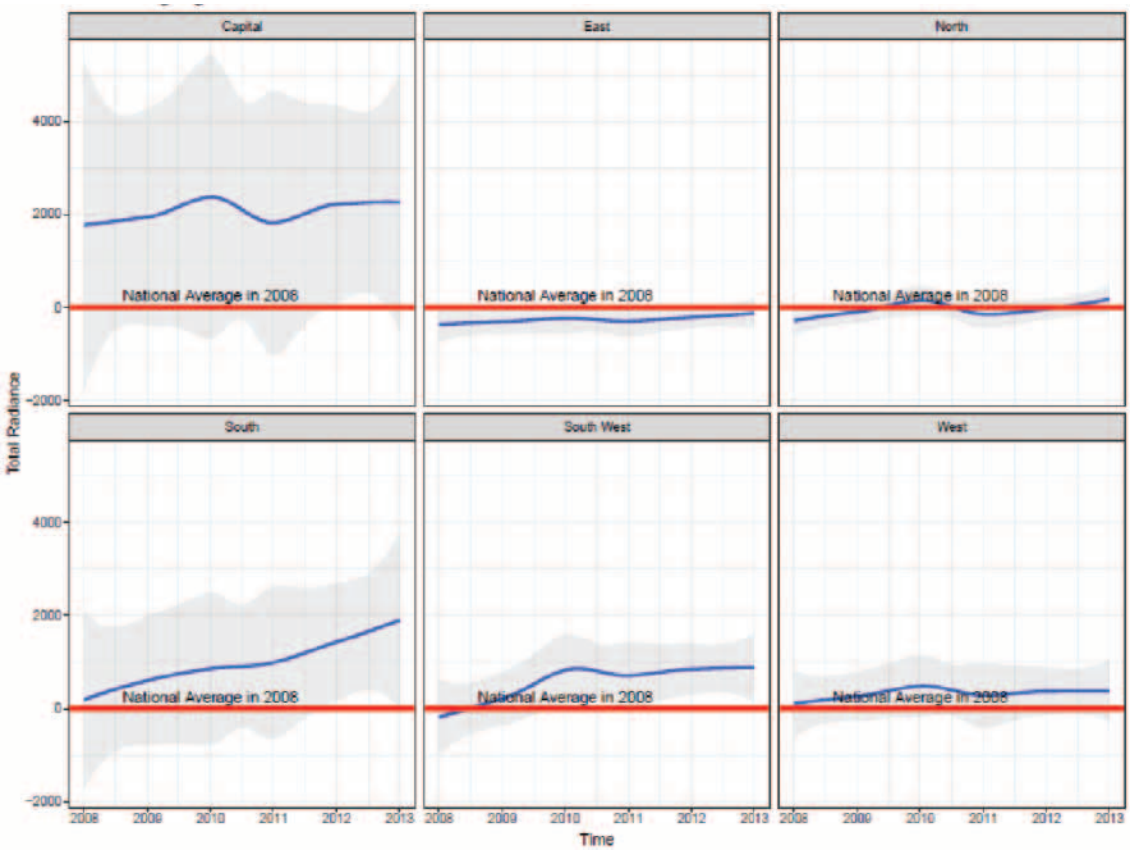
In fact, the economic trends in Afghanistan are largely positive. Figure 3.13 shows the changes in nightlights over the years by region. The general economic trend is moving upwards gradually in some regions and more rapidly in other areas (e.g., the South and South West) with relatively unchanged levels in the North and the East. This measure and associated trends differ from the GDP trends which show growth slowing in 2014 and 2015.[32] While GDP is a standard and important measure of production, economists have long worried about the GDP as a measure of economic activity in developing economies for several reasons. First, the GDP measure is heavily biased toward measuring manufacturing, which can be difficult in agricultural settings like Afghanistan.[33] Second, it excludes shadow and informal markets, which may be sizeable in Afghanistan. Third, it may not measure household wellbeing or activities well (see for example Easterlin, 2010). Our proxy for economic activity, nightlights differs from these measures

[32] For details on GDP and growth statistics see the World Bank bi-annual update on Afghan development. Joya, Mohammad Omar; Nassif, Claudia; Farahi, Mohammad Aman; Haque, Tobias. 2016. Afghanistan development update. Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/953921468196145402/Afghanistan-development-update

[33] For a simple discussion of these issues see for example http://www.economist.com/news/briefing/21697845-gross-domestic-product-gdp-increasingly-poor-measure-prosperity-it-not-even

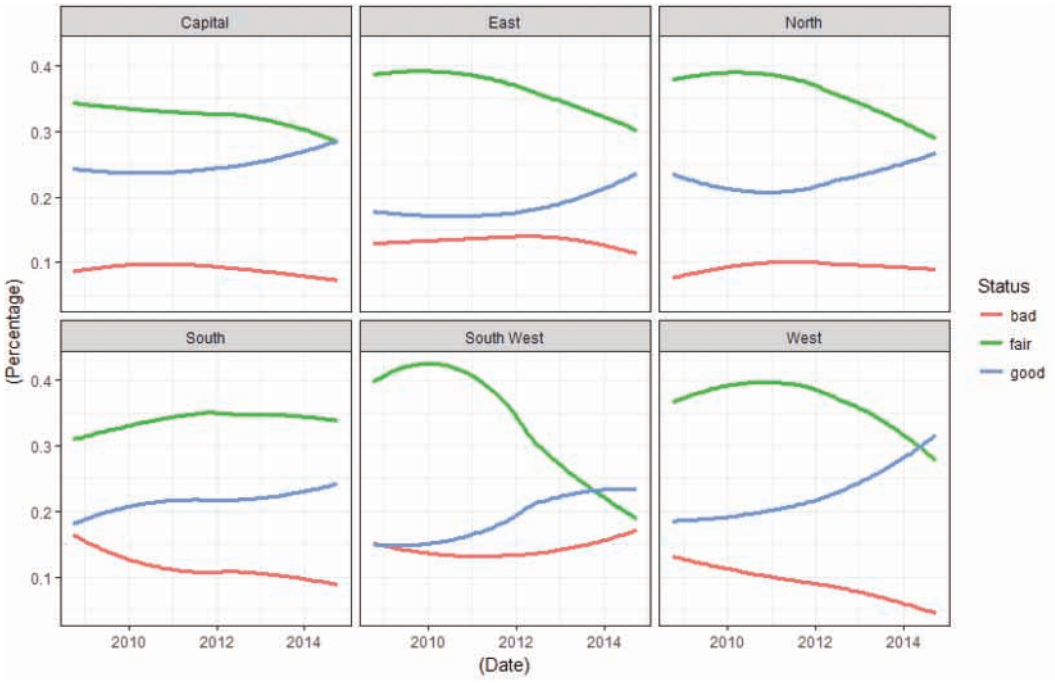
suggesting that while national level production as measured by GDP may not increase, household level economic activity may in fact be increasing. This difference could be explained by increased activity in informal markets or short-term jobs but nevertheless may represent actual increases in household income.

Figure 3.13: Changes in Nightlights over Time



We also observe some improvement in the perception of healthcare access, as shown in Figure 3.14. The perceived quality of healthcare is lowest in the South West region and highest in the Capital region, though these vary substantially over time. The Western region is the only region that appears to have experienced significant and sustained gains. We note here that we can only measure perceived healthcare access which is survey-based and highly subjective. Moreover, we have limited evidence on how changes in healthcare access may be associated with other health outcomes.

Figure 3.14: Regional Variation in Reported Quality of Healthcare



3.3 Time Trends and Regional Variation in Program Activity

In order to better understand the relationship between stabilization efforts and the broad trends observed above, we also reviewed the trends in programmatic activity for several key programs. These include: USAID activity recorded in Afghan Info, programmatic activity by OTI, the NSP operated by the World Bank, LGCD activity as well as the MISTI programs evaluation project, which included USAID projects such as SIKA and Kandahar Food Zone (KFZ).[34] As illustrated in Figure 3.15, OTI and LCGD were concentrated in the South region while NSP was implemented in many districts across Afghanistan.

[34] SIKA and KFZ were programs considered by the MISTI evaluation and though we do not have spending data for these programs we know in which districts these programs operated and consider them in our "project presence" analysis.

Figure 3.15: District Coverage by Program

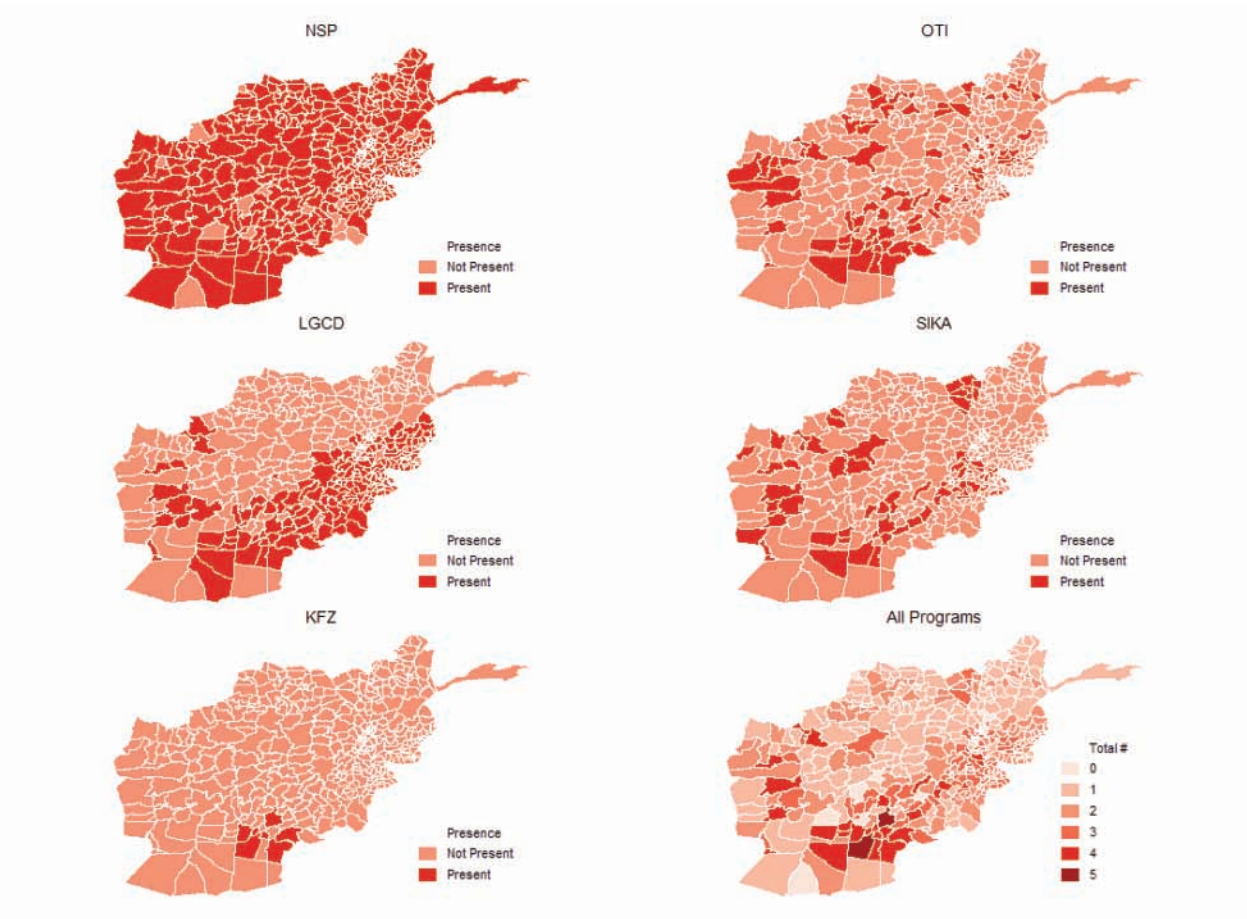


Table 3.3 presents the average annual spending and standard deviation across regions. We note substantial variation both temporally and regionally. OTI and LGCD occurred over shorter periods of time than NSP and spending appears to have been more targeted to particular regions (as evidenced by greater relative differences in disbursement amounts between regions). LGCD spent most, on average, in the East, South, and South West regions. In contrast, OTI spent most heavily in the South West and the Capital regions. NSP had a much larger average spending level and was more consistently spread across regions, with greater emphasis on the West and North regions and the lowest in the South West. The data presented in Afghan Info suggests that most of USAID's recorded stabilization efforts were focused in the South and South West.

Table 3.3: Average Spending by Region for Key Stabilization Programs

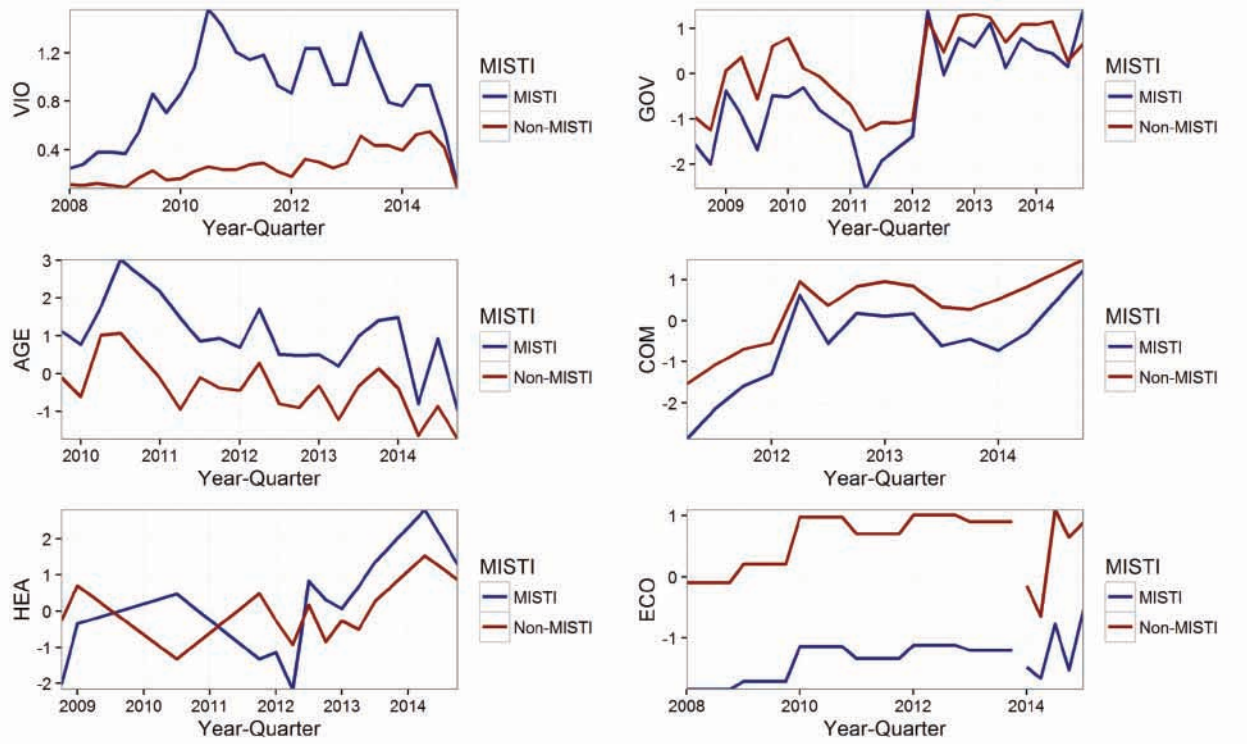
	East	West	North	South	South West	Capital
LGCD	11,390.94	1,819.609	0	5,512.145	4,624.99	1,790.804
	(79,805.36)	(4,128.87)	(0)	(58,583.84)	(30,032.34)	(25,763.97)
OTI	1,427.54	1,903.251	1,013.08	2,169.431	4,501.30	7,832.78
	(17,453.34)	(18,449.74)	(13,060.33)	(24,106.44)	(36,021.79)	(96,363.79)
NSP	57,131.43	92,840.5	71,932.68	52,617.48	47,205.11	58,656.23
	(159,684.8)	(298,413.9)	(181,609.9)	(161,296.7)	(169,606.3)	(147,439.1)
AI	236,117.1	192,562.9	193,676.2	501,682.9	627,204.4	394,2093
	(699,994.1)	(482,112.30)	(564,756.8)	(1,828,525)	(1,869,878)	(15,600,000)

Notes: Each cell contains the mean dollars spent in each region for that program. Standard Errors are in Parentheses.

While we did not conduct a robust analysis of the spending levels for the programs in the MISTI evaluation, we did explore some of the trends in outcomes to place our findings in a broader context. Figure 3.16 compares trends across each of our outcome measures by MISTI and non-MISTI districts. Here we define a 'MISTI district' as a district in which any of the programs considered in MISTI were implemented.[35] For all outcomes except health, MISTI districts were consistently different from non-MISTI districts. On average, districts considered in the MISTI report were more violent, less supportive of the government, more supportive of AGE, less cohesive, and had less economic activity. These differences were significant and substantial in magnitude—in many cases orders of magnitude larger than any program effects.

[35] These programs include CCI), SIKA-W, SIKA-E, SIKA-N, SIKA-S, KFZ, and CDP.

Figure 3.16: Trends by MISTI and Non-MISTI Districts



The observed differences in MISTI and non-MISTI districts have several implications. First, although the MISTI evaluation used a well-designed matching approach, the findings of differences cannot contextualize the size and scope of these effects in the broader context of Afghan stabilization programs. This is relevant because modest gains (or small negative effects) are not likely to substantially change the operating environment or population's well-being, especially in the long run. This is consistent with a broad review of the literature on estimated effects (Iyengar, Shapiro, and Hegarty, 2017).

Second, these districts were more supportive of AGE and had higher levels of violence, increasing the risk that any programs would be targeted or co-opted by insurgent groups. These activities may easily shift targeting from other areas that appear similar (and thus were in the matched sample), creating the appearance of greater levels of violence or increased AGE support in MISTI areas while in fact simply redistributing the total violence and support for AGE to be more concentrated in MISTI areas.

3.4 Summary of Key Trends

The objective of this chapter was to situate some of the findings from the research review and provide context for some of the later correlations by talking about broad trends in outcomes. An

extensive review of the research (see Iyengar, Shapiro and Hegarty, 2017) finds limited evidence that any individual program had substantial impact on key outcomes; in fact, the best evidence suggests that only small programs with limited scope appear to have any impact and these were modest in size and not sustained. We note explicitly here that these trends should be interpreted with caution. They *cannot and should not be* directly attributed to any individual program without a well-designed, causally identified study. What we present are correlations between program activity and these broader trends to highlight the extent to which programs taken as a whole may be temporally associated with overall improvement. Specifically, the broader trends do suggest:

- There has been overall improvement in Afghanistan with declining violence, increasing support for the Afghan government, and modest gains in health and economic outcomes

- Despite relatively high military presence in the South West and South, there were differences in trends in key outcomes, such as support for the Afghan government and health outcomes activity in which the trends have diverged and the South West appears to have trended better.

4. The Relationship between Stabilization Programs and Violence, Attitudes, and Well-being

This chapter discusses the key findings from comparing program activity and the key outcomes of interest (violence, support for Afghan government, support for AGE, community cohesion, health and economic well-being). The goal of this section is to estimate and discuss the relationship between these outcomes and stabilization programs to better understand the context within which these programs operated. However, none of the analyses presented are able to adequately address the multitude of interdependencies and complex relationships between program location, program spending levels, and key outcomes. For instance, conditions on the ground likely affect where programs are located, the ability of programs to disburse funds, and a host of other factors that would also affect the ability of the program to impact the key outcomes. As such, we present the relationships between broad trends.

4.1 Near-term Relationship between Program Activity and Key Outcomes

Our analysis includes a series of regression-based analyses to estimate the conditional correlation between program activity and key outcomes.[36] This analysis builds in complexity and we present the results from several key models defined as:

Raw: We begin our analysis by presenting the correlations between the level of spending (or program presence) and the outcomes of interest controlling only for the time period (i.e. quarter) during which the program operated. This analysis is critical to generating a baseline estimate of how the level of spending or program presence is associated with the outcomes of interest to compare with the subsequent more constrained models.

Fixed Effects (FE): There are a number of area- and time-specific factors that may generate a relationship between the outcomes of interest and project spending (or program presence). Therefore our next analysis estimates the relationship between project spending (or program presence) and the short-term outcomes, accounting for region (district)- and time (quarter)-specific effects that likely drive both spending and violence levels.

First Differences (Change): Several of the variables of interest, including violence, fluctuate significantly over time. Moreover, focusing on these changes may allow us to isolate the changes in outcomes in the presence of strong trends. For instance, while we may see that high levels of violence and project activity are correlated, we may also see that changes in violence and

[36] These specifications follow the Pre-Analysis Plan (PAP) approved by USAID in July 2017. In addition to the specifications discussed here, we also test the degree to which location specific effects, such as road density and ethnicity, are related to differences in outcomes. We do not find any significant differences when adding these controls and thus do not discuss these correlations in detail. The regression results specified in the PAP are presented in the Results Appendix.

changes in project activity are negatively correlated. This would suggest that while more projects exist in violent areas, increases in project activity are associated with reductions in violence.

Security Force Presence (ISAF presence): A key confounding factor in the relationship between violence and programmatic activity is security force levels (which we measure with ISAF military presence).[37] For instance, military presence could serve to reduce violence and make program activity more feasible or military could serve as a target and increase violence, regardless of program activities. We therefore want to estimate the relationship between USAID activities and the outcomes of interest, conditional on security force levels.

Security Force Synergy (ISAF Synergy): Security forces may not simply be a confounding factor in estimating the relationship between various outcomes and project spending or counts; they may amplify or undermine the effectiveness of projects in a given area as suggested by prior work in Afghanistan on CERP spending. We therefore estimate the relationship between project spending (or program presence) on key outcomes allowing for the direct relationship between security and the outcome as well as an interaction between ISAF presence and the program activity.

We estimate each of these specifications for the total spending and project existence variable from the Afghan Info Database, OTI, LGCD, CERP, and NSP data.

4.1.1 Relationship to Violence

As shown in Table 4.1, we find a predictable relationship between violence and program activity, based on how the programs selected districts in which to operate. For programs such as OTI or the stabilization programs in Afghan Info, we find a positive relationship. This is to be expected, given that many of these programs were explicitly targeted at violent areas. In some cases the level of violence in districts where these programs were implemented was 50 percent higher than the typical district in Afghanistan. The notable exception to this is relationship between NSP and violence. There is a consistent negative relationship between NSP and violence, as this program was explicitly directed away from violent areas.

The strong negative relationship found between NSP and violence is particularly instructive. The randomized control trial of NSP, in which NSP villages were compared to a random sample of non-NSP villages, found no effect on the likelihood of attack by AGE.[38] The observed

[37] We note that this measure does not include Afghan National Security Force (ANSF) presence. Currently we do not have reliable information on ANSF force levels. We are working to obtain that information at least at the province level and will include a separate control for ANSF or at least Afghan National Army force levels if feasible.

[38] For a detailed discussion of these findings see Beath, A., Fotini, C., and Enikolopov, R. Randomized Impact Evaluation of Afghanistan's National Solidarity Programme, July 2013.

negative relationship thus highlights the important role that site selection plays in mediating the relationship between violence and outcomes. In fact, the changes column illustrates that while districts with OTI or LGCD programs were on average more violent, this did not change significantly as the level of spending or degree of programmatic presence changed. Overall, it seems NSP illustrates two points: (1) if project activities are located in a community that has not experienced much violence then there may continue to be a negative relationship between program activity and violence due to selection; and (2) based on some combination of location selection and program design, NSP projects did not attract significant additional violence.

Table 4.1: Estimated Relationship between Violence and Program Activity

	Raw	FE	Changes	ISAF Presence	ISAF Synergy
Panel A: Spending					
Afghan Info	(+)	(+)	+	+	(+)
OTI	(+)	0	0	(-)	0
LGCD	0	0	0	0	0
NSP	(-)	(-)	(-)	(-)	(-)
Panel B: Existence					
Afghan Info	0	0	0	0	0
OTI	(+)	0	0	-	0
LGCD	(+)	(+)	(+)	0	0
NSP	(-)	0	0	0	0
CERP	(+)	(+)	0	(+)	(+)

Notes: Estimates based on column 1 of Tables B.1 through B.10 in the Results appendix. Estimates labeled with a "+" (or a "-") found a positive (or negative) relationship significant at or above 5 percent. Estimates labeled with parentheses around the sign –i.e. a "(+)" or "(-)"—were significant at or above the 1 percent level. Dark green and red cells indicate magnitudes in excess of 0.2 standard deviations. Lighter colors indicate a magnitude of greater than 0.1 standard deviations. Green denotes a positive change (an increase in violence) and red a negative change (a decrease in violence).

It is also noteworthy that the relationship changes, and in many cases goes to zero, when controlling for the presence of ISAF forces. This is consistent with the existing literature and with our interviews of program personnel. A number of current and former program staff noted that physical security was a constant concern and a tremendous obstacle to both program design

and implementation. The literature suggests that absent the security provided by military support (and sometimes even with it), the security situation inhibited the execution of even basic tasks needed for program operation (see for example Altai, 2012(a)).[39] This vital function was acknowledged even among those critical of the military's role in the stabilization context. However, the presence of military forces served as both a means of providing basic security and a target for AGE. Thus, consistent with the literature, it is not surprising that we find that the impact of aid on security is highly dependent on initial conditions (see Fishstein and Wilder, 2012; Sexton, 2015).

The fact that physical security itself is a key determinant to successful program implementation and sustainability complicates our ability to assess the impact of development aid on security as an outcome.[40] This is particularly salient when considering how integrated many potential insurgents are in the general community. Indeed, broader studies find that humanitarian assistance in conflict settings does not have uniform effects and the impact of violence on changes in civilian attitudes depends on whether the perpetrator is viewed as part of their in-group.[41] We address this issue in greater detail in Section 4.3.

[39] This refers to a report reviewing ASI programs, specifically Afghanistan Stabilization Initiative (ASI) Third Party M&E and Strategic Support: Sarkani District Report

[40] Derleth, J. W., & Alexander, J. S. (2011). Stability Operations: From Policy to Practice. PRISM Journal for the Center of Complex Operations Vol 2 (3), 125-136

[41] Lyall, J. (2016). Civilian Casualties and the Conditional Effects of Humanitarian Aid in Wartime.
Lyall, J., Blair, G., & Imai, K. (2013). Explaining support for combatants during wartime: A survey experiment in Afghanistan. American Political Science Review, 107(04), 679-705.

Table 4.2: Estimated Relationship between Support for the Afghan Government and Program Activity

	Raw	FE	Changes	ISAF presence	ISAF Synergy
Panel A: Spending					
Afghan Info	0	0	0	-	0
OTI	0	0	0	0	0
LGCD	0	0	0	0	0
NSP	0	0	0	0	0
Panel B: Existence					
Afghan Info	(+)	0	0	0	0
OTI	0	0	0	0	0
LGCD	0	+	+	0	0
NSP	+	0	0	0	0
CERP	0	0	0	0	0

Notes: Estimates based on column 2 of Tables B.1 through B.10 in the Results appendix. Estimates labeled with a "+" (or a "-") found a positive (or negative) relationship significant at or above 5 percent. Estimates labeled with parentheses around the sign –i.e. a "(+)" or "(-)"—were significant at or above the 1 percent level. Dark green and red cells indicate magnitudes in excess of 0.2 standard deviations. Lighter colors indicate a magnitude of greater than 0.1 standard deviations. Green denotes a positive change (an increase in support for government) and red a negative change (a decrease in for government).support

4.1.2 Relationship to Support for the Afghan Government or AGE

We find a reverse relationship when looking at reported support for the Afghan government, as shown in Table 4.2. There is a weakly positive relationship between most of the programs and support in the raw data. However, when addressing geographic variation and military presence, we find no consistent impact. In fact, we consistently find no relationship between programs and the changes data.

We find no consistent relationship with support for AGE, as shown in Table 4.3, though there is an overall tendency towards a negative correlation. For instance, for LGCD as spending

increased, support for AGE decreased. However, these results do not appear robust related to controls and may be artifacts of selecting relatively permissive areas, even within violent districts, to implement stabilization programs.

This limited relationship is again consistent with the existing evidence that finds that the degree to which the programs influenced attitudes was driven by activities outside of the program's control (see, for example, Altai, 2012(b) and Altai, 2012(c)).[42] In fact, a key factor in how programs related to attitudinal changes was the degree to which projects were implicated in government corruption (Carter, 2013; Fishstein, 2012). In the case of NSP, there is only weak evidence that it impacted attitudes or support for the Afghan government (Beath, Christia, and Enikolopov, 2013). During the interviews, respondents noted that corruption was pervasive, with considerable regional variation, and was a significant barrier to successful program implementation. Given these findings, it is unlikely that any individual, localized program could change perceptions of the Afghan government as a whole and more likely programs could only be implemented in places where corruption was relatively low and thus areas where Afghans may also already have a more positive association with the Afghan government. As such, specifications that control for the geographic variation or estimate changes over time find no significant relationship between programs and stated support.

[42] Altia 2012b refers to the Afghanistan Stabilization Initiative (ASI) Third Party M&E and Strategic Support: Barmal-Shkin District Report and Altia 2012c refers to Afghanistan Stabilization Initiative (ASI) Third Party M&E and Strategic Support: Khas Uruzgan District Report

Table 4.3: Estimated Relationship between Support for AGE and Program Activity

	Raw	FE	Changes	ISAF presence	ISAF Synergy
Panel A: Spending					
Afghan Info	0	0	0	0	+
OTI	0	(-)	0	-	(-)
LGCD	+	0	(-)	0	0
NSP	0	0	0	0	0
Panel B: Existence					
Afghan Info	(-)	0	0	0	0
OTI	0	-	0	0	0
LGCD	(+)	0	0	0	0
NSP	(-)	0	0	0	0
CERP	+	0	0	0	0

Notes: Estimates based on column 3 of Tables B.1 through B.10 in the Results appendix. Estimates labeled with a "+" (or a "-") found a positive (or negative) relationship significant at or above 5 percent. Estimates labeled with parentheses around the sign –i.e. a "(+)" or "(-)"—were significant at or above the 1 percent level. Dark green and red cells indicate magnitudes in excess of 0.2 standard deviations. Lighter colors indicate a magnitude of greater than 0.1 standard deviations. Green denotes a positive change (an increase in support for AGE) and red a negative change (a decrease in support for AGE).

4.1.3 Relationship to Community Cohesion

We find no significant relationship between any program and measures of community cohesion. In fact, the measures for community cohesion could not even be estimated for LGCD because LGCD was operated over a period of time for which community cohesion measures are not available. This lack of significant finding highlights the continued lack of evidence on the effect of stabilization programs on community cohesion and resilience (See Iyengar, Shapiro and Hegarty, 2017 for additional details).

Table 4.4: Estimated Relationship between Community Cohesion and Program Activity

	Raw	FE	Changes	ISAF presence	ISAF Synergy
Panel A: Spending					
Afghan Info	0	0	0	0	0
OTI	0	0	0	0	0
LGCD	NA	NA	NA	NA	NA
NSP	0	0	0	0	0
Panel B: Existence					
Afghan Info	+	0	0	0	0
OTI	0	0	0	0	0
LGCD	NA	NA	NA	NA	NA
NSP	+	0	0	0	0
CERP	(-)	(-)	(-)	NA[43]	NA

Notes: Estimates based on column 4 of Tables B.1 through B.10 in the Results appendix. Estimates labeled with a "+" (or a "-") found a positive (or negative) relationship significant at or above 5 percent. Estimates labeled with parentheses around the sign –i.e. a "(+)" or "(-)"—were significant at or above the 1 percent level. Dark green and red cells indicate magnitudes in excess of 0.2 standard deviations. Lighter colors indicate a magnitude of greater than 0.1 standard deviations. Green denotes a positive change (an increase in community cohesion) and red a negative change (a decrease in community cohesion).

4.1.4 Relationship to Health Access and Economic Outcomes

For health, we find a primarily positive relationship between reported availability of healthcare and program spending and presence. In particular, the relationship between LGCD activity and health is positive and significant across all specifications, including changes. The presence of CERP projects is also associated with an increase in the reported improvement in healthcare access in all specifications except changes. This may be due to the presence of CERP in particularly unstable areas, which may also result in lower perceptions of healthcare access.

[43] Military presence and CERP spending are perfectly collinear over the period for which we have Community Cohesion indicators (i.e. CERP only occurs where the military is present). Therefore, CERP coefficients are already conditional on military presence

Table 4.5: Estimated Relationship between Health Access and Program Activity

	Raw	FE	Changes	ISAF presence	ISAF Synergy
Panel A: Spending					
Afghan Info	(+)	0	-	0	0
OTI	0	0	0	0	0
LGCD	+	+	(+)	+	+
NSP	0	0	0	0	0
Panel B: Existence					
Afghan Info	0	0	0	-	0
OTI	0	0	0	0	0
LGCD	0	(+)	0	0	0
NSP	0	0	0	0	0
CERP	(+)	(+)	0	(+)	(+)

Notes: Estimates based on column 5 of Tables B.1 through B.10 in the Results appendix. Estimates labeled with a "+" (or a "-") found a positive (or negative) relationship significant at or above 5 percent. Estimates labeled with parentheses around the sign –i.e. a "(+)" or "(-)"—were significant at or above the 1 percent level. Dark green and red cells indicate magnitudes in excess of 0.2 standard deviations. Lighter colors indicate a magnitude of greater than 0.1 standard deviations. Green denotes a positive change (an increase in quality of healthcare) and red a negative change (a decrease in quality of healthcare). In no specification do we observe relationships of this magnitude

For economic activity, we observe relatively few significant associations, except that aid spending is negatively correlated with economic activity in the raw specification, consistent with targeting of aid dollars to less developed areas. Where significant correlations exist, outside of the raw specification, they appear to be positive. In one specification, OTI is positively associated with economic activity when interacted with military presence. This may be related to the feasibility of conducting OTI activities in the presence of the military which could then drive increased economic activity. CERP is positively associated in all but the raw specification which similarly could be related to the feasibility of executing program activity, which could then affect economic activity.

Table 4.6: Estimated Relationship between Economic and Program Activity

	Raw	FE	Changes	ISAF presence	ISAF Synergy
Panel A: Spending					
Afghan Info	(-)	0	0	0	0
OTI	(-)	0	0	0	+
LGCD	NA	NA	NA	NA	NA
NSP	0	0	0	0	0
Panel B: Existence					
Afghan Info	0	0	0	0	0
OTI	0	0	0	0	0
LGCD	0	(+)	0	0	0
NSP	NA	NA	NA	NA	NA
CERP	-	+	(+)	+	NA

Notes: Estimates based on column 6 of Tables B.1 through B.10 in the Results appendix. Estimates labeled with a "+" (or a "-") found a positive (or negative) relationship significant at or above 5 percent. Estimates labeled with parentheses around the sign –i.e. a "(+)" or "(-)"—were significant at or above the 1 percent level. Dark green and red cells indicate magnitudes in excess of 0.2 standard deviations. Lighter colors indicate a magnitude of greater than 0.1 standard deviations. Green denotes a positive change (an increase in economic activity) and red a negative change (a decrease in economic activity). In no specification do we observe relationships of this magnitude.

4.2 The role of Violence and Security Force Presence

Tables 4.1 to 4.6 present estimates from two specifications which include military presence. In many cases, the presence of international military forces does attract? violence. Across the board, the evidence suggests that military presence has a large, positive, and frequently significant relationship with the level of violence in an area. This is not surprising since the military was likely to be present in more violent areas. To understand what this might mean for the direct relationship between program spending and violence, we look at the size of the estimate in a specification where we control for violence compared to one where we do note control for violence. Controlling for military presence shrinks the size of the relationship between violence and program spending. ().[44] This suggests two things. First, stabilization programs operated in

[44] The full regression tables for this paragraph are presented in Appendix B, Table B.8.

violent districts where the military was present. Second, in some cases, the presence of the military mediated the effect of violence. This is observed among the programs included in the Afghan info data. This effect was much smaller, and sometimes not significant, for other programs—especially OTI and LGCD. ,

This is consistent with both the research literature and the programmatic staff interviews. At a purely tactical level, much of the literature recognizes the importance of military presence during program execution to assist in providing the basic level of security needed for program execution (DoD JCOA, 2006; Felbab-Brown, 2012; ICG_2011; Kapstein, Kathuria, 2012; Sexton, 2015; Taylor, 2010). Absent this support (and sometimes even with it), the security situation inhibited even basic tasks needed for program operation (see for example Altai_2012_5). In the context of programmatic implementation, though, the presence of the military and even basic security was not a magic bullet. A prominent theme across all interviews is an acknowledgement of the tremendous difficulty of implementing programs in an environment such as Afghanistan. One respondent described the dilemma: "We were being asked to successfully implement stabilization projects in the most corrupt country in the world in the middle of a war." Given the obstacles facing the implementation of development aid in Afghanistan, most respondents felt the goals and expectations for stabilization programming efforts were unrealistic, especially regarding long-term sustainability. One respondent noted it was as if they were expected to "magically" stabilize deeply insecure areas suffering from a complex array of problems. This problem was repeatedly noted in the literature; for instance Taylor (2010) highlighted that "security is still the major issue inhibiting project implementation in stabilization contexts. The take-away from this analysis is that the presence of the military is an indicator of whether programs can be conducted in an area or not but does not appear to have a significant effect either amplifying or inhibiting the program's effect directly.

4.3 Sustainability of Program Effects

A key aspect of the research design was the time horizon over which any effects were apparent. In many cases, programs focused on generating rapid effects within a very short window (3 to 6 months). Other programs were focused on medium-term (6 to 18 months) or long-term (18+ months). Some stabilization programs also considered the impact on key indicators over an even longer timeframe (3 to 5 years) of implementation. However, our data was a significant limiting factor, as is the case in many other evaluations. In many cases, programs are not implemented in ways that would enable measurement of effects at different time scales. Doing so requires generating consistent measurement in treatment and control units from baseline through multiple periods of measurement, including endline measurement years after implementation is complete. Such measurements were not budgeted for as far as we could tell from the existing studies, nor did we find studies that looked back on the effects of major efforts after a substantial gap of time. Indeed, one frequently expressed concern about the "securitization of aid" underlying

stabilization programming in Afghanistan was the shift in focus away from longer-run "development", toward short-term, quick impact projects.

Moreover, the dynamic nature of conflict environments makes it hard to find any effects from programming after more than a year. Many other factors may impact the outcomes of interest, and the more time that passes, the more these variables interfere with establishing the cause of each effect. Thus, after six months, it is impossible to establish a causal link between programming and outcomes.[45] As a result, many programs focused primarily on outputs, rather than outcomes. In these circumstances, one of the few means by which to track success was maintaining a high "burn rate" – allocating vast sums of money over a short interval. Though well intentioned (trying to do as much good as rapidly as possible), this approach made monitoring outcomes impractical at best, and impossible at worst.

Table 4.7: Estimated Longer Run Relationship between Program Activity and Outcomes of Interest

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Spending						
Afghan Info	(+)	0	0	0	0	0
OTI	0	+	0	0	0	0
LGCD	+	0	0	0	0	NA
NSP	-	0	(+)	0	0	0
Panel B: Presence						
Afghan Info	0	0	(-)	+	0	0
OTI	0	+	-	0	0	0
LGCD	(+)	+	-	0	0	NA
NSP	0	0	0	0	0	0
CERP	(+)	0	+	0	(+)	0

Notes: Estimates based on column Table B.11 in the Results appendix. Estimates labeled with a "+" (or a "-") found a positive (or negative) relationship significant at or above 5 percent. Estimates labeled with parentheses around the sign –i.e. a "(+)" or "(-)"—were significant at or above the 1 percent level. Dark green and red cells indicate magnitudes in excess of 0.2 standard deviations. Lighter colors indicate a magnitude of greater than 0.1

[45] In impact evaluations in conflict zones, this is a general problem for most any outcome other than basic demographics among geographically stable populations.

standard deviations. Green denotes a positive change (an increase in economic activity) and red a negative change (a decrease in economic activity). In no specification do we observe relationships of this magnitude.

Nevertheless, we wanted to see whether the data show persistent program effects on outcomes, so we looked two years (eight quarters) past the conclusion of the programs. The results are reported in Table 4.7. Not surprisingly, given that programs were intentionally located in "tough" areas, there remains a positive correlation between program activity and violence. This confirms the general finding that violence is geographically concentrated and persistent. We do find consistent evidence of positive effects on support for the Afghan government for both OTI programs and NSP. This could be due to the fact that, in evaluations for NSP and OTI's CCI program, the government was credited with successfully running the programs. This is particularly notable in the case of CCI. While NSP was an Afghan Government run program, CCI which was an "Afghan First" initiative, which focused on specifically engaging Afghan-owned businesses.[46] In our interviews, respondents noted that programs were likely to be more successful—and those successes were likely to be sustained—in areas where the population perceived their local government representatives to be more responsive and less corrupt. When programs were well-administered and the local population attributed the successes to the government, it appears that small, but statistically detectable effects may be observed even over a longer timeframe.

4.4 Summary of Key Findings

Overall we find some evidence of modest gains associated with stabilization programming, but limited evidence that these changes could be sustained. In particular, we found that:

- Many of the programs were present and spent money in tough places—that is places that were more violent and with greater anti-government sentiment.

- Despite being located in these tough locations, we found some evidence of gains associated with stabilization programming, particular with LGCD and health related outcomes. These LGCD estimates are quite robust.

- In the long run, there is little evidence of persistent effects that can be detected with the currently available data. Violence appears to be persistent despite program activities. There is some evidence that OTI programs are associated with longer-run gains in improved attitude towards the Afghan government and reduced support for anti-government elements.

[46] For a more detailed discussion of the Afghan First initiative, see for example Rhyne (2011) http://www.dtic.mil/dtic/tr/fulltext/u2/a543600.pdf

5. Conclusion

The overall findings of this report can be summarized as follows:

- In general, all indicators are improving over time—albeit modestly and with substantial regional variation; violence is persistent and very few programs will have significant impact on it;
- In the near-term, many programs are positively associated with violence, likely related to their placement in violent, unstable areas.
- Some of the smaller programs like OTI's CCI and CERP seem to be correlated with increases in support and LGCD is robustly associated with improvement in perceived access to healthcare. (R1)
- In the longer term, we see more limited evidence of effects of sustainable effects, with some evidence of sustained improvement in support for the government from OTI programs. Violence appears to be persistent in areas that were violent prior to stabilization program activity. (R2)
- The military appears to play a key role in facilitating program activity but there is no evidence that it serves as an amplifying or inhibiting factor for program success. (R3)
- There is some evidence of synergistic effects but we caution that our analysis only looks at a small number of programs (some programs in AI, NSP, MISTI, OTI) and thus may not be generalized to overall synergy between donors (R4).

These findings are consistent with a broad, systematic compilation of evidence from the existing literature and a series of interviews with stabilization-related program staff. One key takeaway from our analysis is just how critical initial program design—including measurement and evaluation—is to allowing specific program evaluation and broad panel analyses like those in this study. Although we deliberately chose to not define "stabilization" in this study due to the range of definitions and goals used among the programs we considered, establishing a definition would be beneficial when undertaking future program design both for facilitating evaluation and for helping those on the ground implementing the program activities. For example, in the interviews conducted by the research team, we noted that most participants found it difficult to answer the question, "How would you define success for stabilization programs generally?" They were able to enumerate specific project successes (e.g., improved delivery of seed, increased participation in local councils) but the lack of a specific and consistent definition of "stabilization" made it difficult for many implementers to collect the metrics necessary to evaluate the broader impact of their activities.

Our analysis also addresses a related question—"What types of programs worked where?"—by trying to determine how a mix of different program activities and external factors may be related to aggregate trends. In such complex and difficult operating environments, the well-designed, internally valid causal estimate presented by MISTI provides only limited insight into

relationships between such programs and the outcomes of interest. Comparing programmatic activity and violence or measures of support, while holding a number of conflating factors constant, can provide a complementary set of information to supplement the MISTI evaluation or similar locally-focused studies. Moreover, in-depth qualitative studies that capture the experiences of program officers and implementers in greater detail could help shed light on specific practices that may be effective under different conditions.

The analytical effort also produced four broader lessons:

- First, while none of the gains are large, and while we cannot establish causal relationships with the same credibility of impact evaluations conducted outside of conflict zones, increased spending on stabilization programming did lead to sustained gains in health and support for the government. These measurable gains are especially notable given that stabilization programming targeted more insecure areas, and suggest that the effort resulted in a small overall improvement in outcomes compared to a scenario where no stabilization programs were run. These modest gains show that the outlook for stabilization aid is not hopeless and must be more rigorously evaluated to enable more effective program designs in the future.
- Second, implementation data are not being recorded with sufficient detail to enable retrospective learning and adaptive management. Pulling the information together for this project required substantial coordination between USAID, USIP, the World Bank, and an academic institution. Despite that effort, data on many programs simply could not be feasibly shared because of the difficulty in combing through a wide range of information collected. The requirement for better, standardized, centralized record keeping needs to be built in to future contracts from day one if the USG hopes to learn from and improve upon its stabilization efforts. These data should be as comprehensive as possible; to include what was done where and when, but also the rationale for decision-making (such as why certain sites were chosen over others).
- Third, structuring evaluations appropriately in conflict zones requires long-term thinking and coordination. The MISTI evaluation was an unprecedented effort to measure stabilization impacts in one place, but it was not set up to learn about key design elements for stabilization programming and was not part of a family of similar efforts that could have probed whether what worked (or did not work) in southern Afghanistan worked (or did not work) elsewhere. In the future, processes to evaluate program design should be an explicitly identified goal within the broader scope of planned evaluations.
- Fourth, there is tremendous potential in using remote sensing data to track outcomes. Modern open-source tools for working with geo-spatial data can be applied to remote sensing data available from the National Oceanic and Atmospheric Administration (NOAA), as well as commercial providers to measure economic conditions and population welfare in even the toughest areas. Doing so requires sensitivity to the quirks

of data collection and to cross-regional differences in the relationship between on-the-ground conditions and what can be seen from space, but the potential exists to measure changes at fine geo-temporal scales in any location on earth. That opens up tremendous opportunities for learning and policy feedback provided that detailed programmatic data is maintained. This technology should be leveraged to provide the most complete information possible to integrate into future evaluations.

The lessons our analysis brought forth will not make future stabilization efforts more effective in and of themselves, but setting future programs up to collect higher quality data, coordinate their evaluation efforts, plan ahead by building in reporting requirements, and make effective use of new technologies may allow programming to be more adaptive to the conditions on the ground. As these best practices are incorporated, policy makers and implementers may be able to learn more about what makes programming effective or ineffective and establish more definitive causal links between programming and outcomes. This can enable better design and implementation of stabilization projects in Afghanistan and other conflict-affected areas in the future.

Appendix A - Data Appendix

A.1 Violence data

Data Processing
As our primary violence dataset we use a unified version of Significant Actions (SIGACTs). This dataset includes approximately 500,000 georeferenced incidents. For each action, we have precise latitude and longitude, time and date, instigator and target, and a brief description. The violence measures that we construct from these data are used throughout our analyses both as a primary outcome variable and as a potential confounder.

Cleaning Process
For each of our violence data sets, each attack is coded as one observation. For use in our analysis, we first run some preliminary data cleaning and aggregate into a usable unit of analysis (cumulative weekly/monthly/quarterly incidents).

1. Identify data entry errors (e.g., the same attack entered multiple times) and drop these observations.

2. Aggregate the cumulative number of events by district over a certain period of time and visually inspect resulting data for extreme outliers that are likely the result of data entry error.

3. Depending on the structure of the data, we normalize the cumulative number of events by logarithmically transforming the number of incidents or the per-capita number of incidents.

4. Compare violence trends across data sources to cross verify data.

Summary Statistics

Table A.1: SIGACTs Summary Statistics

	Mean	Std. Dev.	Min	Median	Max	Observations
Year	2011.03	2.4	2002	2011	2015	483,900
Month	6.69	3.26	1	7	12	483,900
Monthly SIGACTs	9.32	30.86	0	1	1,093	51,912
Monthly SIGACTs PC	1.47	3.77	0	0.23	131.1	51,912
Monthly SIGACTs (Enemy action)	2.84	13.13	0	0	762	51,912
Monthly SIGACTs (Enemy action PC)	0.48	1.65	0	0	71.32	51,912

Note: Monthly SIGACTs PC represents the number of incidence per 10,000 people in a district. The number of Observations per year and month are higher for year and month as these are taken from a dataset in which one action is one observation. The latter statistics are drawn from a dataset of the incidents aggregated to the district-month level.

SIGACTs data for this study covers the period between 2002 and 2015. As can be inferred from the disparity between the mean and median values, the distribution of incidents is highly skewed, with many districts seeing very little violence while a few others experienced much higher levels of conflict. However, these simple averages tell us very little about the nature of the conflict over time and geography.

In this dataset, we observe the greatest number of violent incidents in the year 2010, with the number of enemy actions being nearly double that of the preceding year and more than double that of the following. Of the types of conflict, enemy action was the most frequent, with over 147,000 events logged between 2004 and 2015. The next most frequent type of incident was explosive hazard, with more than 108,000 incidents, followed by friendly actions with approximately 68,000 incidents.

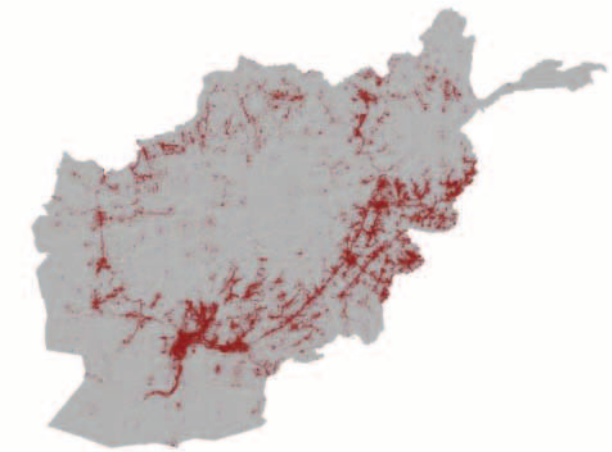
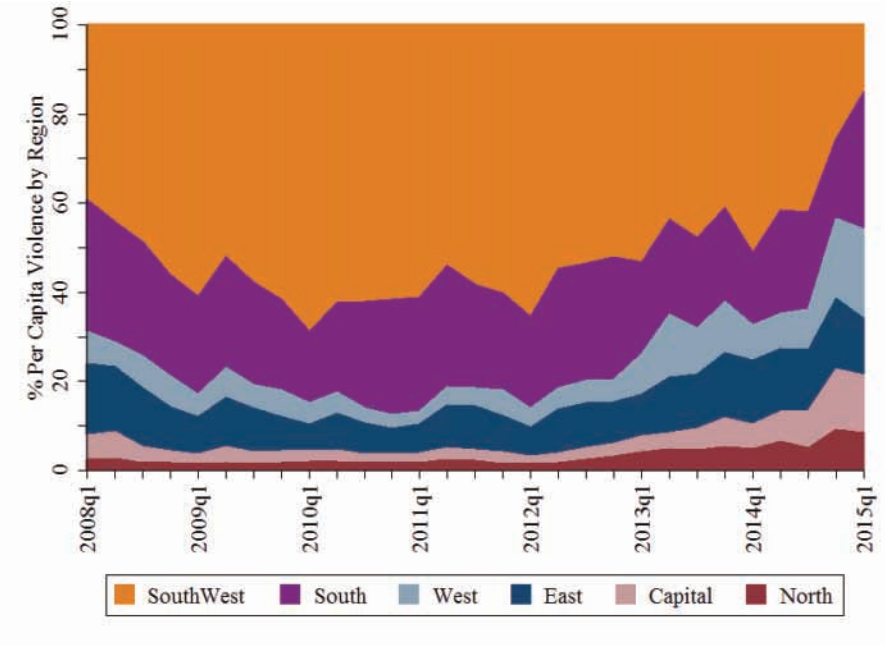


Figure A.1 highlights the geographical distribution of these incidents, where each point in the figure represents an

incident. This gives us some idea of which districts had the highest levels of incidents throughout the conflict. In general, the points data suggest that the violent events are highly targeted–concentrating on specific road networks and cities–while the distribution across districts demonstrates that much of the conflict took place in the South of Afghanistan, in Helmand province in particular.

Figure A.2: Per-Capitized Share of Violence



Over time, we see a general increase in levels of violence. However, we see a marked drop towards the end of the period covered. Looking at the per capita measure, the rank ordering of regions in terms of violence remains stable, i.e. the South West remains the most violent area while the North West region is the least.

We have also compiled a related indicator from the ANQAR survey data. We aggregate district-quarter responses from the survey item "How is the security situation in your Mantaqa?" When compared to the violence indicator generated from SIGACTs, we find that an increase in the level of violence is negatively correlated with self-reported measures of security.

Issues with Data

There are a number of issues related to the collection of violent events data in Afghanistan, the first of which is that the nature of collection is related to the focus of the collecting agency. In this case, ISAF will likely emphasize violent events related to insurgency rather than overall criminal violence. However, it may also be that the emphasis changes (slightly) over time as ISAF's strategic priorities change from COIN operations to rebuilding campaigns. An additional

issue is that the capacity to collect this data is related to presence; as ISAF presence in an area increases, it is better able to identify and record instances of violence. Thus, there may be some selection bias, though we will attempt to control for presence by including measures of troop levels and military spending.

Within the SIGACTs data, we found several data entry errors including several empty observations –with a unique identifier, but values of 0 for numeric variables and empty for string variables. We also found several repeated entry errors.

A.2 Attitudinal Measures

Support for Government Indices

A number of surveys that we have compiled include measures of support for government. However, our preliminary analysis focuses on the Afghanistan Nationwide Quarterly Assessment Research (ANQAR) survey as our primary source for support for government as it covers the broadest period and includes the most potential questions related to an individual's support for government. The purpose of these indices is to generate a proxy for popular support for the Afghan government and to track how those attitudes change over time, across districts, and in response to the level of USAID spending.

Cleaning Process

Our support for government indicator is measured exclusively through survey data. Our general work-flow for these indicators is presented below.

1. Identify relevant questions within each survey and within each round for each indicator.
2. Prioritizing these and other demographic characteristics, clean survey questions and responses.
3. Generate an additive index of all relevant responses within each bin, denoted G^{add}_{jt} ($\forall$ q in a bin, $q_i = 1$ if supportive (i.e. signals support for government), else, $q_i = 0$. Then, G^{add}_{jt} is equal to $\frac{\sum_{i=1}^{n} q_i}{N_{qi}}$) where N_{qi} is the number of questions. Note: this step entails additional cleaning of each question.
4. Perform multiple imputation, prioritizing questions for which 1) we have a strong predictive capacity, and 2) that require as little imputation as possible.
5. Generate first principal components analysis (PCA) index, G^{pca}_{jt}.

For ANQAR we have a total of 51 questions that we identified as potentially indicating an individual's support for the Government of the Islamic Republic of Afghanistan (GIRoA). Fifteen of these questions were asked over a sufficient time period and were strongly correlated.

On average these indicators are highly correlated. However, we identify two indicators[47] which appear relatively uncorrelated, or negatively correlated with the other indicators. This suggests that these questions are relatively uninformative vis-a`-vis support for government and will not be included in our indices.

Generating the first principal components index of support for government was done as follows:

Step 1: We use simple chained regression imputation to fill in missing observations within survey waves, taking the mean of five rounds of imputation.

Step 2: Following the same methodology, we then also impute *across* survey waves to fill in predicted responses for those questions that were not asked in a wave as a function of the 15 that were asked.

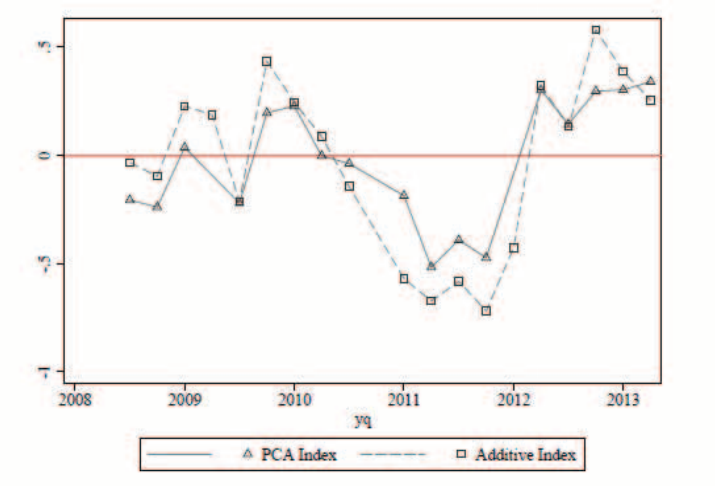
Step 3: We then perform a principal components analysis of these questions to identify the underlying covariance shared between the survey items.

Step 4: We use these factor loadings to generate individual level, continuous scores of support for government, which are then aggregated at the district-quarter level.

The first principal component in the case of support for Government accounts for approximately 37% of the covariance in each of the questions. In theory, PCA is extracting the underlying component of support for government shared between survey questions and leaving behind the extraneous information which varies across questions. Thus, the PCA index should be a purer measure of support for government compared to the additive index. Additionally, it is able to capture support for government in more detail (finer and more continuous). However, the PCA measure is very highly correlated with the additive index at the district-quarter level. As can be seen in Table 4.7, the correlation between the two indices is nearly 90%. Figure 4.12 tracks the country average level of support for government as measured by both indices over time. Slopes are increasing and decreasing over the same period, with only minor discrepancies between the two.

[47] qnew110"do you think the national army will be able to defeat the opposing government...", and qnew058"Between the two opposing anti-government elements and the government, who has mo..."

Figure A.3: PCA Versus Additive Index over Waves 1-20



Issues with Data

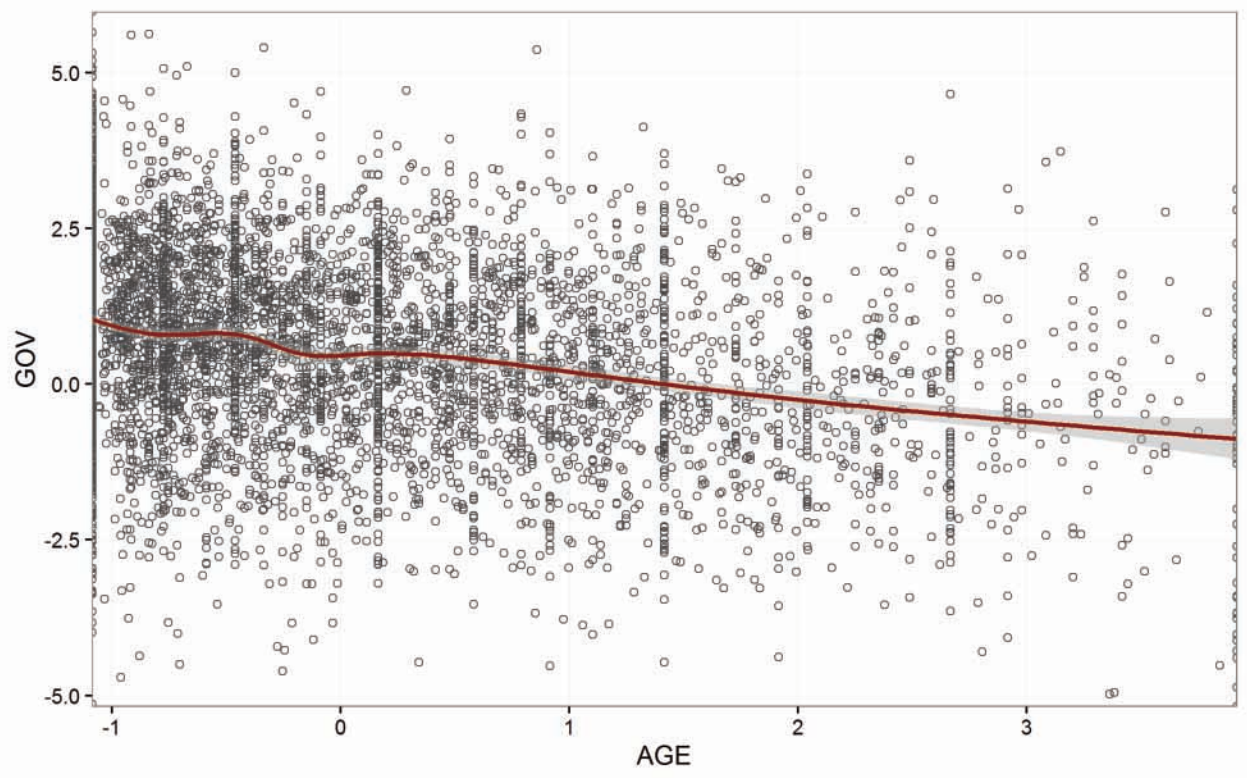
Based on our review of the data, we note several caveats. 1) Across each of the 26 survey rounds that we currently have in our possession questions are numbered sequentially and not in regard to its position in a previous survey. Thus, there is no identifier for each question and questions had to be matched across rounds manually, introducing some potential for error. 2) Small changes in wording across rounds can occur, though this may have been the result of differences in translation and not differences in phraseology in the actual survey. 3) The potential responses for similar questions also change across waves. For instance, in some rounds Likert scales held three values and in others seven. Additionally, across waves these scales could range from neutral to increasingly positive, or from negative to positive (with neutral as a midpoint) for the same question. In each of these cases and across 650+ unique questions, responses were amended to the lowest common denominator to ensure maximum coverage. For instance, if the potential responses for a question in one round were "supportive," "neutral," and "not supportive" and "very supportive," "supportive," "neutral," "not supportive," and "very not supportive," the responses in the latter set would have been simplified. So any responses for "very supportive" would become "supportive" as, logically, someone expressing strong support for something can be inferred to be generally supportive of the same. 4) Coding methodologies for "Do Not Know" and "Refused" responses also varied across survey rounds. These were standardized and given a common value throughout the survey rounds.

Support for AGE

There are fewer questions asked about support for AGE and only one of these is consistently present throughout the 26 rounds. This question asks, "in your opinion, if the Taliban were to

return to power and govern Afghanistan, would this be a good thing or a bad thing for the country?" Responses to this question ("good" or "bad") were averaged at the district-quarter level to get a measure of support for Anti-Government Elements. In line with intuition, support for government and support for anti-government are significantly and negatively correlated.

Figure A.4: District-Quarter Relationship between Support for GOV and AGE



At the regional level, support for AGE is consistent over time, though decreases modestly in most cases. The exception to this is the South West, where support for AGE started relatively high and dramatically decreased from 2010 to 2015.

Figure A.5: Regional Variation in Support for AGE

Community Cohesion

Potential questions for Community Cohesion were less frequently asked in the ANQAR surveys. For this reason, our community cohesion indicator only begins in the second quarter of 2011. Community Cohesion and resilience is a fairly amorphous concept, but based on a reading of the literature and interviews with stabilization programming implementers, the goal of community cohesion would be 1) to improve attitudes towards and performance of local governance, 2) to provide mechanisms for local dispute resolution, and 3) to maintain security in the community. We develop an indicator for each of these three components of community cohesion. First, we generate a PCA index of support for local government (using several measures and the same method to generate the Support for Government Indicator). Second, we take responses from questions asking "If you had a dispute, who would you take it to?" Responses indicating a willingness to take the dispute to either a local Shura/Jirga or a state court at the individual level demonstrate the availability of local conflict resolution mechanisms. Third, our measure of community resilience was generated from a survey item asking whether or not respondents believed that the government would be capable of

maintaining security in the future. These three sub-indicators were then aggregated using the same principal component method that we employ to generate our measure of support for government. The outcome of this PCA we use as our primary Community Cohesion indicator at the individual level. Each of the three sub-indicators is significantly and positively correlated. As with our other attitudinal measures, these measures were aggregated at the district-quarter level for analysis.

Figure A.6: Community Cohesion Sub-Indicators over Time

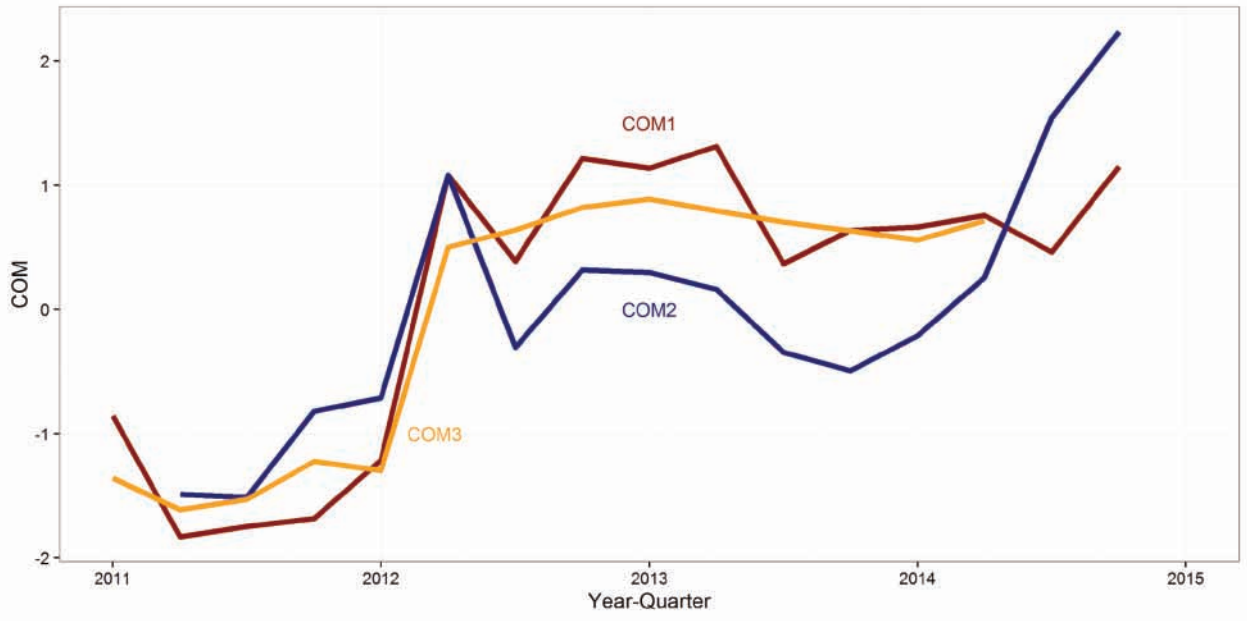


Figure A.6 displays the national average for each of these indicators over time. Despite the limited time period for this variable, we observe a distinct upward trend across each of the sub-indicators. COM1 is the support for local government measure, COM2 indicates how capable the local authority will be in maintaining security after ISAF leaves, and COM3 indicates willingness to use a state court or local Jirga to resolve disputes.

A.3 Economic Activity Data

Economic Activity Measured Using Satellite Images

The night illumination landscapes (hereon referred to as nightlights) can be shown to proxy for local economic activity, urbanization levels, and population density. The nightlights data plays a useful role in policy evaluations and understanding local economic scenarios, especially in places such as Afghanistan, which has few reliable measures of economic activity. Thus there is very little chance of confounding the nightlight variable in proxying economic activity due to such factors. We develop a nightlights indicator based on data from two sources of satellite imagery that are not directly comparable.

First, from 2008 to 2013 the DMSP collected visible and thermal infrared data (night and day) to form yearly composite images. The composites contain lights from sites with persistent lightings. Ephemeral events such as short term fires or flares are discarded. The background noise is replaced with 0 and data values of radiance range from 1 to 63. Second, in 2014, the DMSP satellite was replaced by a more sophisticated VIIRS satellite (Visible Infrared Imaging Radiometer Suite) which reports actual radiance values in nanoWatts/cm2/sr. The composite images thus comprise night time light in an area filtered to exclude lightning, lunar illumination, and cloud cover. Both the DMSP and VIIRS geo-referenced files are forced onto the WGS84 projection. However, the images are not directly comparable, generating a break in the time series but still permitting regional, within year comparison.

Overview of Data

From 2008 to 2013 the DMSP collected visible and thermal infrared data (night and day) to form yearly composite images. The composites contain lights from sites with persistent lightings. Ephemeral events such as short term fires or flares are discarded. The background noise is replaced with 0 and data values of radiance range from 1-63.

Data Processing

The GeoTiff image files[48] are downloaded from NOAA (National Oceanic and Atmospheric Administration) for all the available time periods and a coordinate based geo-referencing system is applied. The spatial reference defines how geographic data is mathematically mapped onto a flat map with the least amount of distortion.[49] Once the data is transformed using this projection it can be combined, intersected, interacted with any other map (i.e. the district shapefiles, population land-scans etc.) having those projections, which makes

[48] Raster image data comprises of a rectangular grid of pixels and each pixel is a data point defined by latitude/ longitude and illumination level.

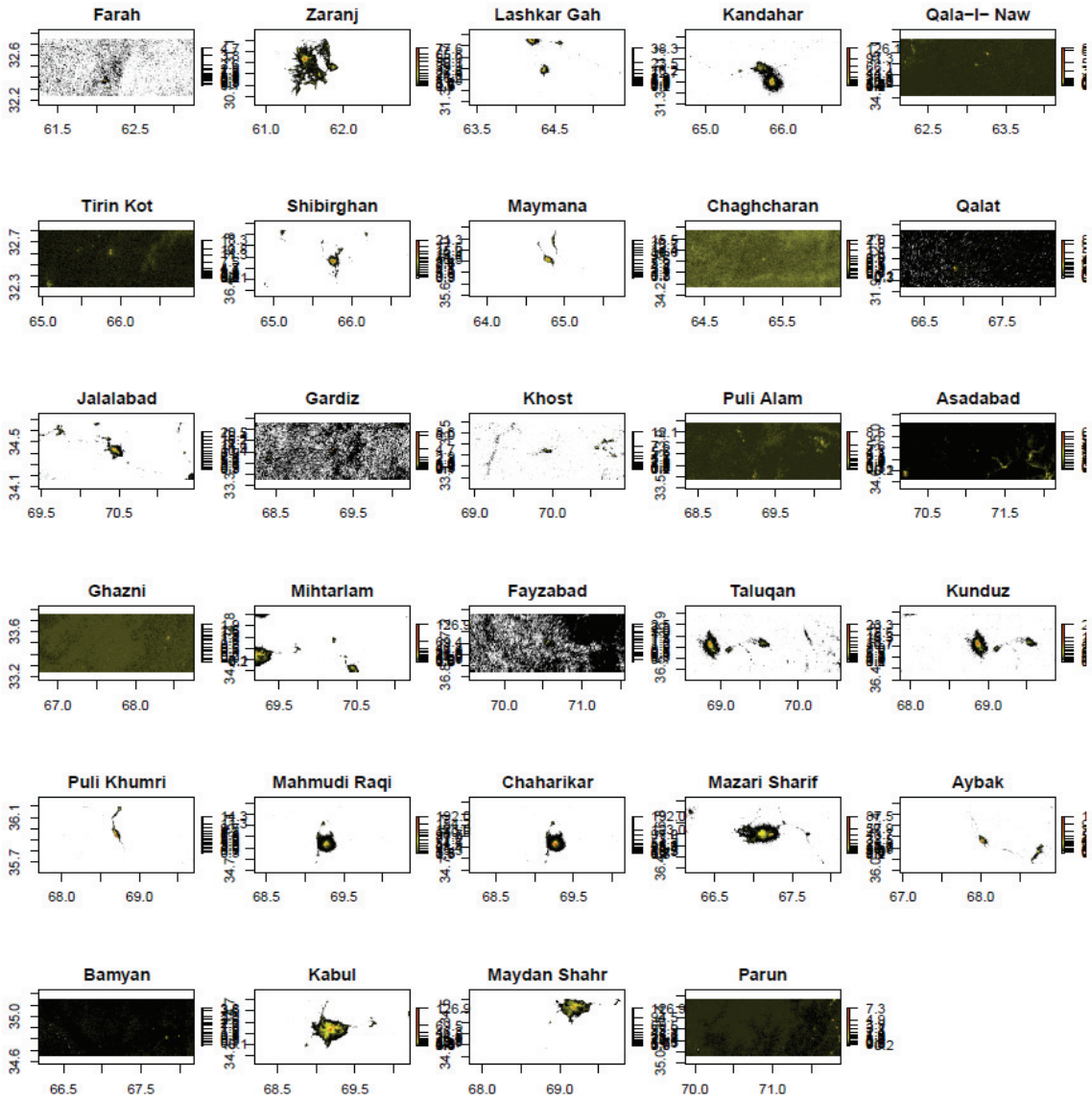
[49] Here we use WGS84 spatial projection.

borrowing data from other maps possible.

Figure A.7 (below) compares district capitals across the provinces. At first we find the center of the city using Google Maps. From the city center we draw a rectangular boundary (by limiting latitude and longitude values). Then different levels of nightlight illumination are clustered together to visualize the spatial distribution across the sites. It is important to note the different scales on the Y axis which gives a fair idea of the maximum and minimum levels of illumination in the city center.

For example in Kandahar we observe the full range of nightlights spread across the city. In comparison, Bamyan only has a maximum value one-tenth of Herat across a small spot.

Figure A.7: Regional Variation in Nightlight Measure



To compare values across the years we need to move beyond visual comparisons and need to unify the raster data into matrices and vectors. The VIIRS files give absolute radiance values in Nano Watts/cm2/sr multiplied by 1E9. The DMSP values are rescaled to fit the range 163. Although comparison cannot be done across the two measuring methods, comparison within methods is possible. (Side note- DMSP captures nightlights from 2008 to 2013, VIIRS composites are now available from January 2014 onwards for every month). The vector form of the data is cleaned and preprocessed to - have a row identifier (ObjectID, ProvinceID, or DistrictID) the exact location in latitude and longitude and the average radiance at this particular location.

Time-Series of Nightlight Illumination Across Provinces

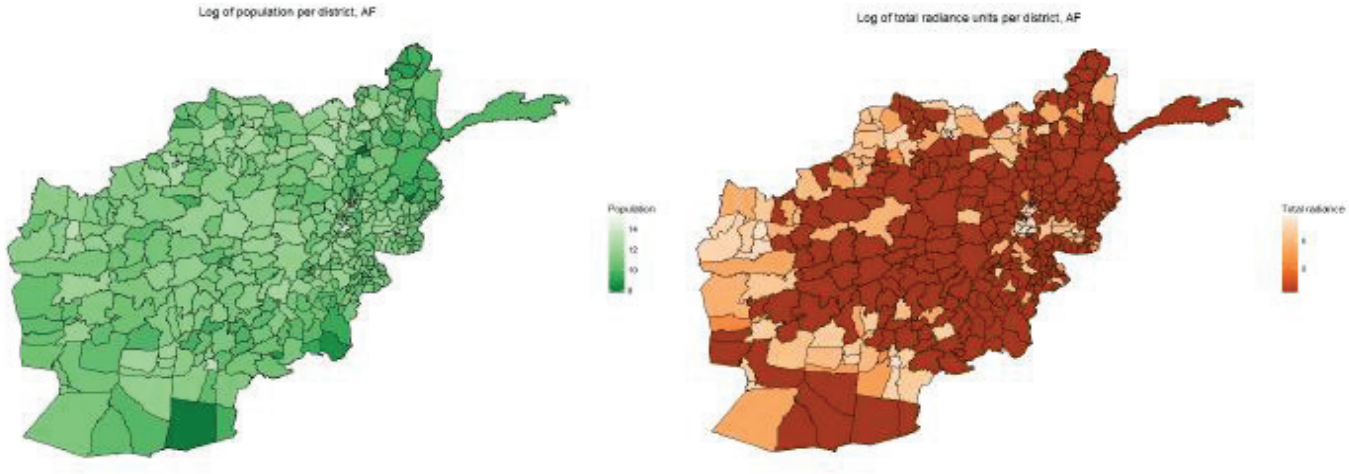
It should be noted that in Graph A.7 all the cities have different scales to emphasize the rise and fall in radiance levels. Our aim is to analyze if the rise and fall of difference radiance levels can be associated with increases / decreases in international stabilization efforts. If a strong correlation can be established it will significantly help our understanding of policies which have had positive impact in improving local economic conditions. The data above in the table helps us recreate this time series graph.

Nightlight Adjusted by Population

So far using the above pre-processing steps we have defined district and province boundaries and aggregated the nightlight data points within these boundaries to give an estimation of the general economic activity in these locations. This estimate can be improved significantly if adjusted for population. The normalized night illumination will be a much better indicator of economic activity as compared to just the radiance levels. It is expected that an area with sparse population will have low night illumination and thus does not necessarily indicate poor local economy. But an area with high population yet low nightlight illumination suggests the exact opposite.

The first two images given below represent the population of the district and the total radiance in the district respectively. Figure A.8 gives the radiance per capita in the district. Some districts such as Ghoryan, Kohsan and Qalat do not have very high population but still have good comparative night illumination. This indicates a better economic situation as compared to most central Afghanistan districts which all have population but negligible (0 radiance throughout the year) nightlight. Kabul and some district capital can be seen to clearly stand out with high population and high concentration of nightlights.

Figure A.8: Comparison of Illumination and Population Density



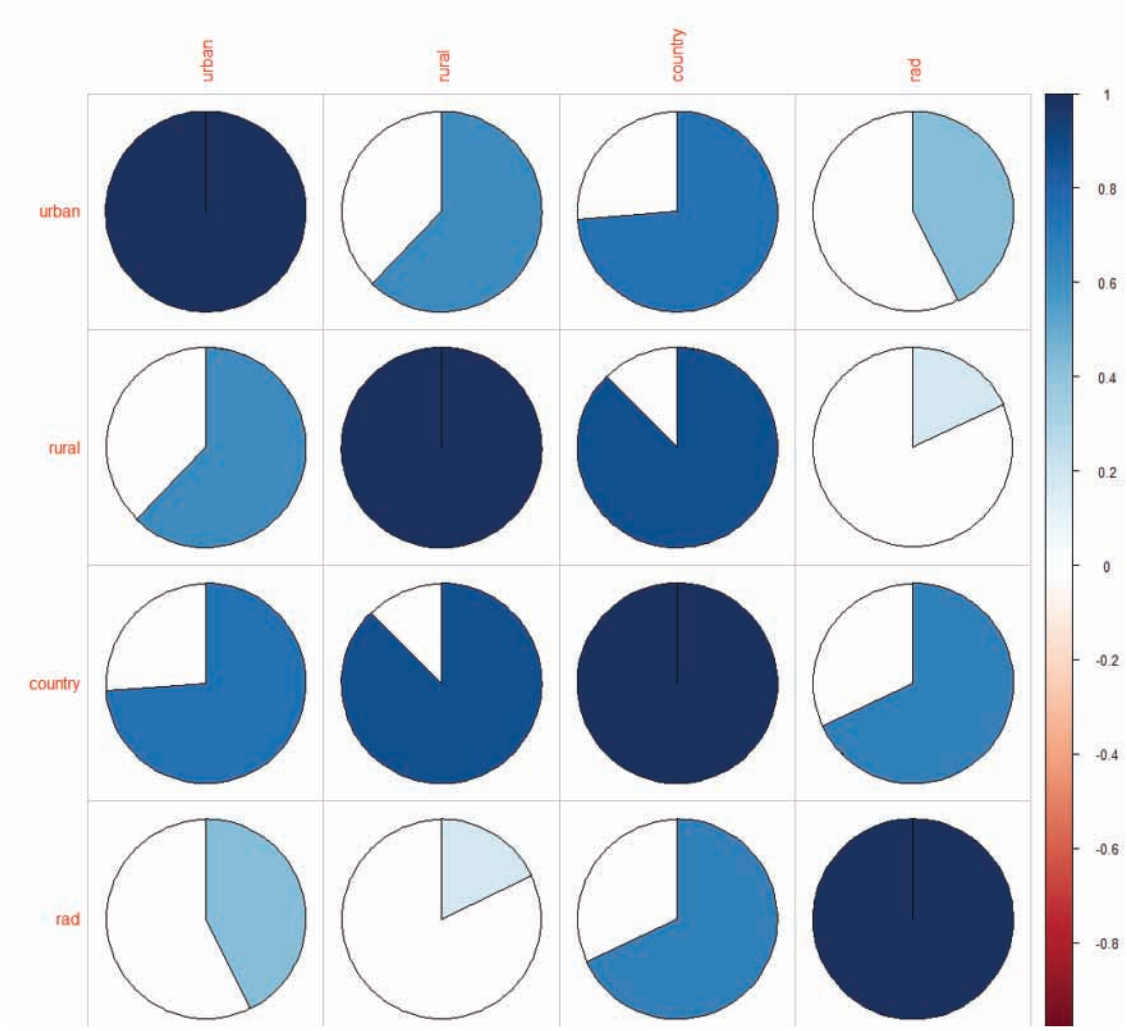
Cross Validation

The nightlights numbers were cross validated against various survey metrics for robustness. As expected there is a high positive correlation between survey economic indicators and the nightlight numbers both at the province and district levels. Figure A.9 shows the correlation between nightlights and these economic measures. The last column shows correlation of nightlight radiance with Urban, Rural and Country Wealth Index. Overall there is a positive correlation between nightlights and economic measures with high correlations for many of the household income measures.

The results of cross validation are as follows:

1. **MISTI Survey**- The nightlights have a **71.05%** positive correlation with "*Total Household Monthly Income from All Sources*".
2. **NRVA** – Compared cross sectionally (for the year 2011- 2012) the nightlights are **60.05%** correlated with *"Total Household Income"* in Afghanistan.
3. **DHS** – DHS gives a wealth index score (up to 5 decimal points) for rural areas, urban areas and the country as a whole. As expected, nightlights in rural areas were positively correlated but the correlation was weak (18.1 %). In urban area the correlation was much stronger (42.57 %). As a whole, nightlights were **67.98%** correlated with wealth index score across the whole country.

Figure A.9: Comparison of Nightlights Measure with other Economic Survey Measures



Correlation with Violence Indicators and Troop presence

We also tested whether the nightlights in Afghanistan are truly an accurate indication of the economic activity or whether the results are biased by violent activity and/or ISAF troop presence. The bias could come from key security features of bases such as base security lighting as well as the economic activity that surrounds bases such as logistic services to the base. Figures A.10 and A.11 show the correlation with the most violence and least violent districts respectively.

Figure A.10: Comparison of Nightlights Measure with other Economic Survey Measures

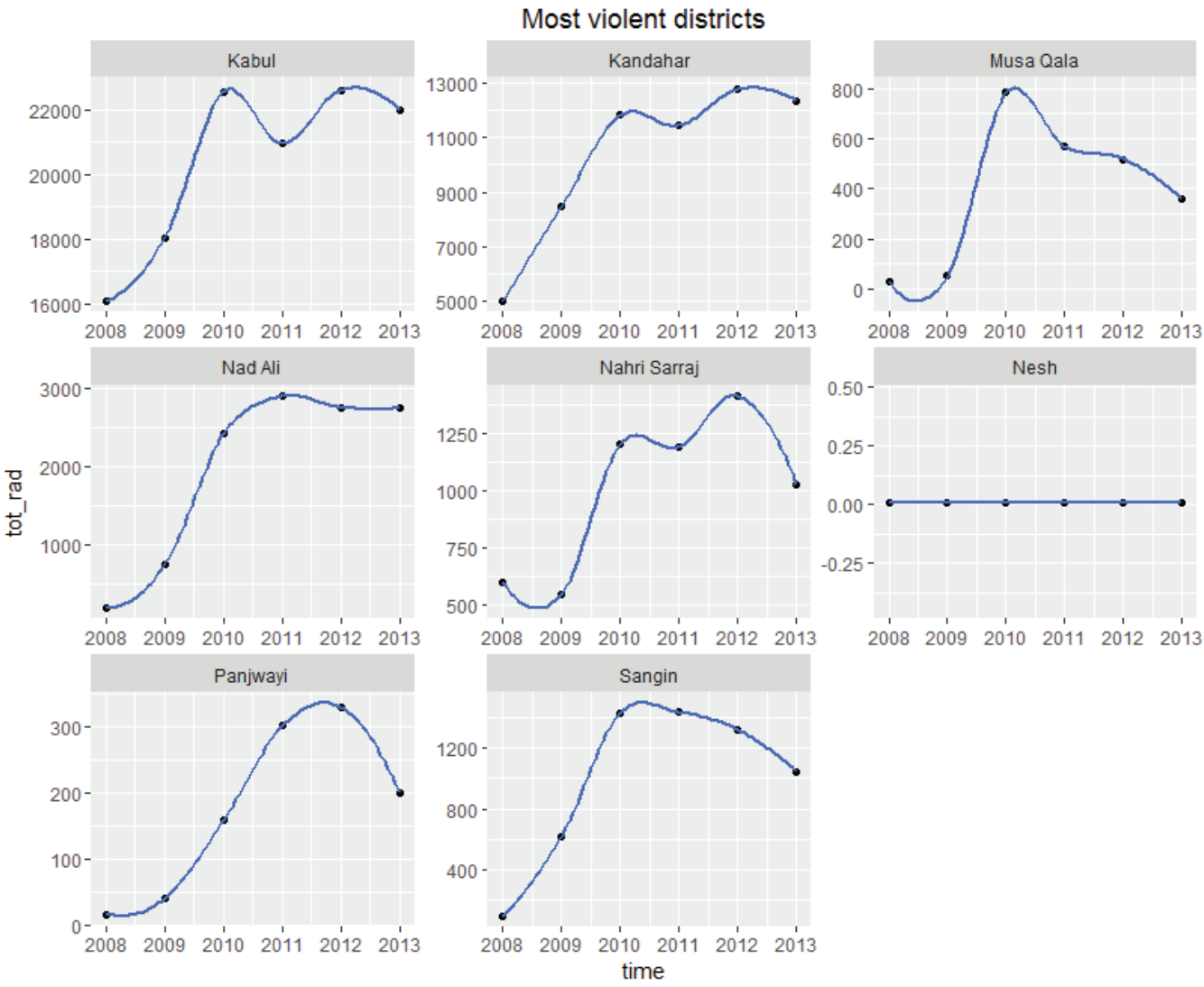


Figure A.11: Nightlights in Least Violent Districts

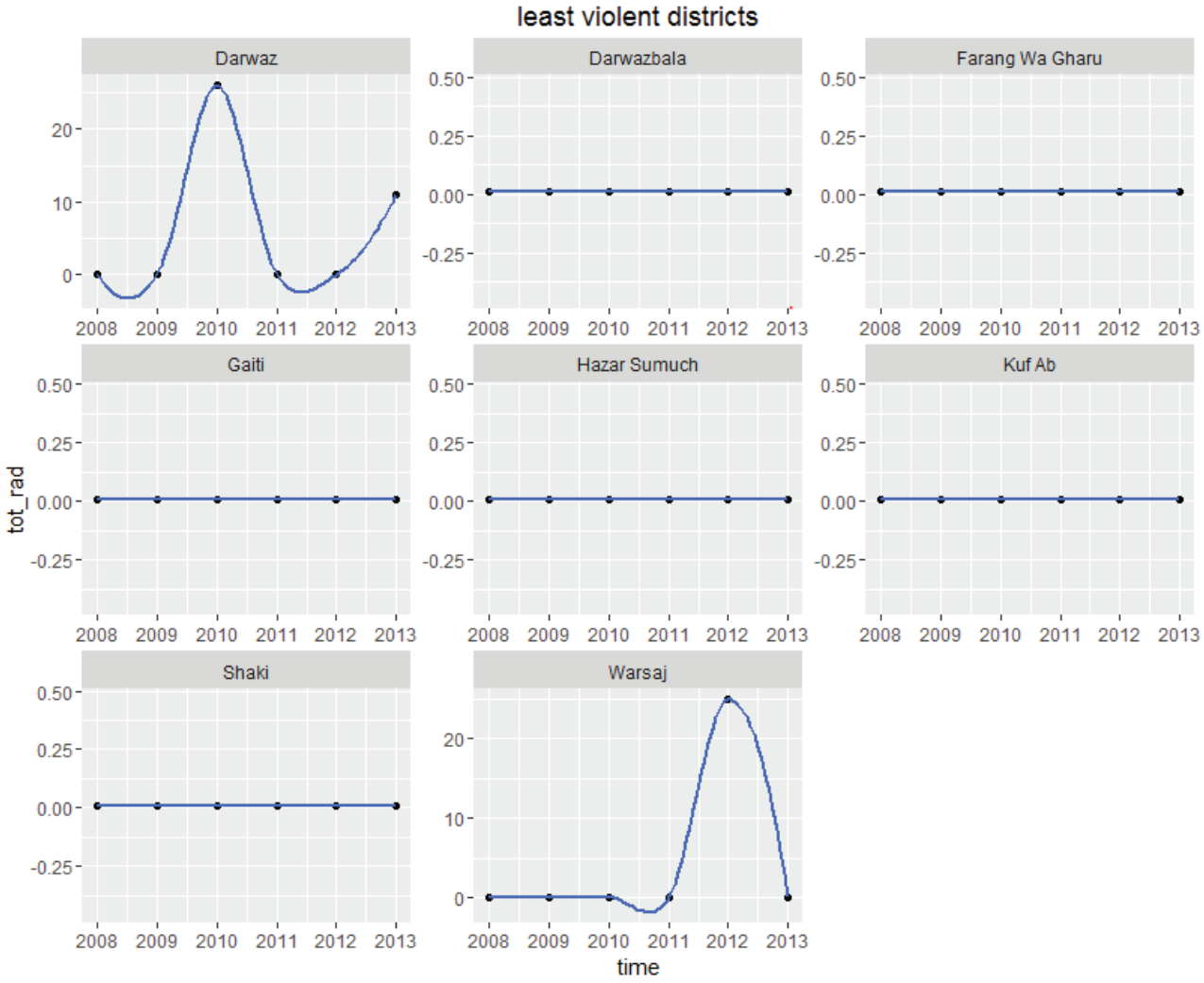


Figure A.12 shows the relationship between violence and the nightlights measure overall. The plot shows the correlation between violence (SIGACTs) and Nightlights (rad). Although there is a significant correlation ~34%, this number drops substantially after controlling for population. Thus violence per capita (log_sig_pc) and nightlights per capita (log_rad_pc) have a very weak correlation (~11%). A significance test (0.95) suggests that relationship is not statistically significant.

Figure A.12. Correlation between Nightlights and SIGACTs

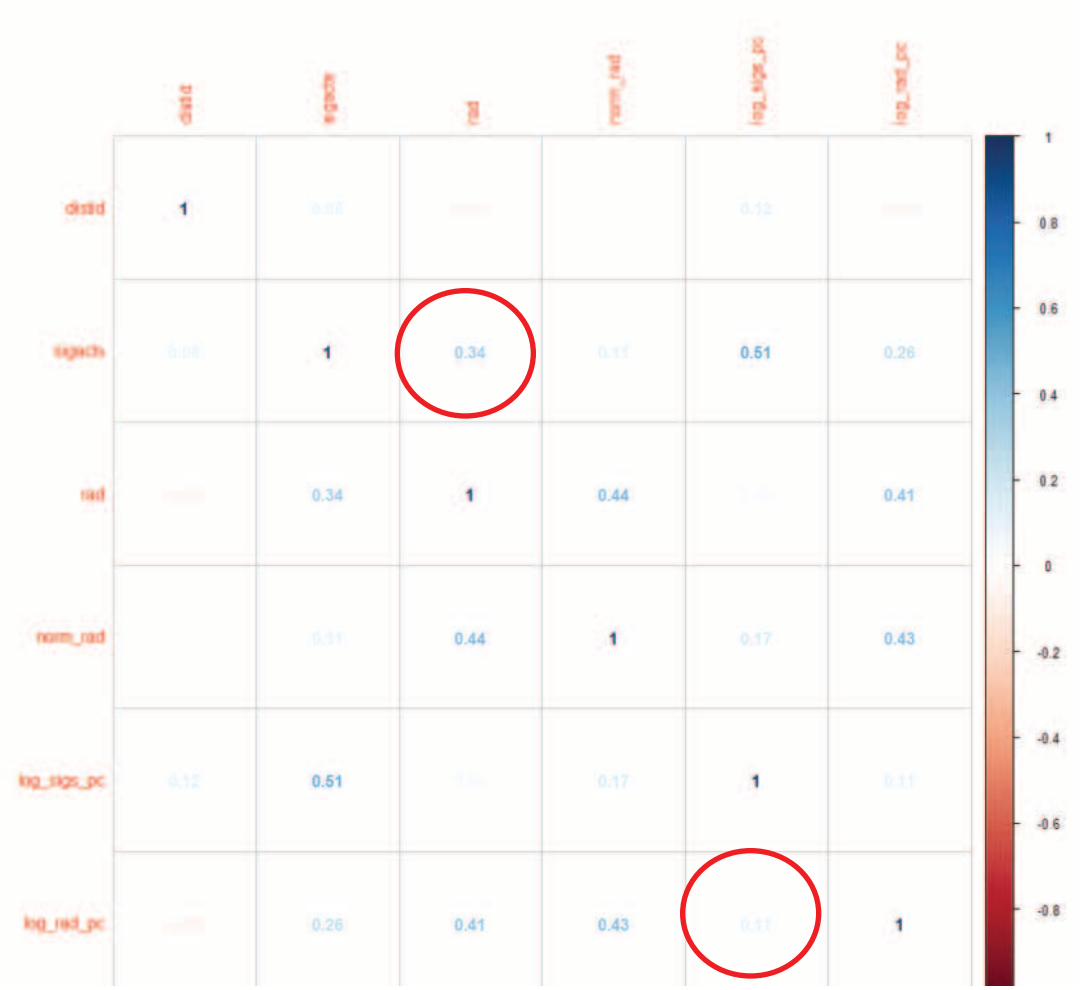
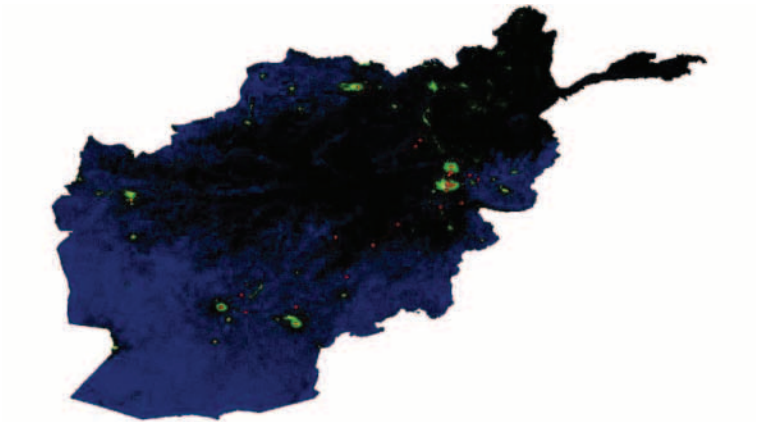


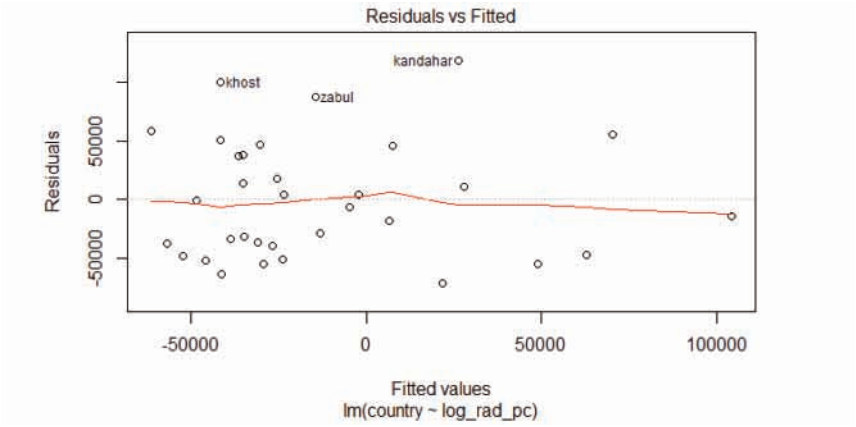
Figure A.13 compares troop presence and nightlights. Visually, while there is overlap, there are a number of low troop density areas with nightlights and high troop density areas with low radiance

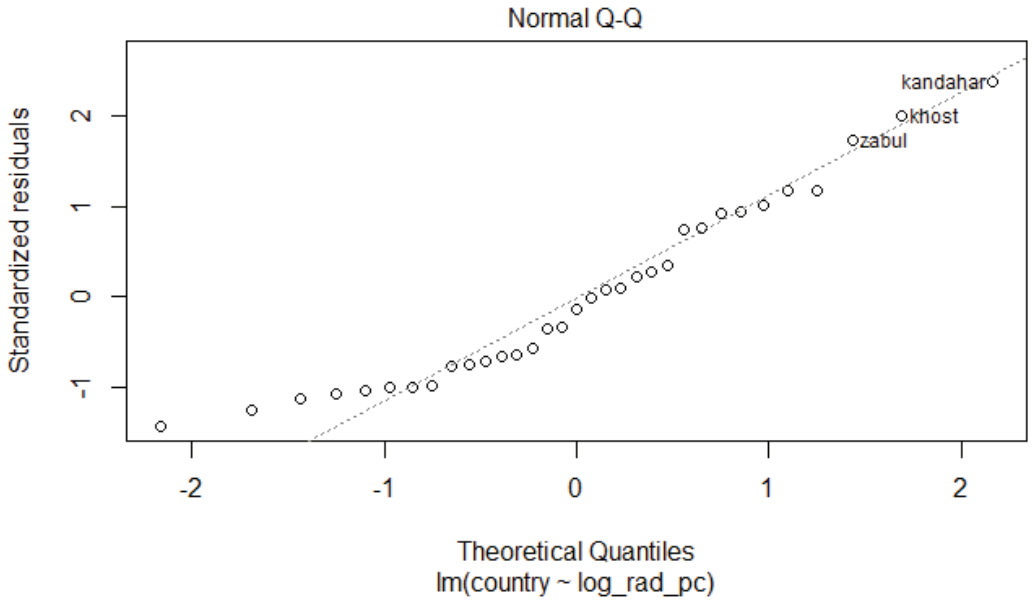
Figure A.13: Bright Nightlights Spots Overlaid with Troop Presence



Lastly we estimate how well nightlights are predicting economic activity and if any nightlights data is being contaminated by nearby violence or ISAF troop presence. Thus we run *a linear regression to understand how well nightlights are predicting local economic activity* (in the particular test DHS country wealth index score has been used). The summary and diagnostics (with some outlier points) of this regression are plotted below. It can be seen that in most cases nightlights can capture very effectively economic activity. Only when the province is in the state of extreme violence or is primarily rural do we notice a significant difference between lights and economic activity.

Figure A.14: Correlation between Nightlights and Regression Residuals

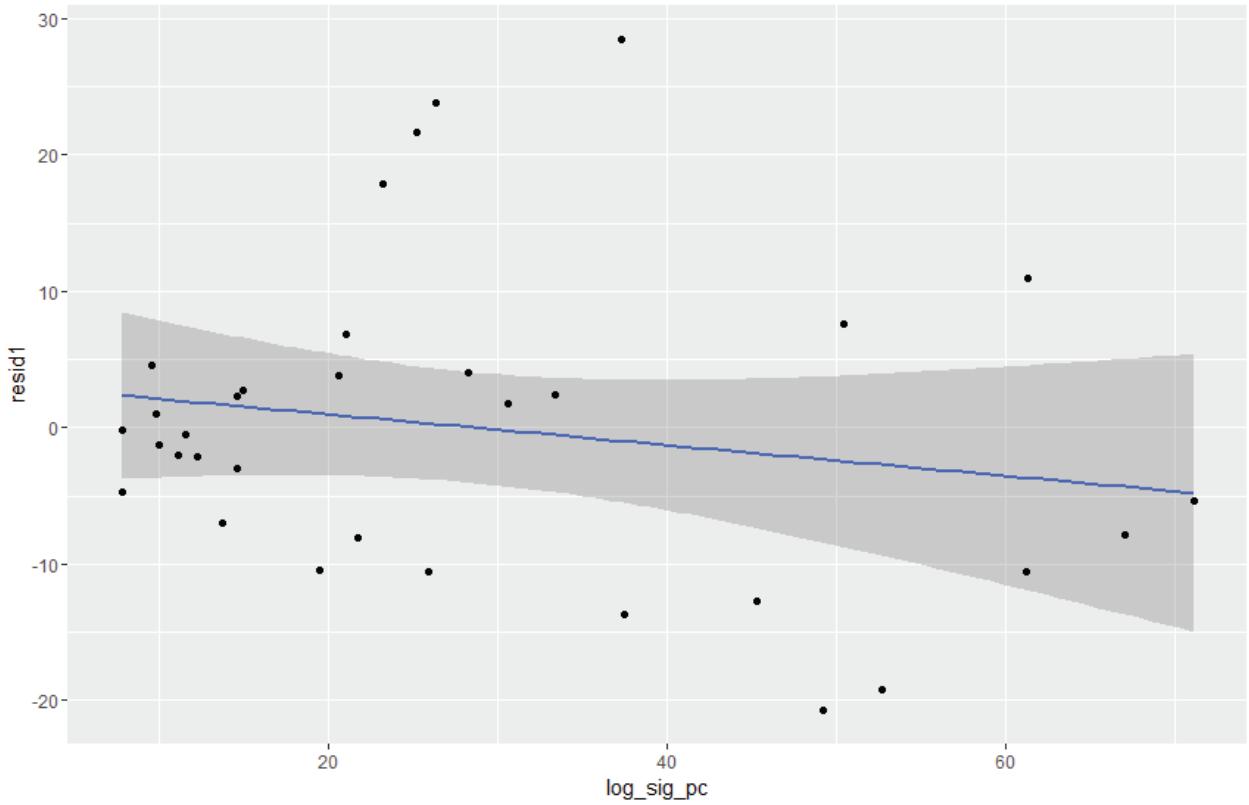




The adjusted R squared value for the above regression is 0.361 and this is a statistically significant finding with a p value of 0.000124.

To understand if any of these residuals could be explained by violence we ran a regression of violent events (log of SIGACTs per capita) versus residuals (residuals of wealth as explained by nightlights).

A.15: Correlation between SIGACTS and Nightlights Residuals



The regression has a very small R squared value (0.0029) and a very high p value (0.3039). Thus the most likely most of the variance or deviation in the residuals (residuals of wealth as explained by nightlights) ***cannot*** be effectively be explained by violence.

Data limitations

In the analysis conducted so far we have only been using both the DMSP (for a larger portion) and VIIRS data. Some Inherent drawbacks of the DMSP data are:

1. Coarse spatial resolution
2. Saturation on bright lights
3. No access to actual data values just re-scaled values which have a fixed range of 0-64

Despite the above limitations, historic nightlights data at present is best represented by DMSP. Although VIIRS data has significant advantage over DMSP, the technology and relevant satellites have been launched very recently. Comparison across the two measuring techniques will not yield meaningful results. Thus the two are evaluated and compared separately using exactly the same processing. When appropriate, we will do additional analyses using VIIRS data.

Furthermore the composites available freely to public are aggregated at a yearly level. If the data could be given at a monthly / quarterly level, a more accurate time series could be formed which could help us understand better if the rises and drops in illuminations correlate with stabilization efforts.

A.4 Health Access Outcomes

The PAP specified use of the DHS and NRVA as potential sources to measure health and wellbeing of the Afghan population. In particular, the DHS has a robust, internationally recognized, and regularly validated survey with a host of relevant health questions that can be tracked over time and across Afghanistan including measures of maternal mortality. However, the latest DHS wave results (2015) were not publicly available at the time of this analysis, and the previously available dates (2005, 2009) did not match with the times of the available program data to facilitate the broad panel analysis. We also explored using the NRVA, which has a section on maternal and child health. In both surveys, we used questions related to: birth rates for males and females, child mortality, antenatal visit by a doctor, nurse, child care worker, or midwife; birth assisted by doctor; and experience with recent health conditions including fever, cough, breathing difficulties, and diarrhea.

As shown in Table A.2, the relevant questions from DHS and NRVA section well correlated. However, the NRVA was also not available for the later years of program data.

Table A.2: Correlation between Various Surveys with Health Measures

	DHS	NRVA	ANQAR
DHS	100.0%	56.8%	67.1%
NRVA	56.8%	100.0%	37.5%
ANQAR	67.1%	37.5%	100.0%

We therefore focused on the ANQAR which asked the question, “How would you rate the quality of healthcare available in your area?” We note that this question is related to health services not health outcomes and primarily relates to the quality (and availability) of these services. This question correlated well with the DHS and somewhat with the NRVA questions.

Appendix B – Detailed Results

This appendix provides the detailed findings related to the regressions detailed in the pre-analysis plan (PAP). In the notation of the PAP, regressions are estimated using a panel dataset structured in location-time pairs such that a unit of observation is a variable value for area j (a district) at time t (a quarter). We define our variables of interest as the vector P_{jt} = *[Project spending* $_{jt}$*, Project count* $_{jt}$*]*. To describe these data, we will conduct a basic graphic analysis to show the regional variation in total level of project spending and sector specific project existence as defined above.

For outcome variables, we define V_{jt} as a per capita count of violent incidents, G_{jt} as the location-time index measure of support for the Afghan government in area j at time t, A_{jt} as the location-time index measure of support for AGE in area j at time t, C_{jt} as the location-time pair measure of community cohesion in area j at time T, H_{jT} as location-time pair measure of health outcomes in area j at time T and E_{jT} as location-time pair measure of economic outcomes in area j at time T. Health (H_{jT}) and economic outcomes (E_{jT}) are longer-run measures (denoted in the equations with a T instead of t) and thus we anticipate estimating their relationship with programmatic spending using annual measures.

We used indices to measure attitudinal changes for support of the Afghan government, support for anti-government elements (AGE) and community cohesion $C_{jt} = \{C_{jt}^{Add}, C_{jt}^{PCA}\}$. For ease of notation, we define the complete set of these variables Y, so that $Y_{jt} = \{G_{jt}, A_{jt}, C_{jt}\}$

The description below outlines our forms of analysis and lists on the left-hand side of each equation the relevant outcome variables.

B.1 Initial Estimation of Correlation in Program Activity and Key Outcomes

We begin our analysis by presenting the raw correlations between the level of USAID spending (or program presence) and the outcomes of interest (number of violent incidents per capita, the index of government support, the index of AGE support, and the index of community cohesion). This analysis will help us understand how the level of spending or program presence is associated with the outcomes of interest and serve as a baseline for comparison with other conditional correlations estimated in subsequent analysis. While this will not address any confounding factors, comparison between this relationship, which we label β_1, is useful in understanding the degree to which the observed correlation between USAID spending (or sector-specific program presence) and various outcomes (violence, popular support, etc.) is actually due to a third, related factor (such as troop presence or other donor activities).

We estimate the coefficient β_1 from the linear regression specified in (1) which can be interpreted as the best fitting line through the scatter plots described above.

Precisely, for violence we will estimate:

$$\{V_{jt}, Y_{jt}\} = \beta_0 + \boldsymbol{\beta_1 P_{jt}} + \tau_t + \varepsilon_{jt} \quad (1)$$

Table B.1 presents the relationships between spending across a number of projects and each of our outcomes using parsimonious specifications with only quarter fixed effects to control for the different time period in which each of the programs operate.[50] Coefficients represent a $1,000 change in spending.

Table B.1: Raw Relationships Between Program Spending and Key Outcomes

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.128**	0.004	0.000	0.001	0.008***	-0.001***
	(0.043)	(0.003)	(0.001)	(0.001)	(0.002)	(0.000)
R^2	0.109	0.075	0.026	0.144	0.019	0.010
Observations	21041	6604	5399	3979	3634	3573
Panel B: Total (OTI, LGCD, NSP)						
Total	-0.185***	0.016	-0.005	0.006	0.002	-0.001
	(0.045)	(0.011)	(0.007)	(0.005)	(0.008)	(0.001)
R^2	0.103	0.075	0.026	0.144	0.016	0.009
Observations	21041	6604	5399	3979	3634	3573
Panel C: OTI						
OTI	3.579***	0.127	0.023	-0.021	-0.055	-0.009***
	(1.076)	(0.101)	(0.068)	(0.042)	(0.062)	(0.002)
R^2	0.107	0.075	0.026	0.144	0.016	0.010
Observations	21041	6604	5399	3979	3634	3573
Panel D: LGCD						
LGCD	0.531	0.090	0.162*	.	0.081*	.
	(0.278)	(0.052)	(0.064)	.	(0.034)	.
R^2	0.103	0.076	0.028	.	0.017	.
Observations	21041	6604	5399	.	3634	.
Panel E: NSP						
NSP	-0.293***	0.011	-0.007	0.006	0.000	-0.001
	(0.057)	(0.011)	(0.007)	(0.005)	(0.008)	(0.001)
R^2	0.105	0.075	0.027	0.144	0.015	0.009
Observations	21041	6604	5399	3979	3634	3573

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. All coefficients represent response for a $100,000 increase in spending from the specified data source. Each regression includes quarter fixed effects to account for changes in survey design over time. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

From the raw regressions, there are relatively few consistent findings across programs. In general, stabilization programming appears to have been directed towards more violent areas (with the exception of NSP) and areas with less economic activity. Programs also appear to be implemented in areas with higher support for government.

Table B.2 presents the relationships between spending across a number of projects and each of our outcomes using a parsimonious specification with only quarter fixed effects to control for the different time period in which each of the programs operate. Coefficients represent a $1,000

[50] For each specification, Kabul observations have been dropped because the high concentration of spending and other activities drove the overall results. Kabul is added back in for the fixed effects regressions

change in spending. In most cases, although the magnitude of the coefficients may vary, the direction of the relationship is consistent. This is not the case for Afghan Info which shows a significant positive relationship between spending and violence, for example, but a negative insignificant relationship when looking at program presences in Table B.2. Similarly there is a positive association with support for government, a negative association for support for AGE, and a positive association with community cohesion. This is likely due to the differences in the Afghan Info sample available for analysis with spending data. These differences are discussed in detail in Iyengar, Shapiro, and Mao, 2017 but in many cases the programs with spending data are also programs which USAID identified as "stabilization" programs. The differences in these raw relationships support the notion that the programs with available spending information were targeted at more violent areas, while the broader set of programs were in more permissive, supportive settings.

Table B.2: Raw Relationships Between Program Presence and Key Outcomes

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	-0.482	0.361**	-0.301**	0.175*	0.127	-0.028
	(0.731)	(0.139)	(0.115)	(0.074)	(0.124)	(0.015)
R^2	0.102	0.079	0.034	0.151	0.017	0.016
N	21306	6711	5485	4039	3690	3618
Panel B: Total (NSP, OTI, LGCD)						
TOT	-0.355	0.169*	-0.069	0.080*	-0.007	-0.015
	(0.315)	(0.078)	(0.059)	(0.037)	(0.058)	(0.008)
R^2	0.102	0.078	0.028	0.148	0.016	0.011
N	21306	6711	5485	4039	3690	3618
Panel C: OTI						
OTI	6.021**	0.162	0.069	-0.058	0.008	-0.023***
	(1.932)	(0.174)	(0.140)	(0.086)	(0.124)	(0.006)
R^2	0.107	0.075	0.027	0.145	0.016	0.010
N	21306	6711	5485	4039	3690	3618
Panel D: LGCD						
LGCD	4.792***	-0.037	0.533***	.	0.083	.
	(0.928)	(0.160)	(0.150)	.	(0.145)	.
R^2	0.110	0.075	0.034	.	0.016	.
N	21306	6711	5485	.	3690	.
Panel E: NSP						
NSP	-0.976**	0.170*	-0.113*	0.094*	-0.014	-0.012
	(0.319)	(0.078)	(0.056)	(0.037)	(0.057)	(0.007)
R^2	0.103	0.078	0.030	0.149	0.016	0.011
N	21306	6711	5485	4039	3690	3618
Panel F: CERP						
CERP	0.482***	-0.002	0.012*	-0.843***	0.022***	-0.016*
	(0.144)	(0.006)	(0.006)	(0.080)	(0.005)	(0.007)
R2	0.122	0.075	0.029	0.145	0.019	0.009
N	21306	6711	5485	4039	3690	3618

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. All coefficients represent the conditional correlation between outcomes and the presence of the specified program in that district quarter (binary). Each regression includes quarter fixed effects to account for changes in survey design over time. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

Unlike the more stable measures of public attitudes, violence is a relatively volatile measure with large shifts from month to month as well as from year to year. As such, we also estimate this outcome in first-differences. In this specification, rather than estimating the relationship in the level of *V* and vector, ***P***, we will construct the measure $\Delta V = V_{jt} - V_{jt-1}$ and $\Delta P = P_{jt} - P_{jt-1}$ and then estimate

$$\Delta V_{jt} = \beta_0 + \boldsymbol{\beta_1}(\Delta \boldsymbol{P_{jt}}) + \varepsilon_{jt} \quad (2)$$

For completeness, we show this first difference specification for each set of outcomes in Table B.3 where the coefficients here represent the correlation between changes in spending and changes in outcomes. For instance, a $1,000 increase in spend in period *t* compared to period *t*-1 is associated with a decrease of 0.0007 violent events per capita in that quarter for total OTI, LGCD, and NSP spending. In general, we do observe that as spending increases, violence decreases.

Table B.3: First Differences of Program Spending on Key Outcomes

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.020*	-0.004	-0.002	-0.003	-0.015*	0.000
	(0.008)	(0.002)	(0.002)	(0.002)	(0.007)	(0.000)
R^2	0.075	0.048	0.044	0.056	0.013	0.038
Observations	20644	5136	4211	2943	1926	3176
Panel B: Total (OTI, LGCD, NSP)						
Total	-0.073***	-0.003	0.008	-0.006	-0.020	-0.000
	(0.018)	(0.009)	(0.009)	(0.004)	(0.013)	(0.000)
R^2	0.075	0.048	0.044	0.056	0.013	0.038
Observations	20644	5136	4211	2943	1926	3176
Panel C: OTI						
OTI	-0.496	-0.054	-0.062	-0.010	-0.004	0.001
	(0.338)	(0.065)	(0.033)	(0.029)	(0.103)	(0.001)
R^2	0.075	0.048	0.044	0.055	0.011	0.038
Observations	20644	5136	4211	2943	1926	3176
Panel D: LGCD						
LGCD	-0.036	0.022	-0.202***	.	0.150***	.
	(0.037)	(0.039)	(0.046)	.	(0.042)	.
R^2	0.075	0.048	0.046	.	0.013	.

Observations	20644	5136	4211	.	1926	.
Panel E: NSP						
NSP	-0.075***	-0.003	0.011	-0.006	-0.024	-0.000
	(0.020)	(0.009)	(0.009)	(0.004)	(0.013)	(0.001)
R^2	0.075	0.048	0.045	0.056	0.013	0.038
Observations	20644	5136	4211	2943	1926	3176

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. All coefficients represent the estimated change in outcome for a $10,000 change in spending from the specified data source from the previous period. Each regression includes district and quarter fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

Table B.4 finds very limited evidence of any systematic differences in outcomes when looking at first differences and program presence. The similarity in outcomes in the levels and first differences specification is useful in confirming the use of levels and ruling out differential trends.

Table B.4: First Differences of Program Presence on Key Outcomes

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.590	-0.071	0.040	0.030	-0.121	0.005
	(0.521)	(0.108)	(0.090)	(0.056)	(0.216)	(0.003)
R^2	0.074	0.047	0.043	0.054	0.011	0.037
N	20909	5243	4297	3003	1982	3221
Panel B: Total (NSP, OTI, LGCD)						
TOT	-0.080	-0.034	0.002	-0.037	-0.105	-0.002
	(0.133)	(0.058)	(0.043)	(0.029)	(0.079)	(0.002)
R^2	0.074	0.047	0.043	0.054	0.012	0.037
N	20909	5243	4297	3003	1982	3221
Panel C: OTI						
OTI	0.356	-0.160	0.012	-0.070	-0.257	-0.008
	(0.835)	(0.161)	(0.066)	(0.075)	(0.310)	(0.004)
R^2	0.074	0.047	0.043	0.054	0.011	0.037
N	20909	5243	4297	3003	1982	3221
Panel D: LGCD						
LGCD	-0.087	0.243*	-0.177	.	0.231	.
	(0.316)	(0.119)	(0.165)	.	(0.298)	.
R^2	0.074	0.048	0.044	.	0.011	.
N	20909	5243	4297	.	1982	.
Panel E: NSP						
NSP	-0.172	-0.072	0.031	-0.031	-0.132	-0.001
	(0.133)	(0.061)	(0.044)	(0.029)	(0.078)	(0.002)
R^2	0.074	0.047	0.043	0.054	0.013	0.037
N	20909	5243	4297	3003	1982	3221
Panel F: CERP						
CERP	0.035	-0.006	-0.001	-0.536***	-0.020	0.007**
	(0.020)	(0.004)	(0.002)	(0.075)	(0.029)	(0.002)

R2	0.074	0.047	0.043	0.054	0.011	0.037
N	20909	5243	4297	3003	1982	3221

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. All coefficients represent the estimated change in outcome for a 0 to 1 change in presence from the specified data source from the previous period. Each regression includes district and quarter fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

There are a number of area and time specific factors which may generate a spurious relationship between the outcomes of interest and project spending (or program presence). We will therefore next estimate the relationship between project spending (or program presence) and the short-term outcomes, accounting for region and time specific effects that likely drive both spending and violence levels. This net effect, which we call β_1^C, can be compared to β_1, to understand how much of the relationship between spending/program existence and other outcomes is related to the location and time factors that underlie both aspects.

For each of the short-term outcomes we estimate β_1^C with a fixed effects regression controlling for location and time effects:

$$\{V_{jt}, Y_{jt}\} = \beta_0^C + \boldsymbol{\beta_1^C P_{jt}} + \delta_j + \tau_t + \varepsilon_{jt} \quad (3)$$

The conditional correlation β_1^C, should be interpreted as the average, contemporaneous relationship between violence and project spending (or sector-specific program existence).

Table B.5: Fixed Effect Regressions (General Controlled Correlation) of Program Spending on Key Outcomes

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.112***	-0.003	0.000	-0.000	-0.003	0.000
	(0.033)	(0.002)	(0.001)	(0.001)	(0.003)	(0.000)
R^2	0.441	0.346	0.413	0.525	0.290	0.793
Observations	21041	6604	5399	3979	3634	3573
Panel B: Total (OTI, LGCD, NSP)						
Total	-0.109***	0.006	0.004	0.002	-0.005	-0.001
	(0.032)	(0.007)	(0.005)	(0.003)	(0.007)	(0.001)
R^2	0.438	0.346	0.413	0.525	0.290	0.793
Observations	21041	6604	5399	3979	3634	3573
Panel C: OTI						
OTI	1.128	0.086	-0.106**	0.005	-0.031	0.000
	(0.696)	(0.081)	(0.038)	(0.032)	(0.061)	(0.001)
R^2	0.438	0.346	0.414	0.525	0.290	0.793
Observations	21041	6604	5399	3979	3634	3573
Panel D: LGCD						
LGCD	0.193	0.084	-0.019	.	0.119**	.
	(0.109)	(0.046)	(0.051)	.	(0.041)	.
R^2	0.437	0.347	0.413	.	0.292	.
Observations	21041	6604	5399	.	3634	.
Panel E: NSP						
NSP	-0.152***	0.001	0.006	0.001	-0.010	-0.001

	(0.036)	(0.007)	(0.006)	(0.003)	(0.007)	(0.001)
R^2	0.438	0.346	0.414	0.525	0.290	0.793
Observations	21041	6604	5399	3979	3634	3573

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. All coefficients represent response for a $100,000 increase in spending from the specified data source. Each regression includes district and quarter fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

As shown in Table B.5 there is very limited effect on program activity on any of the outcomes of interest. As discussed in the main body of the report, the negative correlation between violence and program activity for NSP is likely due to the selection of NSP sites in more permissive settings. This is consistent with estimates from program existence shown in Table B.6.Only CERP appears to have any significant relationship. Not surprisingly, CERP is associated with higher levels of violence—likely due to the selection of where to execute the programs by the military. CERP is also associated with lower levels of community cohesion, again potentially based on the selection of where to conduct programs. CERP does appear associated with a relatively small, positive difference in health and economic outcomes which may be due to the program or to a host of other intervening activities in the area.

Table B.6: Fixed Effect Regressions (General Controlled Correlation) of Program Presence on Key Outcomes

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	-0.508	-0.071	-0.020	0.032	-0.142	0.002
	(0.515)	(0.115)	(0.069)	(0.052)	(0.108)	(0.002)
R^2	0.434	0.344	0.414	0.525	0.290	0.793
N	21306	6711	5485	4039	3690	3618
Panel B: Total (NSP, OTI, LGCD)						
TOT	0.293	0.071	-0.055	0.023	-0.025	-0.005
	(0.236)	(0.057)	(0.035)	(0.024)	(0.051)	(0.003)
R^2	0.434	0.345	0.414	0.525	0.290	0.793
N	21306	6711	5485	4039	3690	3618
Panel C: OTI						
OTI	2.161	0.052	-0.171*	-0.027	0.031	-0.006
	(1.126)	(0.131)	(0.079)	(0.057)	(0.114)	(0.004)
R^2	0.435	0.344	0.414	0.525	0.290	0.793
N	21306	6711	5485	4039	3690	3618
Panel D: LGCD						
LGCD	1.959**	0.251*	0.066	.	0.359**	.
	(0.624)	(0.113)	(0.109)	.	(0.132)	.
R^2	0.435	0.345	0.414	.	0.292	.
N	21306	6711	5485	.	3690	.
Panel E: NSP						
NSP	0.026	0.052	-0.035	0.029	-0.066	-0.004
	(0.241)	(0.057)	(0.034)	(0.024)	(0.049)	(0.003)
R^2	0.434	0.345	0.414	0.525	0.290	0.793
N	21306	6711	5485	4039	3690	3618
Panel F: CERP						
CERP	0.326**	-0.005	0.005	-0.975***	0.025***	0.005*

	(0.126)	(0.006)	(0.005)	(0.186)	(0.006)	(0.002)
R2	0.442	0.345	0.414	0.526	0.292	0.793
N	21306	6711	5485	4039	3690	3618

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. All coefficients represent response for a $100,000 increase in spending from the specified data source. Each regression includes district and quarter fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

B.2 Addressing Security Force Presence as a Confounding Factor

A key confounding factor in the relationship between violence and programmatic activity is security force levels (which we measure with ISAF military presence).[51] We therefore estimate the relationship between stabilization program activities and the outcomes of interest, conditional on security force levels, a value we call β_1^{SF}. These relationships are estimated through a linear regression of the form:

$$\{V_{jt}, Y_{jt}\} = \beta_0^{SF} + \boldsymbol{\beta_1^{SF} P_{jt}} + \gamma^{SF} F_{jt} + \delta_j + \tau_t + \varepsilon_{jt} \quad (4)$$

Table B.7 presents the fixed effects regression controlling for presence of international security forces. It is clear that the presence of military actors does have a substantial confounding effect on violence. Across the board, military presence has a large, positive, and frequently significant effect on the level of violence in an area.

Table B.7: Regressions of Program Spending on Key Outcomes Controlling for Security Force Presence

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.076*	-0.006*	0.003	-0.002	-0.001	-0.000
	(0.037)	(0.003)	(0.001)	(0.001)	(0.003)	(0.000)
M	0.410	0.017	-0.012	0.008	-0.022	0.001*
	(0.221)	(0.011)	(0.006)	(0.006)	(0.013)	(0.001)
R^2	0.660	0.372	0.452	0.576	0.242	0.886
N	1500	1045	913	654	561	540
Panel B: Total (OTI, LGCD, NSP)						
TOT	-0.350**	0.003	0.005	-0.003	-0.012	-0.000
	(0.118)	(0.009)	(0.008)	(0.005)	(0.017)	(0.000)
M	0.451*	0.011	-0.010	0.005	-0.023*	0.001*
	(0.218)	(0.010)	(0.006)	(0.006)	(0.011)	(0.001)
R^2	0.660	0.368	0.450	0.575	0.242	0.886
N	1500	1045	913	654	561	540
Panel C: OTI						
OTI	-3.858**	-0.005	-0.080*	-0.031	-0.106	-0.000
	(1.287)	(0.116)	(0.040)	(0.038)	(0.064)	(0.001)

[51] We note that this measure does not include Afghan National Security Force (ANSF) presence. Currently we do not have reliable information on ANSF force levels. We are working to obtain that information at least at the province level and will include separate controls for ANSF or at least Afghan National Army force levels if feasible.

M	0.446*	0.012	-0.010	0.004	-0.024*	0.001*
	(0.217)	(0.010)	(0.006)	(0.006)	(0.011)	(0.001)
R^2	0.662	0.368	0.451	0.576	0.244	0.886
N	1500	1045	913	654	561	540
PANEL D: LGCD						
LGCD	0.976	-0.015	0.022	.	0.258**	.
	(0.499)	(0.127)	(0.083)	.	(0.080)	.
M	0.441*	0.012	-0.010	0.005	-0.023	0.001*
	(0.216)	(0.010)	(0.006)	(0.005)	(0.012)	(0.001)
R^2	0.658	0.368	0.450	0.575	0.256	0.886
N	1500	1045	913	654	561	540
PANEL E: NSP						
NSP	-0.340**	0.004	0.007	-0.002	-0.021	-0.000
	(0.120)	(0.009)	(0.008)	(0.006)	(0.017)	(0.000)
M	0.449*	0.011	-0.010	0.005	-0.022*	0.001*
	(0.217)	(0.010)	(0.006)	(0.006)	(0.011)	(0.001)
R^2	0.660	0.368	0.450	0.575	0.242	0.886
N	1500	1045	913	654	561	540

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $100,000 increase in spending from the specified data source. Each regression controls for Military presence and includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

The observed relationship with violence is even more robust in when focusing on program existence in specific districts. In these cases, the military presence is uniformly associated with higher levels of violence. This is not surprising both because the military was likely present in more violent areas and because any reporting bias in the SIGACTs data would be biased towards more reports in areas with the military present. Military presence also appears to be directly related to negative health and economic outcomes in some cases. It is noteworthy that controlling for military presence shrinks the size of the relationship between violence and program activity in both the spending and the presence regressions (Table B.8). This suggests that although stabilization programs operated in violent districts, they typically did so with the military present.

Table B.8: Regressions of Program Presence on Key Outcomes Controlling for Security Force Presence

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	-2.114	-0.147	-0.015	-0.115	-0.473*	0.006
	(1.688)	(0.245)	(0.168)	(0.084)	(0.225)	(0.004)
M	0.508*	0.008	-0.008	0.002	-0.026*	0.001*
	(0.233)	(0.010)	(0.007)	(0.006)	(0.011)	(0.001)
R^2	0.656	0.364	0.445	0.577	0.252	0.888
N	1600	1124	983	704	602	576
Panel B: Total (OTI, LGCD, NSP)						

TOT	-0.682	0.185	-0.061	0.041	0.038	-0.000
	(0.895)	(0.150)	(0.074)	(0.046)	(0.110)	(0.001)
M	0.517*	0.008	-0.007	0.003	-0.025*	0.001*
	(0.236)	(0.010)	(0.007)	(0.006)	(0.011)	(0.001)
R^2	0.656	0.366	0.446	0.577	0.244	0.886
N	1600	1124	983	704	602	576
Panel C: OTI						
OTI	-6.222*	-0.231	-0.067	-0.145	-0.080	-0.002
	(2.892)	(0.201)	(0.129)	(0.079)	(0.160)	(0.004)
M	0.517*	0.009	-0.007	0.002	-0.025*	0.001*
	(0.234)	(0.011)	(0.007)	(0.006)	(0.011)	(0.001)
R^2	0.659	0.364	0.445	0.579	0.244	0.886
N	1600	1124	983	704	602	576
PANEL D: LGCD						
LGCD	1.246	0.090	0.112	.	0.256	.
	(1.545)	(0.211)	(0.148)	.	(0.342)	.
M	0.510*	0.008	-0.007	0.003	-0.025*	0.001*
	(0.229)	(0.010)	(0.007)	(0.006)	(0.011)	(0.001)
R^2	0.656	0.364	0.445	0.576	0.246	0.886
N	1600	1124	983	704	602	576
PANEL E: NSP						
NSP	-0.175	0.219	-0.076	0.069	0.006	-0.000
	(0.974)	(0.133)	(0.069)	(0.045)	(0.105)	(0.001)
M	0.516*	0.007	-0.007	0.003	-0.025*	0.001*
	(0.236)	(0.010)	(0.007)	(0.006)	(0.011)	(0.001)
R^2	0.656	0.367	0.446	0.578	0.244	0.886
N	1600	1124	983	704	602	576
PANEL F: CERP						
CERP	0.346**	-0.003	0.004	.	0.023**	.
	(0.106)	(0.007)	(0.004)	.	(0.008)	.
M	0.537*	0.008	-0.008	0.003	-0.020*	0.001*
	(0.230)	(0.011)	(0.007)	(0.006)	(0.009)	(0.001)
R^2	0.668	0.364	0.446	0.576	0.255	0.886
N	1600	1124	983	704	602	576

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Acts per Capita. Spend is $100,000. Kabul Dropped. ECO is VIIRS.

However, security forces may be more than a confounding factor in estimating the relationship between various outcomes and project spending or counts; they may actually amplify or undermine the effectiveness of projects in a given area as suggested by prior work in Afghanistan on CERP spending. We therefore estimated the relationship between project spending (or sector-specific program existence) on key outcomes in two components: the first estimate measures the direct relationship between the program activity and the outcome of interest, which we will call β_1^{IE}, and second estimates the additional marginal effect of aid given a one unit increase in security forces, φ^{IE}, holding fixed the level of security forces. Precisely, we estimate:

$$\{V_{jt}, Y_{jt}\} = \beta_0^{IE} + \boldsymbol{\beta_1^{IE} P_{jt}} + \gamma^{IE} F_{jt} + \varphi^{IE}\left(P_{jt} * F_{jt}\right) + \delta_j + \tau_t + \varepsilon_{jt} \quad (5)$$

Table B.9 shows the results of this interaction. Notably, the interaction appears important for the broader set of programs in Afghan Info, but less relevant for the specific programs of OTI or LGCD. The direct effect of military presence on violence remains in most cases but there does not appear to be a significant interactive effect either amplifying or inhibiting the relationships.

Table B.9: Regressions of Program Spending on Key Outcomes including Security Force Presence Synergy

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.155**	-0.005	0.004*	-0.003	-0.005	0.000
	(0.046)	(0.005)	(0.002)	(0.005)	(0.005)	(0.000)
M	0.481*	0.019	-0.010	0.007	-0.026	0.001*
	(0.223)	(0.012)	(0.007)	(0.008)	(0.015)	(0.001)
AIxM	-0.004***	-0.000	-0.000	0.000	0.000	-0.000*
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.664	0.372	0.452	0.576	0.243	0.887
Panel B: Total (LGCD, OTI, NSP)						
TOT	-0.308**	0.005	0.004	-0.001	0.010	0.000
	(0.115)	(0.009)	(0.008)	(0.006)	(0.019)	(0.000)
M	0.471*	0.013	-0.011	0.006	-0.017	0.001**
	(0.225)	(0.010)	(0.006)	(0.006)	(0.009)	(0.001)
TOTxM	-0.019	-0.001	0.001	-0.000	-0.004**	-0.000**
	(0.015)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
R^2	0.660	0.368	0.451	0.576	0.252	0.888
Panel C: OTI						
OTI	-5.110	0.017	-0.121**	-0.026	-0.113	0.001*
	(2.933)	(0.127)	(0.038)	(0.041)	(0.082)	(0.001)
M	0.439*	0.012	-0.010	0.004	-0.024*	0.001*
	(0.220)	(0.010)	(0.006)	(0.006)	(0.011)	(0.001)
OTIxM	0.111	-0.002	0.004*	-0.000	0.001	-0.000
	(0.114)	(0.005)	(0.002)	(0.001)	(0.002)	(0.000)
R^2	0.663	0.368	0.452	0.576	0.244	0.886
Panel D: LGCD						
LGCD	0.542	-0.029	0.020	.	0.263**	.
	(0.392)	(0.130)	(0.089)	.	(0.080)	.
M	0.447*	0.012	-0.010	0.005	-0.023	0.001*
	(0.216)	(0.010)	(0.006)	(0.006)	(0.012)	(0.001)
LGCDxM	0.290	0.009	0.001	.	-0.003	.
	(0.240)	(0.007)	(0.005)	.	(0.005)	.
R^2	0.659	0.368	0.450	0.575	0.256	0.886
Panel E: NSP						
NSP	-0.292**	0.006	0.005	-0.001	-0.000	-0.000
	(0.108)	(0.009)	(0.008)	(0.006)	(0.019)	(0.000)

M	0.473*	0.013	-0.012	0.006	-0.016	0.001*
	(0.226)	(0.010)	(0.006)	(0.006)	(0.009)	(0.001)
NSPxM	-0.023	-0.001	0.001	-0.000	-0.004**	-0.000**
	(0.017)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
R^2	0.660	0.368	0.451	0.576	0.243	0.887
N	1500	1045	913	654	561	540

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $100,000 increase in spending from the specified data source. Each regression controls for Military presence, includes an interaction term between spending and military presence (a positive interaction denotes an increase in the effect observed in the non-interaction term) and includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

A similar, but weaker, relationship between military presence and violence in the Afghan Info data is evidence in the program existence specification, shown in Table B.10.

Table B.10: Regressions of Program Presence on Key Outcomes including Security Force Presence Synergy

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	-0.686	0.138	-0.148	-0.056	-0.252	0.003
	(1.277)	(0.270)	(0.173)	(0.095)	(0.231)	(0.003)
M	0.735*	0.079***	-0.041**	0.018*	0.024	-0.000
	(0.325)	(0.019)	(0.012)	(0.009)	(0.016)	(0.000)
AIxM	-0.276	-0.070***	0.034***	-0.015*	-0.047**	0.002*
	(0.244)	(0.014)	(0.009)	(0.006)	(0.014)	(0.001)
R^2	0.658	0.375	0.449	0.579	0.261	0.893
Panel B: Total (LGCD, OTI, NSP)						
TOT	-0.733	0.202	-0.076	0.065	0.136	0.001
	(1.211)	(0.162)	(0.083)	(0.044)	(0.117)	(0.001)
M	0.513*	0.010	-0.009	0.004	-0.017	0.002**
	(0.236)	(0.011)	(0.006)	(0.006)	(0.009)	(0.000)
TOTxM	0.015	-0.005	0.004	-0.005	-0.023**	-0.001*
	(0.146)	(0.009)	(0.006)	(0.004)	(0.009)	(0.000)
R^2	0.656	0.366	0.446	0.578	0.252	0.888
Panel C: OTI						
OTI	-6.509	-0.308	-0.064	-0.155	-0.113	0.001
	(3.819)	(0.228)	(0.156)	(0.083)	(0.184)	(0.004)
M	0.514*	0.008	-0.007	0.002	-0.025*	0.001**
	(0.238)	(0.011)	(0.007)	(0.006)	(0.011)	(0.001)
OTIxM	0.045	0.011	-0.000	0.002	0.005	-0.001*
	(0.167)	(0.012)	(0.005)	(0.004)	(0.006)	(0.000)
R^2	0.659	0.365	0.445	0.579	0.244	0.888
Panel D: LGCD						
LGCD	-0.751	0.062	0.123	.	0.279	.
	(1.182)	(0.224)	(0.166)	.	(0.354)	.
M	0.520*	0.009	-0.007	0.003	-0.026*	0.001*

	(0.226)	(0.010)	(0.007)	(0.006)	(0.012)	(0.001)
LGCDxM	0.753	0.010	-0.003	.	-0.010	.
	(0.520)	(0.012)	(0.008)	.	(0.017)	.
R^2	0.661	0.364	0.445	0.575	0.246	0.886
Panel E: NSP						
NSP	0.160	0.234	-0.075	0.088	0.128	0.001
	(1.107)	(0.143)	(0.080)	(0.044)	(0.116)	(0.001)
M	0.540*	0.009	-0.007	0.004	-0.014	0.001*
	(0.247)	(0.011)	(0.007)	(0.007)	(0.009)	(0.001)
NSPxM	-0.101	-0.004	-0.000	-0.005	-0.033***	-0.001*
	(0.125)	(0.010)	(0.006)	(0.004)	(0.010)	(0.000)
R^2	0.656	0.367	0.446	0.579	0.257	0.888
Panel F: CERP						
CERP	0.328***	-0.008	0.006	.	0.062***	.
	(0.046)	(0.007)	(0.003)	.	(0.018)	.
M	0.539*	0.009	-0.008	0.003	-0.021*	0.001*
	(0.237)	(0.010)	(0.007)	(0.006)	(0.009)	(0.001)
CERPxM	0.003	0.001	-0.000**	.	-0.002***	.
	(0.016)	(0.001)	(0.000)	.	(0.001)	.
R^2	0.668	0.365	0.446	0.576	0.262	0.886
N	1500	1045	913	654	561	540

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Overall, the findings appear to be consistent with the literature that military presence is an important factor in establishing baseline levels of security but that it may not affect the broader set of issues that many implementers face on the ground.

B.3 Longer Run Effects and Sustainability

We next turn to the effect of project spending on future levels of violence. We measure this by estimating the effect of violence six months to a year after a set of projects. This helps illustrate how future violence is related to spending (or sector-specific program presence) over a longer period of time. As illustrated in equation (6), we include controls for the total number of military forces present over this time period

$$\{V_{jt+n}, Y_{jt+n}\} = \beta_0^L + \boldsymbol{\beta_1^L}(\textstyle\sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \boldsymbol{P_{t+i}}) + \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \gamma_n^F F_{\boldsymbol{t+i}} + \delta_j + \tau_t + \varepsilon_{jt} \qquad \text{for } n = 1,4 \qquad (6)$$

In equation (5), depending on the results from the raw correlation analysis, we may also elect to extend the number of forward time periods beyond one year (currently one year is represented as n=4, assuming the dataset is constructed at a quarterly level).

The parallel to this for the longer-run outcomes of health and economic well-being will be to estimate the outcomes two and three years after program completion (when feasible).

$$\{H_{jT+1}, E_{jT+1}\} = \beta_0^C + \boldsymbol{\beta_1^{LT}} \boldsymbol{P_{jT}} + \gamma^{LT} F_{jT} + \delta_j + \tau_T + \varepsilon_{jT} \qquad (7)$$

In equation (6) we parallel the estimate in (4) to adjust for the presence of security forces during the period when programming was ongoing. This is consistent with recent World Bank research which finds a direct relationship between economic well-being and security force levels. Table B.11 presents the results for spending on key outcomes over a longer time horizon.

Table B.11a: Regressions of Program Spending on Key Outcomes over Longer Time Horizon

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.025***	-0.000	0.001	-0.000	-0.001	-0.000
	(0.007)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
R^2	0.462	0.346	0.414	0.525	0.290	0.793
N	19453	6604	5399	3979	3634	3573
Panel B: Total (LGCD, NSP, OTI)						
TOT	-0.030*	0.003	0.003	0.001	-0.001	-0.000
	(0.013)	(0.002)	(0.002)	(0.001)	(0.002)	(0.000)
R^2	0.459	0.346	0.414	0.525	0.290	0.793
N	19453	6604	5399	3979	3634	3573
Panel C: OTI						
OTI	0.038	0.074*	-0.041	0.016	-0.018	-0.001
	(0.331)	(0.034)	(0.021)	(0.014)	(0.022)	(0.000)
R^2	0.459	0.348	0.414	0.526	0.290	0.793
N	19453	6604	5399	3979	3634	3573
Panel D: LGCD						
LGCD	0.182**	0.045***	0.006	-0.004	0.014	.
	(0.062)	(0.013)	(0.014)	(0.031)	(0.011)	.
R^2	0.460	0.349	0.413	0.525	0.290	.
N	19453	6604	5399	3979	3634	.
Panel E: NSP						
NSP	-0.045***	-0.001	0.003*	0.001	-0.002	-0.000
	(0.013)	(0.002)	(0.001)	(0.001)	(0.002)	(0.000)
R^2	0.460	0.346	0.414	0.525	0.290	0.793
N	19453	6604	5399	3979	3634	3573

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $100,000 increase in cumulative spending from the previous year from the specified data source includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

Table B.11b: Regressions of Program Spending on Key Outcomes over Longer Time Horizon

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	-0.377	-0.009	-0.064**	0.043*	-0.046	-0.001
	(0.207)	(0.039)	(0.024)	(0.021)	(0.034)	(0.001)
R^2	0.456	0.344	0.415	0.527	0.290	0.793
N	19698	6711	5485	4039	3690	3618

Panel B: Total (LGCD, NSP, OTI)						
TOT	0.074	0.023	-0.022	0.018*	-0.008	-0.001
	(0.080)	(0.020)	(0.011)	(0.008)	(0.015)	(0.001)
R^2	0.456	0.345	0.414	0.526	0.290	0.793
N	19698	6711	5485	4039	3690	3618
Panel C: OTI						
OTI	0.159	0.096*	-0.073*	0.015	-0.019	-0.002
	(0.467)	(0.045)	(0.030)	(0.023)	(0.041)	(0.001)
R^2	0.456	0.345	0.415	0.525	0.290	0.793
N	19698	6711	5485	4039	3690	3618
Panel D: LGCD						
LGCD	0.955***	0.069	0.012	0.001	0.027	.
	(0.176)	(0.037)	(0.031)	(0.048)	(0.036)	.
R^2	0.459	0.345	0.414	0.525	0.290	.
N	19698	6711	5485	4039	3690	.
Panel E: NSP						
NSP	-0.013	0.017	-0.018	0.014	-0.010	-0.001
	(0.078)	(0.018)	(0.011)	(0.007)	(0.014)	(0.001)
R^2	0.456	0.345	0.414	0.526	0.290	0.793
N	19698	6711	5485	4039	3690	3618
Panel D: CERP						
NSP	0.131**	-0.001	0.003*	-0.001	0.006**	-0.008
	(0.047)	(0.002)	(0.001)	(0.003)	(0.002)	(0.004)
R^2	0.472	0.344	0.414	0.525	0.293	0.793
N	19698	6711	5485	4039	3690	3618

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $100,000 increase in cumulative presence level (number of quarters in the previous year) from the specified data source includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

Finally, we sought to estimate the effect of cumulative funding on outcomes six to 18 months after project implementation. The goal was to take the program relationships to medium term outcomes and partition these programs into the parallel set of four bins: programs with medium-term gains that were sustained, programs with medium-term gains that were not sustained, programs with no medium-term or longer-term gains, and programs with no medium-term gains but with longer terms gains. However, given the data limitations, longer-run analysis was not feasible.

B.4. Other Intervening Factors

Security forces are not the only source of interacting effects that may impact the relationship of stabilization programs with violence or public support. Given the large number of donors and broad swathes of efforts in Afghanistan over the past decade, we also estimate the relationship between aid by one program and other ongoing programs, conditional on security force levels, a value we call β_1^D. Paralleling the approach used for security forces, we can decompose this as:

$$\{V_{jt}, Y_{jt}\} = \beta_0^D + \beta_1^D P_{jt} + \gamma^D F_{jt} + \varphi^D (P_{jt} * D_{jt}) + \theta^D D_{jt} + \delta_j + \tau_t + \varepsilon_{jt} \quad (12)$$

Much like interaction with security forces, we can also decompose the effect of project spending (or project counts) on other outcomes of interest into a portion that is directly due to aid on violence, holding fixed an average level of security forces and the program specific spending, which we'll call *D* and the additional effect of more military forces, φ^D. Table B.12 presents the results for project presence. Overall, we find very limited evidence of any interactive effects.

Table B.12: Regressions of Program Presence on Key Outcomes including Other Program Synergy

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: OTI and NSP						
OTI	0.101	0.006	-0.009*	-0.001	-0.005	-0.000
	(0.078)	(0.009)	(0.004)	(0.004)	(0.006)	(0.000)
NSP	-0.016***	-0.000	0.001	0.000	-0.001	-0.000
	(0.004)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
OTIxNSP	0.001	0.000	-0.000	0.000*	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	21041	6604	5399	3979	3634	3573
R^2	0.438	0.346	0.415	0.526	0.290	0.793
Panel B: CERP and NSP						
CERP	0.257**	-0.005	0.004	-0.976***	0.023***	0.006*
	(0.089)	(0.005)	(0.005)	(0.188)	(0.005)	(0.003)
NSP	-0.014***	0.000	0.001	0.000	-0.001	-0.000
	(0.004)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
CERPxNSP	-0.004	0.000	0.000	.	0.002	.
	(0.003)	(0.000)	(0.000)	.	(0.001)	.
N	21041	6604	5399	3979	3634	3573
R^2	0.442	0.346	0.414	0.526	0.293	0.793
Panel C: CERP and OTI						
OTI	0.118	0.008	-0.011**	0.000	-0.002	0.000
	(0.069)	(0.008)	(0.004)	(0.003)	(0.006)	(0.000)
CERP	0.243**	-0.003	0.004	-0.979***	0.026***	0.005*
	(0.084)	(0.005)	(0.005)	(0.187)	(0.006)	(0.002)
OTIxCERP	.	.	.	.	.	.
	.	.	.	.	.	.
N	21041	6604	5399	3979	3634	3573
R^2	0.442	0.346	0.414	0.526	0.293	0.793

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $10,000 increase in cumulative spending from the previous year from the specified data source includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped.

We also estimated the relation between programs overall and MISTI districts to see if areas with additional USAID presence or spending had an amplifying effect. These results are shown in Table B.13.

Table B.13: Regressions of Program Presence on Key Outcomes in MISTI Districts

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	-0.065**	0.003	-0.002	0.003	0.001	19.282
	(0.024)	(0.005)	(0.003)	(0.004)	(0.005)	(10.336)
MISTI	-1.351	0.070	-0.041	0.012	0.037	28.360
	(0.879)	(0.056)	(0.043)	(0.014)	(0.056)	(15.119)
AIxMISTI	0.266***	-0.009	0.004	-0.004	-0.005	-22.697
	(0.061)	(0.005)	(0.003)	(0.004)	(0.005)	(12.138)
R^2	0.445	0.345	0.414	0.525	0.290	0.817
Panel B: Total (OTI, LGCD, NSP)						
TOT	-0.186***	0.005	0.003	0.006	-0.003	-18.688
	(0.045)	(0.007)	(0.006)	(0.003)	(0.009)	(11.296)
MISTI	-0.532*	0.007	-0.018	0.087	0.057	-132.900*
	(0.264)	(0.051)	(0.049)	(0.052)	(0.104)	(61.051)
TOTxMISTI	0.209**	-0.002	0.003	-0.010*	-0.008	24.296*
	(0.080)	(0.011)	(0.009)	(0.005)	(0.015)	(10.658)
R^2	0.435	0.344	0.414	0.526	0.290	0.817
Panel C: OTI						
OTI	-0.666	-0.317	-0.189*	-0.193	-0.126	-249.063*
	(0.532)	(0.277)	(0.089)	(0.157)	(0.123)	(107.052)
MISTI	-0.048	-0.027	-0.008	-0.025	-0.011	-3.211
	(0.046)	(0.029)	(0.010)	(0.028)	(0.018)	(2.638)
OTIxMISTI	1.957*	0.431	0.092	0.209	0.100	263.925*
	(0.982)	(0.297)	(0.104)	(0.163)	(0.144)	(105.499)
R^2	0.434	0.345	0.414	0.526	0.290	0.817
Panel D: LGCD						
LGCD	0.111	0.115	-0.138*	.	0.125*	.
	(0.089)	(0.066)	(0.066)	.	(0.050)	.
MISTI	-0.089	0.017	-0.000	.	0.007	0.000
	(0.136)	(0.022)	(0.000)	.	(0.028)	(0.000)
LGCDxMISTI	0.149	-0.082	0.190*	.	-0.021	.
	(0.211)	(0.082)	(0.081)	.	(0.077)	.
R^2	0.434	0.345	0.414	0.525	0.292	0.815
Panel E: NSP						
NSP	-0.212***	-0.001	0.005	0.006	-0.007	-17.397
	(0.047)	(0.007)	(0.007)	(0.003)	(0.009)	(11.351)
MISTI	-0.345	-0.005	-0.017	0.095	0.057	-124.656*
	(0.229)	(0.048)	(0.049)	(0.055)	(0.099)	(61.301)
NSPxMISTI	0.179*	0.001	0.003	-0.011*	-0.009	22.840*
	(0.090)	(0.011)	(0.009)	(0.005)	(0.015)	(10.783)
R^2	0.435	0.344	0.414	0.526	0.290	0.817
N	21306	6711	5485	4039	3690	3618

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $10,000 increase in cumulative spending from the previous year from the specified data source includes

quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped. MISTI takes the value of 1 if a MISTI program (identified in the wave 5 MISTI report) operated in that district and does not vary over time.

We anticipate that support for the government may vary significantly across regions, given variations in ethnicity, culture, geographic features, and a host of other issues. Based on existing research and currently available data, we prioritized these factors along three dimensions: percent Pashtun[52] (defined as greater than 80% Pashtun and labeled *e*), illicit agricultural intensity[53] (defined as opium intensity controlling for ethnicity *a*), and distance from major roads (defined as distance from paved highways or roads and labeled *r*).[54] When feasible, we also coded the percent of programs that overlapped geospatially in a given district-time unit. Based on the literature review in Section 2 and the initial analysis of the data, we may expand this set of location specific factors.[55] For ease of discussion, suppose that the variable $\boldsymbol{\beta_1^G}$ would include three estimates then: β_1^e, β_1^a, and β_1^r

$$\{V_{jt}, Y_{jt}\} = \beta_0^G + \boldsymbol{\beta_1^G P_{jt}} + \gamma^G F_{jt} + \delta_j + \tau_t + \varepsilon_{jt} \quad \text{for } G = e, a, r \tag{14}$$

$$\{H_{jT+1}, E_{jT+1}\} = \beta_0^G + \boldsymbol{\beta_1^{GT} P_{jT}} + \gamma^G F_{jT} + \delta_j + \tau_T + \varepsilon_{jT} \quad \text{for } G = e, a, r \tag{15}$$

The features above listed were explored and reported but also reported a host of other descriptive variables, including crop-type, geographic features, and sociodemographic characteristics. For those that show significant relationships, results will be presented as exploratory and included in the overall findings. These results are shown in Tables B.14 and B.15.

Table B.14: Regressions of Program Presence on Key Outcomes by Road Density

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	0.148***	-0.004	0.001	-0.002	-0.005	-0.000
	(0.034)	(0.003)	(0.001)	(0.002)	(0.003)	(0.000)
RD	10.200***	0.504	1.178**	0.276	-1.602*	0.049***
	(2.169)	(0.790)	(0.355)	(0.216)	(0.724)	(0.003)
AIxRD	-0.000***	0.000	-0.000*	0.000*	0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.443	0.346	0.414	0.526	0.290	0.815

[52] For evidence of the salience of intergroup bias on support for the Afghan government, see Lyall, J., et al. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." On the interaction between government-Pashtun relations and stabilization, see Jasins, A. "The Afghan Government's Relationship with the Pashtun Community and its effect on Stability; a Comparative Approach."

[53] On the importance of and variability in the connection between illicit agricultural activity and the emergence of conflict, see Chouvy, P., and Laniel, L. "Agricultural Drug Economies: Cause or Alternative to Intra-State Conflicts?"

[54] When feasible based on the data, we will assess how outcomes and program factors based on distance to military bases (e.g. FOB/COP) and district centers

[55] Some key factors we will explore: how rural the area is, distance of project from provincial or district center, distance of project from security services (national or international), whether a district services team operated there, and whether the area was a key terrain district.

Panel B: Total (OTI, LGCD, NSP)						
TOT	-0.099**	0.011	0.007	0.005	0.003	-0.001
	(0.036)	(0.008)	(0.007)	(0.004)	(0.008)	(0.001)
RD	10.184***	0.516	1.191***	0.284	-1.603*	0.046***
	(2.151)	(0.784)	(0.353)	(0.213)	(0.724)	(0.005)
TOTxRD	-0.000	-0.000	-0.000	-0.000	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.438	0.346	0.414	0.526	0.290	0.815
Panel C: OTI						
OTI	1.279	0.153	-0.124**	0.022	0.004	0.001
	(0.732)	(0.080)	(0.043)	(0.033)	(0.072)	(0.001)
RD	10.295***	0.498	1.180**	0.275	-1.606*	0.049***
	(2.192)	(0.792)	(0.353)	(0.216)	(0.723)	(0.003)
OTIxRD	-0.000	-0.000**	0.000	-0.000*	-0.000*	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.438	0.347	0.414	0.526	0.290	0.815
Panel D: LGCD						
LGCD	0.222	0.090	-0.054	.	0.127**	.
	(0.120)	(0.052)	(0.063)	.	(0.046)	.
RD	10.292***	0.494	1.178**	0.275	-1.611*	0.049***
	(2.186)	(0.787)	(0.356)	(0.216)	(0.721)	(0.003)
LGCDxRD	-0.000*	-0.000	0.000	.	-0.000	.
	(0.000)	(0.000)	(0.000)	.	(0.000)	.
R^2	0.437	0.347	0.414	0.525	0.292	0.815
Panel E: NSP						
NSP	-0.150***	0.011	0.010	0.005	-0.003	-0.001
	(0.042)	(0.008)	(0.007)	(0.004)	(0.008)	(0.001)
RD	10.120***	0.514	1.196***	0.284	-1.610*	0.046***
	(2.139)	(0.785)	(0.352)	(0.213)	(0.727)	(0.005)
NSPxRD	-0.000	-0.000***	-0.000*	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.438	0.347	0.414	0.526	0.290	0.815
N	21041	6604	5399	3979	3634	3573

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $10,000 increase in cumulative spending from the previous year from the specified data source includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped. RD is a measure of road density and does not vary over time. The total effect of spending is determined by both the coefficient for spending and its interaction with road density. For instance, the positive relationship between spending in Afghan Info and violence is reduced by road density (areas with a greater length of road per square mile).

Table B.15: Regressions of Program Presence on Key Outcomes with Pashtun Specific Estimates

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Afghan Info						
AI	-0.131*	0.004	-0.001	0.001	0.007	28.486*
	(0.056)	(0.004)	(0.003)	(0.004)	(0.006)	(11.916)
P	8.241***	0.412	0.958**	0.224	-1.300*	263.388***

	(1.761)	(0.643)	(0.289)	(0.176)	(0.589)	(43.795)
AIxP	0.381***	-0.010	0.003	-0.002	-0.017	-49.125**
	(0.090)	(0.007)	(0.004)	(0.006)	(0.010)	(17.705)
R^2	0.447	0.347	0.413	0.525	0.291	0.816
Panel B: Total (OTI, LGCD, NSP)						
TOT	-0.238***	0.003	0.012	0.003	-0.009	-28.607
	(0.049)	(0.012)	(0.007)	(0.004)	(0.009)	(16.627)
P	7.960***	0.405	0.992***	0.230	-1.324*	145.484
	(1.737)	(0.644)	(0.287)	(0.175)	(0.594)	(91.905)
TOTxP	0.276***	0.005	-0.017	-0.003	0.011	41.831
	(0.079)	(0.017)	(0.011)	(0.008)	(0.018)	(33.017)
R^2	0.438	0.346	0.414	0.525	0.290	0.816
Panel C: OTI						
OTI	-1.267	0.034	-0.076	-0.021	0.065	-139.566**
	(0.692)	(0.130)	(0.058)	(0.069)	(0.114)	(47.800)
P	8.370***	0.404	0.960***	0.223	-1.305*	248.636***
	(1.780)	(0.645)	(0.287)	(0.176)	(0.588)	(41.344)
OTIxP	3.586*	0.078	-0.046	0.037	-0.135	216.217***
	(1.450)	(0.189)	(0.099)	(0.097)	(0.173)	(61.152)
R^2	0.438	0.346	0.414	0.525	0.290	0.815
Panel D: LGCD						
LGCD	0.196	0.097	0.020	.	0.160	.
	(0.178)	(0.070)	(0.166)	.	(0.103)	.
P	8.369***	0.402	0.960**	0.223	-1.310*	249.241***
	(1.777)	(0.640)	(0.290)	(0.176)	(0.586)	(41.079)
LGCDxP	-0.004	-0.016	-0.046	.	-0.045	.
	(0.185)	(0.090)	(0.182)	.	(0.113)	.
R^2	0.437	0.347	0.413	0.525	0.292	0.815
Panel E: NSP						
NSP	-0.252***	0.003	0.012	0.003	-0.009	-26.880
	(0.049)	(0.012)	(0.007)	(0.004)	(0.009)	(16.971)
P	7.979***	0.415	0.990***	0.231	-1.313*	151.870
	(1.732)	(0.643)	(0.287)	(0.174)	(0.594)	(90.513)
NSPxP	0.225**	-0.006	-0.014	-0.003	-0.002	38.532
	(0.080)	(0.017)	(0.011)	(0.008)	(0.019)	(33.986)
R^2	0.438	0.347	0.414	0.526	0.290	0.815
N	21041	6604	5399	3979	3634	3573

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: VIO is violent incidents per 10,000 people within district. ECO is a standardized measure from VIIRS. All coefficients represent response for a $10,000 increase in cumulative spending from the previous year from the specified data source includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510). Observations for Kabul have been dropped. P is a measure of percentage Pashtun within district and does not vary over time. The total effect of spending is determined by both the coefficient for spending and its interaction with percentage of the population Pashtun. For instance, the positive relationship between spending in Afghan Info and violence is increased by percentage Pashtun.

3.2.5.6 - Programmatic Feature Effects: Similarly, we envision that program level characteristics will have differential impact on our outcomes of interest. For this reason, we include measures of USAID funding for specific projects as well as cumulative USAID funding for specific projects

or groups of projects which share similar programmatic features, namely: whether the program was officially designated a "stabilization" program; the contract type (fee for service, cost-plus, etc.), project size (where projects will be divided into quantiles and compared), type of implementing partner (contractor, NGO, etc.) and other features as determined in consultation with USAID.[56] In addition, when information is available, we will code information on delays in execution (many of which may be due to issues with vetting).

For each of these features we will construct a new project spending total which is project spending for feature *p* in district *j* at time *t*. We will also construct the sector-specific existence variable which is 1 for each programmatic feature type. So for example, if there were 3 agricultural programs one which was implemented by a contractor and 2 by local NGOs there would be an indicator that is 1 for Ag-contractor and 1 for Ag-NGO. While in another area with three programs all implemented by NGOs, there would be an indicator that is 1 for Ag-NGO while the Ag-Contractor variable would be zero.

As before, we would estimate the relationship between project spending by feature (or sector-specific project existence by feature) and the outcome of interest. In these specifications, the parameter $\boldsymbol{\beta_1^p}$ could be compared across different programmatic features to understand the differential relationship between programs of feature *p* and the outcome of interest.

$$\{V_{jt}, Y_{jt}\} = \beta_0^P + \sum_{\forall \boldsymbol{p}} \boldsymbol{\beta_1^p P_{cjt}} + \gamma^p F_{jt} + \delta_j + \tau_t + \varepsilon_{jt} \quad (16)$$

$$\{H_{jT+1}, E_{jT+1}\} = \beta_0^P + \sum_{\forall \boldsymbol{p}\boldsymbol{\beta_1^p}} \boldsymbol{P_{pjT}} + \gamma^p F_{jT} + \delta_j + \tau_T + \varepsilon_{jT} \quad (17)$$

Table B.16: Regressions of Program Existence on Key Outcomes by Key Sectors

	VIO	GOV	AGE	COM	HEA	ECO
Panel A: Total Afghan Info						
AI	0.076*	-0.006*	0.003	-0.002	-0.001	-0.000
	(0.037)	(0.003)	(0.001)	(0.001)	(0.003)	(0.000)
R^2	0.660	0.372	0.452	0.576	0.242	0.886
N	1500	1045	913	654	561	540
Panel B: Afghan Info "Stabilization Unit"						
AI	2.047	-0.020	0.031	-0.011	-0.081	.
	(1.620)	(0.044)	(0.027)	(0.015)	(0.087)	.
R^2	0.660	0.368	0.450	0.575	0.243	.
N	1500	1045	913	654	561	.
Panel C: Afghan Info programs identified as Stabilization by USAID						
AI	0.102*	-0.008***	0.003	-0.002	-0.001	-0.000*
	(0.042)	(0.002)	(0.002)	(0.001)	(0.003)	(0.000)
R^2	0.661	0.373	0.451	0.576	0.242	0.886
N	1500	1045	913	654	561	540

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[56] Some options raised by USAID in the development of this plan include whether the program had cash-for-work components, the degree to which the project is monetized (ex. vouchers that have monetary value for seeds vs. training), the extent to which goods and/or services provided new to the community of recipients or already familiar with receiving, intensity of management by USAID, the average scope/scale/timeframe of projects, the time elapsed between project identification and implementation,

Note: VIO is violent incidents per 10,000 people within district. All coefficients represent response for a $100,000 increase in spending from the specified data source. Each regression controls for Military presence and includes quarter and district fixed effects. Standard errors are clustered at the Province-Year level (N=510).

B.5 Additional Results Not Pre-Specified

The nature of the SIGACTs data allows us to conduct additional analysis on the correlations between program spending and violence. In this section, we first present a robustness check on different ways of measuring violence: raw, per capita, or a logged-measure. Second, we test whether or not the presence of the military acts as a potential mediating or moderating factor by incorporating a measure of military presence (unit-months) and an interaction term between military presence and program spending. Finally, we present analogous results for different types of violence. Program spending may have been targeted only to reduce certain types of violence – for instance, several of the governance programs were also intended to impact criminal activity. The different types of significant actions we consider are: 1) Combat Events, 2) Counter-Insurgency Operations, and 3) Criminal Events.

Considering different types of violence, we find largely similar results, with direction of the relationship being the same for almost all specifications. However, we note that per-capitized violence is likely the most appropriate measure. For these, we find no significant effects in the first differences specification.

Table B.17: Different Measures of Violence

	Raw Violence			Violence Per Capita			ln(Violence)		
	RAW	**FE**	**Diff**	**RAW**	**FE**	**Diff**	**RAW**	**FE**	**Diff**
Panel A: Afghan Info									
AI	0.0000***	0.0000***	0.0000	0.0000***	0.0000***	0.0000	0.0000***	0.0000***	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.076	0.740	0.001	0.007	0.690	0.000	0.052	0.841	0.000
N	8316	8316	7920	8316	8316	7920	8316	8316	7920
Panel B: Total (OTI, LGCD, NSP)									
TOT	0.0000***	-0.0000	0.0000	0.0000***	-0.0000***	0.0000	0.0000***	0.0000	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.005	0.433	0.000	0.000	0.438	0.000	0.028	0.729	0.001
N	21041	21041	20644	21041	21041	20644	21041	21041	20644
Panel C: OTI									
OTI	0.0005***	0.0001**	-0.0000	0.0001***	0.0000*	-0.0000	0.0000***	0.0000***	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.014	0.435	0.000	0.010	0.438	0.001	0.018	0.729	0.000
N	21041	21041	20644	21041	21041	20644	21041	21041	20644
Panel D: LGCD									
LGCD	0.0001***	0.0000**	0.0000	0.0000***	0.0000**	-0.0000	0.0000***	0.0000***	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

R^2	0.002	0.434	0.000	0.001	0.437	0.000	0.007	0.730	0.000
N	21041	21041	20644	21041	21041	20644	21041	21041	20644
Panel E: NSP									
NSP	0.0000***	-0.0000	0.0000	0.0000	-0.0000***	0.0000	0.0000***	-0.0000	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.002	0.434	0.000	0.000	0.438	0.000	0.019	0.729	0.001
N	21041	21041	20644	21041	21041	20644	21041	21041	20644

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also compare the effects of program spending on measures of violence disaggregated by type. These include Combat, Criminal, and Counter-Insurgency Operations. Again, results appear to be largely consistent across these different types of significant actions and we find relatively little relationships in the first differences specification.

Table B.18: Heterogeneous Types of Violence

	Combat			Criminal			Counter Insurgency		
	RAW	**FE**	**Diff**	**RAW**	**FE**	**Diff**	**RAW**	**FE**	**Diff**
Panel A: Afghan Info									
AI	0.0000***	0.0000***	-0.0000	0.0000***	0.0000**	-0.0000	0.0000***	0.0000**	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.081	0.693	0.000	0.094	0.514	0.000	0.057	0.337	0.001
N	8316	8316	7920	8316	8316	7920	8316	8316	7920
Panel B: Total (OTI, LGCD, NSP)									
TOT	0.0000***	0.0000	0.0000	0.0000***	0.0000	0.0000	0.0000***	-0.0000*	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.003	0.388	0.000	0.008	0.321	0.000	0.001	0.283	0.000
N	21041	21041	20644	21041	21041	20644	21041	21041	20644
Panel C: OTI									
OTI	0.0001**	0.0001	0.0000	0.0000***	0.0000	-0.0000	0.0000	-0.0000***	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.013	0.389	0.000	0.008	0.321	0.000	0.000	0.284	0.000
N	21041	21041	20644	21041	21041	20644	21041	21041	20644
Panel D: LGCD									
LGCD	0.0000***	0.0000**	-0.0000	0.0000***	0.0000**	0.0000**	0.0000***	0.0000***	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.002	0.388	0.000	0.002	0.321	0.000	0.008	0.285	0.000
N	21041	21041	20644	21041	21041	20644	21041	21041	20644
Panel E: NSP									
NSP	0.0000***	-0.0000	0.0000	0.0000***	-0.0000	0.0000	0.0000	-0.0000***	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

R^2	0.001	0.388	0.000	0.005	0.321	0.000	0.000	0.284	0.000
N	21041	21041	20644	21041	21041	20644	21041	21041	20644

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

The relationship between spending and violence remains somewhat unclear. Broadly, programs appear to spend more in areas that are violent, in line with their goals. However, there is little relationship between changes in spending and changes in violence. The reason for this, of course, might be that the impact is being lost in aggregation. Many of these programs, OTI's CCI in particular, were designed to have effects at the local level. If spending was targeted to one village within a district and violence was somewhat reduced there, but not in neighboring villages, the effects would be somewhat diminished.

Table B.19: Community Cohesion Sub-Indicators

	COM1: Local Governance			COM2: Security			COM3: Dispute Resolution		
	RAW	**FE**	**Diff**	**RAW**	**FE**	**Diff**	**RAW**	**FE**	**Diff**
Panel A: Afghan Info									
AI	-0.0000**	-0.0000*	-0.0000*	-0.0000***	-0.0000	-0.0001	-0.0000***	0.0000	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.001	0.361	0.000	0.001	0.424	0.001	0.003	0.320	0.000
N	6600	6600	5129	4271	4271	3505	4204	4204	3157
Panel B: Total (OTI, LGCD, NSP)									
TOT	0.0001*	0.0000	-0.0000	0.0001*	0.0000	-0.0001	0.0000	0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.001	0.361	0.000	0.001	0.425	0.001	0.001	0.320	0.000
N	6600	6600	5129	4271	4271	3505	4204	4204	3157
Panel C: OTI									
OTI	0.0006*	0.0004	-0.0002	-0.0000	-0.0004	-0.0005	0.0002***	0.0001	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.000	0.361	0.000	0.000	0.425	0.000	0.001	0.320	0.000
N	6600	6600	5129	4271	4271	3505	4204	4204	3157
Panel D: NSP									
NSP	0.0001*	-0.0000	-0.0000	0.0001*	0.0000	-0.0001	0.0000	0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.001	0.361	0.000	0.001	0.425	0.001	0.001	0.320	0.000
N	6600	6600	5129	4271	4271	3505	4204	4204	3157

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table B.20: Economic Indicators

	VIIRS (quarterly)			VIIRS w/ Military			DMSP (yearly)			DMSP w/ Military		
	RAW	**FE**	**Diff**	**RAW**	**FE**	**Diff**	**RAW**	**FE**	**Diff**	**RAW**	**FE**	**Diff**
Panel A: Afghan Info												
AI	-0.0000***	0.0000	0.0000	-0.0000**	0.0000	0.0000	0.0000***	0.0000*	0.0000*	0.0000***	0.0000	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
M				0.0031***	0.0014	0.0004				0.0008***	0.0006***	0.0002*
				(0.001)	(0.001)	(0.000)				(0.000)	(0.000)	(0.000)
AIxM				-0.0000**	-0.0000					-0.0000**	-0.0000**	-0.0000*
				(0.000)	(0.000)					(0.000)	(0.000)	(0.000)
R^2	0.010	0.793	0.038	0.186	0.887	0.230	0.009	0.793	0.002	0.298	0.858	0.104
N	3573	3573	3176	540	540	480	2382	2382	1985	2382	2382	1985
Panel B: Total Spending (OTI, LGCD, NSP)												
TOT	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	-0.0000	0.0000	0.0000	-0.0000	0.0000*	0.0000*	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
M				0.0038***	0.0015	0.0003				0.0008***	0.0006***	0.0002
				(0.001)	(0.001)	(0.000)				(0.000)	(0.000)	(0.000)
TOTxM				-0.0000**	-0.0000					-0.0000*	-0.0000*	-0.0000
				(0.000)	(0.000)					(0.000)	(0.000)	(0.000)
R^2	0.009	0.793	0.038	0.211	0.888	0.226	0.007	0.793	0.000	0.283	0.857	0.097
N	3573	3573	3176	540	540	480	2382	2382	1985	2382	2382	1985
Panel C: OTI												
OTI	-0.0001***	0.0000	0.0000	-0.0000*	0.0000	0.0000	0.0000	-0.0000	-0.0000***	0.0000	0.0000	-0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
M				0.0030***	0.0013	0.0004				0.0007***	0.0005***	0.0001
				(0.001)	(0.001)	(0.001)				(0.000)	(0.000)	(0.000)
OTIx.M				-0.0000**	-0.0000					-0.0000*	-0.0000	-0.0000
				(0.000)	(0.000)					(0.000)	(0.000)	(0.000)
R^2	0.010	0.793	0.038	0.177	0.886	0.227	0.007	0.793	0.001	0.272	0.853	0.089
N	3573	3573	3176	540	540	480	2382	2382	1985	2382	2382	1985
Panel D: NSP												
NSP	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0000	-0.0000*	0.0000*	0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
M				0.0032**	0.0013	0.0003				0.0007***	0.0006***	0.0001
				(0.001)	(0.001)	(0.000)				(0.000)	(0.000)	(0.000)
NSPxM				-0.0000*	-0.0000					-0.0000*	-0.0000*	-0.0000
				(0.000)	(0.000)					(0.000)	(0.000)	(0.000)
R^2	0.009	0.793	0.038	0.180	0.887	0.228	0.007	0.793	0.001	0.269	0.856	0.091
N	3573	3573	3176	540	540	480	2382	2382	1985	2382	2382	1985

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.20 presents the relationship between spending across the various projects and economic indicators using raw, fixed-effects and first-differences specifications, respectively. As economic indicators are represented by nightlights which could be heavily influenced by military presence, separate results which control and mediate military presence are also included in the table. The coefficients represent a $1,000 change in spending. The data is broken up by mode of observation - VIIRS high resolution quarterly data (recent years January 2014 to January 2016) and DMSP low resolution yearly data (early years 2008 to 2013).

From the raw regressions without controlling for military activity we consistently observe more stabilization aid spending is directed towards areas that have low economic activity. But only in the case of OTI (in more recent years) and Afghan Info (in the early years) were these findings statistically significant. During the early years we observe statistically significant military

presence in most areas of relatively high economic activity. Lastly, we examine the first difference specification where coefficients represent the relationship between changes in spending and changes in economic activity. Here we observe no statistically significant relationship between changes in economic activity and spending for recent years. In the early years we observe a positive link between higher economic activity and spending for Afghan Info, and a negative relation in OTI and NSP. These effects are sustained even after accounting for military presence and are highly statistically significant for Afghan Info and OTI, but not so for NSP. In the early years we observe a small but significant positive relationship between Afghan Info spending and economic activity, which is different from what is observed in the recent years. This trend is repeated for other programs as well, but is not significant.

Overall trends indicate that in the more recent years developmental efforts are loosely focused on areas with low economic activity which is in line with intuition. There is statistically significant military presence throughout the timeline in areas of high economic activity across all program areas.