



USAID
FROM THE AMERICAN PEOPLE

PHASE 3 EVALUATION INCEPTION REPORT

March 2013

Rule of Law Stabilization Program – Informal Component (RLS-I)
Contract Number: AID-306-C-12-00013



An RLS-I enumerator interviews a resident of Takhta Pul (Kandahar)

26 March 2013

This publication was produced for review by the United States Agency for International Development. It was prepared by Checchi and Company Consulting, Inc.



USAID
FROM THE AMERICAN PEOPLE

Phase 3 Evaluation Inception Report

March 2013

Rule of Law Stabilization Program – Informal Component

Contract Number: AID-306-C-12-00013

Submitted 26 March 2013 by:

David Greer

Acting Chief of Party

USAID/Afghanistan Rule of Law Stabilization Program – Informal Component

4th District, Ansari Square

2nd Street, House No. 149

Kabul

dgreer@chechiconsulting.com

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.



Table of Contents

| | |
|---|-----------|
| Executive Summary | 2 |
| Introduction | 4 |
| Background | 5 |
| Development problem and theory of change..... | 5 |
| Summary of hypotheses..... | 6 |
| Methodology | 8 |
| Respondent selection | 8 |
| Elder selection | 8 |
| Disputant selection..... | 8 |
| Data collection tools and indicators..... | 9 |
| Evaluation Designs | 10 |
| Phase 2 districts that continue to receive programming | 11 |
| Phase 2 comparison districts adopted as Phase 3 program districts | 12 |
| Evaluation in new Phase 3 districts..... | 13 |
| Measurement | 14 |
| Variation in treatment and group..... | 14 |
| Measurement of RLS-I impact | 15 |
| Factor analysis of RLS-I impact measures | 15 |
| Modes of inference | 16 |
| Balancing observable background characteristics | 17 |
| Using contextual variables that provide sources of random variation | 17 |
| Controlling for other factors related to program impacts..... | 18 |
| Data types and analysis | 19 |
| Pooled cross-sectional data..... | 20 |
| Matched sample (panel) data..... | 21 |
| Statistically-matched data..... | 21 |
| Simultaneous equations/multilevel modeling..... | 22 |
| Power analyses | 24 |
| Power analysis for independent samples..... | 24 |
| Power analysis for correlated samples..... | 26 |
| Conclusion | 27 |
| Products to be delivered..... | 27 |
| Annexes | 28 |
| Annex 1: Effects of explanatory variables on the RLS-I treatment effect | 28 |
| Annex 2: Informal justice supply / demand factors..... | 29 |
| Annex 3: Factor analysis of disputant assessment measures..... | 30 |
| Glossary | 33 |



Executive Summary

In the end, the measure of our success will not be predicated on the number of evaluations done, or stored within a database, or even solely upon the quality of the findings. We'll be successful if and when the evaluation work of USAID contributes to greater development effectiveness.

-USAID Administrator Rajiv Shah

How can USAID improve development effectiveness through the use of rigorous and thoughtful evaluation practice? The last several years have witnessed renewed emphasis on performance measurement and subjecting development programs to more rigorous and scientific scrutiny. USAID's 2011 evaluation policy captures this innovation, and requires, among other things, that evaluation be integrated into project design, unbiased in measurement and reporting, based on the best available methods, and oriented toward reinforcing local capacity. The evaluation policy goes on to require that evaluation designs and analysis plans be pre-specified and publicly registered.

This document presents the RLS-I evaluation design and analysis plan. There are three separate evaluation designs, two primary and one secondary data collection protocols, at least two methods for generating the counterfactual, and three approaches to seeking robust inference in identifying an RLS-I treatment effect. Salient features of the RLS-I evaluation include the following:

The RLS-I evaluation is based on observational data

As a result, attention must be paid to differences between treatment and comparison groups that differ in ways that are unrelated to RLS-I, but related to differences in evaluative measures. The possibility of not having a valid counterfactual by which to estimate a treatment effect should not be ruled out. In this case, the counterfactual will be assessed more for learning purposes, while measurements will focus more on within-treatment effects. This limitation is treated at length in the Phase 2 evaluation documents.

The primary means of correcting for non-random treatment assignment will be through the use of differencing, regression-adjustment, and propensity matching

The evaluation baseline report will empirically establish the observable differences between treatment and comparison groups. These differences will then be balanced through (a) differencing both treatment and comparison group over time, thus removing time-invariant sources of bias, (b) including variables in a regression equation that partitions the effect of each confounding variable, thus isolating the RLS-I treatment effect, and (c) selecting only those cases from the treatment and comparison groups that match on observable characteristics, and using these cases to generate the RLS-I treatment effect.

Any remaining differences between treatment and comparison groups that are unobserved cannot be adjusted for, and could bias estimates of the RLS- treatment effect. See the [Modes of inference](#) and [Data types and analysis](#) sections for detail.

Individual-level dynamics will help identify differential effects of RLS-I treatment, while contextual variables will help identify how RLS-I is received based on the enabling environment

The reception and effect of RLS-I is highly sensitive to context. For example, individual learning may be subject to peer effects or the composition of activity participants. At the level of village or district, the extent of state presence or general perceptions of state legitimacy and religiosity may affect individual learning and the traction of sustained behavior change within peer groups or larger geographic aggregations. Also see [Modes of inference](#) and [Data types and analysis](#).

Correlation in response data may weaken the ability to detect an RLS-I treatment effect

It has been observed that some informal justice providers exert influence across a given geographic area. This may result in a stronger correlation between respondents within a given sphere of influence, which in turn could complicate the effort to estimate an RLS-I treatment effect. See [Power Analyses](#) for detail.

--

Checchi's effort to evaluate the impact of RLS-I reflects ongoing direction from USAID to critically self-examine the RLS-I development hypothesis, objectives, and activity implementation as RLS-I expands into new districts. This evaluation does not formally apply the USAID evaluation policy, which requires that impact evaluations be conducted by external experts. Rather, this evaluation research is properly seen as an adjunct to the program M&E system, while also ensuring that in the event a formal impact evaluation takes place, a different evaluator using the same methods would arrive at similar findings and conclusions (see Evaluation Policy Preface). In so doing, Checchi and Company Consulting places itself at the forefront of development and evaluation practice.

Introduction

This report establishes the hypotheses, methodology, survey instruments, and analyses of the Rule of Law Stabilization – Informal Component (RLS-I) Phase 3 evaluation. It explains the mechanics of how the evaluation will build on prior tests of the RLS-I development hypotheses, and presents the specific analyses the evaluation will perform. This has the advantage of fixing the theory and its measurements prior to the actual data analysis and reporting. When an evaluation has a defined hypothesis to be tested but the data analysis is allowed to be exploratory, these additional “researcher degrees of freedom” may result in findings that are the result of a search for significance in the data rather than validating an expected program result generated by a theory of change.¹

While RLS-I Phase 1 (April 2010 – August 2011) originated as a pilot research and implementation program exploring the then nebulous concept of dispute resolution outside the state justice system, RLS-I Phase 2 (September 2011 – July 2012) focused on developing a monitoring and evaluation function and added an impact evaluation to test the RLS-I development hypothesis in anticipation of further scale-up. Data quality issues complicated the longitudinal measurements across both time and treatment and comparison groups. However, a dose-response analysis within only the treatment group showed that RLS-I program participants who attend the core program of network meetings and learning workshops are predicted to improve their knowledge of Afghan statutory law by 9% and *Shari'ah* law by 18%. Similarly, the perceptions of disputants who seek the mediation efforts of RLS-I program participants are predicted to improve 31% for procedural justice and 25% for justice of the outcome. These findings may not fully validate the RLS-I development hypothesis, but do provide clear evidence in its support.

RLS-I Phase 3 (July 2012 – January 2014) maintains the baseline research and impact monitoring function of RLS-I Phase 2, with incorporation of elements of the Phase 2 impact evaluation into the Phase 3 Performance Monitoring Plan (PMP). The RLS-I PMP lists the primary indicators of elder knowledge and disputant perception as outcome- and impact-level indicators. This inception report provides the full details of these and many other evaluation indicators, the additional variables thought to help explain the treatment effects of interest, and the different methods employed to isolate a true treatment effect from other changes in the environment that are unrelated to RLS-I.

The USAID evaluation policy (January 2011) calls for an impact evaluation of at least one large program per operating unit, as well as for any activity involving untested hypotheses that are being considered for potential scale-up. The policy defines impact evaluations as models of cause and effect requiring a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. The RLS-I Phase 2 impact evaluation, while not formally applying the USAID evaluation policy, reflected its intent.

The USAID Request for Proposals (RFP) for RLS-I Phase 3 did not explicitly call for an impact evaluation, but did refer to the new USAID evaluation policy and further stated that “the program will build on the impact evaluation data and techniques used in the previous RLS-I programs.” RLS-I

¹ A number of social science researchers have recently called attention to the problems of data mining techniques entering into research designs. See, for example, [False-Positive Psychology](#), which shows through simulation that flexibility in data collection, analysis, and reporting may dramatically increase the chance of declaring a false-positive result. To help address this, the Poverty Action Lab offers a [Hypothesis Registry](#) where randomized evaluation designs in development economics may be publicly stored and time-stamped prior to the data analysis and reporting.

Phase 3 offers the opportunity to build on lessons learned from Phase 2 while incorporating the most effective elements of the impact evaluation into the monitoring and evaluation (M&E) system for Phase 3. The scope of the evaluation is reduced from six treatment districts in Phase 2 to three treatment districts in Phase 3; however, the evaluation will still have the capability of generating robust inference that may be generalized to other Phase 3 districts. This evaluation design continues the primary design from Phase 2, and is explored under [Evaluation in new Phase 3 districts](#).

An additional focus for Phase 3 is experimentation both in evaluation design and in the mix of programmatic inputs and timelines to help determine the nature, scope, and duration of programming needed to ensure sustainability of RLS-I impacts. These “learning designs” distinguish between designs that may not be sufficient for robust inference but are still valuable for learning for development effectiveness, and more standard experimental designs providing an unbiased estimate of program impact.

Successful prosecution of the Phase 3 learning designs requires steady collaboration between program and M&E teams, and is in keeping with the Phase 2 evaluation recommendations to (a) shift some emphasis from robust inference to understanding deep context, and (b) continue to test assumptions regarding critical mass and saturation. The learning designs are covered under [Phase 2 maintenance programming](#) and [Phase 2 comparison districts adopted as Phase 3 program districts](#). [Evaluation in new Phase 3 districts](#) repeats the difference-in-differences (d-i-d) design from the Phase 2 impact evaluation.

Readers seeking additional background on RLS-I or previous evaluation efforts are referred to the Phase 2 Impact Evaluation Plan (October 2011), the Impact Evaluation Baseline Report (revised June 2012), the Impact Evaluation Final Report (August 2012), and the Phase 3 Performance Monitoring Plan (February 2013).

Background

Development problem and theory of change

Over 30 years of war has left Afghanistan’s informal and formal justice institutions weakened, limiting access to equitable justice and effective dispute resolution. Traditional dispute resolution (TDR) remains the primary forum for the public’s dispute resolution needs, with village, district, tribal, or religious elders handling most disputes, either by direct request of disputants or by referral from district authorities. State justice institutions remain weak or nonexistent in many districts, failing to offer mediation services and lacking the capacity for application or enforcement of criminal penalties. Informal justice providers, meanwhile, often rely on local customary law that is consistent with neither *Shari’ah* nor Afghan statutory law, sometimes resulting in unjust, un-Islamic, and illegal decisions. Finally, though most elders are regarded as honest and unbiased, corruption, tribal and socioeconomic discrimination, and the influence of local powerbrokers undermines confidence in local justice. These factors reduce citizen access to justice and are recognized as continuing drivers of instability.

In light of these challenges, RLS-I addresses the primary objectives of (1) strengthening TDR mechanisms, including strengthening women’s roles in TDR as disputants, witnesses, and decision makers, (2) enhancing linkages between the formal and informal justice, and (3) facilitating the

resolution of longstanding and destabilizing disputes. These objectives fall under broader USG rule of law and stabilization objectives as well as national development strategies of the Government of the Islamic Republic of Afghanistan (GIRoA).

Summary of hypotheses

The RLS-I development hypothesis is that skills- and knowledge-building of informal justice providers, combined with networking opportunities to share experience and build solidarity around improved TDR practices, increases stability through increased access to justice and citizen confidence in TDR mechanisms. This is measured through a village and district panel design in which a cross-section of elders and disputants are surveyed at program inception and again at conclusion. Impact is then defined as the difference in mean scores on various measures from baseline to endline, and between the treatment group (elders passing through the RLS-I core program) and comparison group (elders who do not pass through the RLS-I core program).²

Tentative findings from the Phase 2 impact evaluation suggest that there is not a simple relationship between improved knowledge and change in adjudication and social norms around harmful practices. Rather, in districts without a state justice presence especially, RLS-I activities may play some role in strengthening community-based governance that helps engender change in adjudication without necessarily a change in knowledge.³ Furthermore, the Phase 2 impact evaluation found that network effects were potentially strong. Regardless of knowledge, elders' attendance at RLS-I activities positively affected disputants' assessments of the process. The number of elders passing through the RLS-I core curriculum was likewise associated with both knowledge gains among elders and improvement in disputant perception of the process and outcome of informal dispute resolution. This defines a definite role for network and peer effects in program success, exactly as is supposed by the development hypothesis.

The Phase 2 impact evaluation was organized around four key hypotheses:

1. The intervention will result in TDR decisions that better reflect and/or are based in Afghan statutory law, *Shari'ah*, and human rights norms
2. The intervention will result in TDR decisions and *shural/jirga* members being perceived as more impartial
3. The intervention will result in a decrease in the number of TDR decisions that negatively impact women and children
4. The intervention will result in an increased role for women in TDR processes as disputants, witnesses or decision-makers

Assumptions underlying this theory of change include the following:

- Workshop content effectively imparts knowledge
- Participants are willing and able to change their attitudes and practices that may conflict with Afghan statutory law and *Shari'ah*

² See the Phase 2 evaluation documents for a full presentation of the evaluation design.

³ A stronger statement of this point is that knowledge is not the binding constraint to improved adjudication or social norms. In fact, critical knowledge may be known to TDR practitioners, but the community remains bound by social constraints that are not well understood.



- Participants will be able to use their new knowledge effectively in context, upon returning to their communities
- Participation will generate a critical mass of elders in a given community sufficient to effect change in adjudication reflective of Afghan statutory law, *Shari'ah* and human rights norms
- Improper influence and interference with informal dispute resolution by local power brokers will gradually lessen as a result of security and governance gains
- Threats from anti-government elements (AGE) fail to deter program participation
- The programming environment is stable enough to enable social change

The primary measurements for Hypothesis 1 were tests of program participants' knowledge of Afghan statutory law and *Shari'ah*, as presented in RLS-I learning workshops. The primary measurements for Hypothesis 2 were assessment scores from disputants who found mediation and resolution through informal justice. The primary measurements for Hypotheses 3 and 4 were attitudinal measures of elders and citizens.

For Phase 3, partly due to budget constraints and partly in response to Phase 2 evaluation findings, the evaluation will focus on Hypotheses 1 and 2. Hypotheses 3 and 4 will be measured and reported separately through the interview of RLS-I *spinsary* groups. An additional activity not captured here is the evaluation of the RLS-I outreach program. Recipients of Phase 2 outreach material in Chora district (Uruzgan province) realized a 9% knowledge gain relative to a comparison group in Shah Joy (Zabul province). Recipients in Pule Komri (Baghlan) realized a 7% knowledge gain relative to a comparison group in Aybak (Samangan). A similar evaluation will measure the effect of the RLS-I outreach campaigns in Phase 3, separate from the primary evaluation of elder knowledge and disputant perception. This survey of households will also measure attitudinal items relevant to Hypotheses 3 and 4.

In addition to the primary hypotheses, there are several secondary research questions of interest, such as the following:

- What is the requisite amount of exposure to RLS-I activities before change in behavior might be effected?
- What is the time frame governing any treatment effect, and for how long does any treatment effect persist?
- What is the requisite number of participants from a given community to effect a change in dispute adjudication and outcomes in the community as a whole?
- Do RLS-I activities for women provide an indirect means of affecting dispute prevention, adjudication, and outcomes?
- Is the distinction between real and imposed elders⁴ a meaningful one in the context of RLS-I treatment effect?

⁴ An "imposed" elder refers to elders who may occupy some official representative role in their community, but are not necessarily the most legitimate leaders in the eyes of the community. Imposed elders may be part of the government *malikan* system or members of development committees such as the Community Development Councils (CDCs).

The Phase 2 impact evaluation was able to shed light on these questions, but continued investigation is needed both to learn about the dynamics of informal dispute resolution and to establish practical benchmarks for assessment of district graduation.⁵

Methodology

Respondent selection

Elder selection

The process of mobilization in new RLS-I Phase 3 districts follows a standard approach of, first, fact-gathering about district characteristics and the collection of various lists from which to select program participants conforming to the ethnic, tribal, geographic, and population characteristics of the district. District lists include rosters of registered *malikan* (village headmen and liaison between government and the village), *mullayan* (local religious leaders), members of local development committees such as Community Development Councils (CDCs), and district-level bodies such as District Development Assemblies (DDAs) or the Independent Directorate for Local Governance (IDLG) shuras. A selection of approximately 120 male elders is culled from these lists in consultation with and support from the district government. RLS-I program staff interview the selected participants to gather background information and confirm their commitment to participate.

Once program participants have been identified, 60 elders are randomly selected for the baseline assessment. The RLS-I survey research partner is then responsible for locating the selected elders with support from RLS-I M&E staff. The survey research partner is also responsible for selecting an additional 10-20 elders through direct field work. Where elders are not identified through district lists, enumerators identify elders directly through villagers in one of two directed queries. First, enumerators may ask villagers who normally helped mediate disputes in their village. Second, enumerators may ask villagers whom they would trust to help mediate a dispute. The first query is meant to identify elders in general, while the second is meant to identify elders who are trusted by villagers, but who may not necessarily be an officially recognized, or even traditional, leader of the village.

Data from elders who are within treatment districts but not directly targeted as program participants is intended to provide an estimate of any spillover or network effects from RLS-I. See [Modes of inference and data types](#) for detail.

Disputant selection

While RLS-I targets 60-80 elders per district, the quota sample for disputants is 80-100. The majority of disputes are identified during the elder interviews. After the elders are interviewed and identify what disputes they have helped mediate in the past several months, they are then asked to refer parties to the disputes they helped mediate. Interviews of opposing parties to the same dispute are possible. As a secondary identification method, elders are asked if they could refer disputants they were aware of even if the referring elder had not played any role in mediation. An additional method

⁵ District graduation refers to the point at which an RLS-I district achieves specified program objectives and is ready to continue to pursue those objectives without further RLS-I assistance or with only limited maintenance support for a limited period of time.

of disputant identification is through random walks in the community. Enumerators query citizens whether they had recently resolved a dispute at public centers such as the mosque, bazaar, transport depot, etc. The final identification method is by snowball sampling. Once a disputant was identified and interviewed, the disputant was queried whether they in turn knew of and could refer another disputant in the village.

Data collection tools and indicators

The tools used to measure the primary hypotheses, as well as specific indicators measuring different components of the main hypotheses, are as follows:

- *Elder interview.* Key measures are individual knowledge, attitudes, and practices, with specific examples of application of training content in local dispute resolution. There are also general questions on the structure and mapping of dispute resolution in a given village and district, as well as querying for the direct experience of the respondent in resolving disputes.
- *Disputant case assessment.* The disputant case assessment tool provides details on specific cases resolved through the informal justice system and perceptual assessments of various aspects of the process of resolution and the case outcome.

Related data collection tools that will be applied separately include:

- *Citizen perception survey.* Key questions are attitudes toward informal justice and the possible identification of disputants. The primary objective of this study will be to detect any change in citizen perception as a result of critical messaging from RLS-I outreach activities. A secondary objective is to gauge citizen opinion on the role of women in TDR.
- *Spinsary group interview.* This survey queries female participants in RLS-I activities on the extent and incidence of harmful practices such as *baad* and forced marriage, the role of women in TDR, and female attitudes towards such roles. The Phase 2 impact evaluation collected such data from elders; for Phase 3, women will be the primary source of information on those aspects of informal justice that most affect them.

The following table summarizes the primary measurements for each hypothesis and for each data collection tool:

| RLS-I Impact Evaluation Indicator Summary | | |
|--|-----------------|----------------------|
| Hypothesis 1: The intervention will result in TDR decisions that better reflect Afghan statutory law, <i>Shari'ah</i> , and human rights norms | Elder interview | Disputant assessment |
| Percentage of decisions recorded | x | x |
| Percentage of decisions registered with district or other government entity | x | x |
| Percentage of elders responding change in adjudication compared to one year ago | x | |
| Knowledge increase: Constitutional law, criminal law, family law, inheritance law, property law, property deeds law | x | |
| Attitudinal change: harmful practices, role of women in <i>jirga</i> | x | |
| Number, percentage of respondents perceiving Afghan statutory law, <i>Shari'ah</i> , or customary law as source of adjudication | x | x |
| Hypothesis 2: The intervention will result in TDR decisions and <i>shural/jirga</i> members being perceived as more impartial | Elder interview | Disputant assessment |
| Extent of external influence over process, outcome | x | x |
| Number, percentage of cases where parties could exercise veto right on decision-makers | | x |
| Number, percentage of cases where bond collected | | x |
| Number, percentage of case resolutions accepted by parties | | x |
| Number, percentage of respondents disagreeing with some aspects of decision, regardless of whether they accepted | | x |
| Number, percentage respondents voicing satisfaction with process and outcome of dispute resolution | | x |
| Procedural Index | | x |
| Subversion Index | x | x |
| Justice Index ⁶ | | x |

These measurements will be combined with various background characteristics to examine the dynamics of the treatment effect. See [Simultaneous equations / multilevel modeling](#) for details.

Evaluation Designs

The Phase 2 impact evaluation took baseline and endline measurements in six new program districts and two Phase 1 districts. These were then compared to corresponding measurements in ten comparison districts. With a mix of new districts, maintenance programming⁷ in Phase 1 and Phase 2

⁶ See Measurement of RLS-I Impact below for further detail on these indices. The Phase 2 Baseline Evaluation Report also includes a complete treatment.

⁷ Maintenance programming refers to low-level implementation in a district that has already received a full RLS-I program cycle in order to consolidate gains and prepare for transition to the volunteer network of elders

districts, and a longer implementation horizon across two discrete program cycles⁸, RLS-I Phase 3 has a variety of options for applying different evaluation designs, each with specific features for constructing a counterfactual estimate of what would have happened in the absence of treatment, and each having varying degrees of validity in being able to provide robust inference as to the RLS-I treatment effect.

The following sections identify the three specific evaluation designs planned for Phase 3. As mentioned in the introduction, two of the designs are more for learning about program context, critical mass and other questions pertaining to program effectiveness, while the third design attempts to measure a program effect through the (d-i-d) design from the Phase 2 impact evaluation. In the tables that follow, “O” refers to observation (i.e., measurement), while “X” refers to the “treatment” of RLS-I programming – either the core program in the first 4-6 month program cycle or both the core program and the follow-on maintenance programming that continues for 4-6 months after the core program.

Phase 2 districts that continue to receive programming

Under Phase 3, 16 Phase 1 and Phase 2 districts will continue to receive maintenance programming. Of these districts, RLS-I has identified Chora (Uruzgan) as a special case of a Phase 2 district that will receive not only maintenance programming for the Phase 2 cohort, but will also recruit and graduate a new cohort. Chora has a combination of low assessment scores and relatively good data that make it suitable to continue a role in longitudinal measurement. Each cohort plays a unique role in the RLS-I program implementation and evaluation regime, as follows:

| Chora Phase 2 cohort | | | | |
|----------------------|-----------------------|---------------------|---------------------|-----------------------|
| Phase 2 | | | Phase 3 | |
| Jan 2012 | Jan-May 2012 | June 2012 | Jul 2012 – Apr 2013 | May 2013 |
| O | X | O | X | O |
| Baseline measurement | Core program | Endline measurement | Maintenance program | Follow-up measurement |
| Chora Phase 3 cohort | | | | |
| Nov 2012 | Dec 2012 - April 2013 | May 2012 | May – Sep 2012 | Oct 2013 |
| O | X | X | O | O |
| Baseline measurement | Core program | Midline measurement | -- | Endline measurement |

The Phase 2 cohort will continue to receive maintenance programming and a follow-up evaluation measurement. The Phase 3 cohort will be mobilized, measured at baseline, pass through the core curriculum and undergo a midline measurement. A final endline measurement for the Phase 3 cohort will take place in October 2013, and will help answer questions of consolidation of gains, persistence of treatment effect over time, and the seasonal effect of degraded security and more active Taliban presence in the spring and summer months.

While this design has the advantage of learning more about the intensity and duration of treatment effects over time, and will also provide estimates of the treatment effect upon the treated (ToT), there is no robust estimate of a counterfactual. A comparison district in Kandahar province will offer

⁸ Measurements are taken after two program cycles (the core and maintenance programs), with each cycle lasting 4-6 months. In the Phase 2 impact evaluation, measurements were taken after one program cycle lasting 2-4 months.

some opportunity for comparison of trends, but will not be sufficiently powered as to allow robust inference with any confidence.

Rather, the purpose of this design is for the comparison of two cohorts within the same district, despite their progression over different periods of time, different versions of RLS-I training materials, and an expected higher competencies of RLS-I trainers.⁹ These differences are in fact the primary measures of interest with specific regards to Chora district. Another measure of interest is whether the total number of elders reached in a district plays a cascading role in improving disputant perception, in which the effects on elders who participated in RLS-I and the effects on elders exposed to RLS-I content through the participating elders interact to reach a critical mass of social and behavioral change throughout the district. Finally, note that the Phase 3 cohort lacks a maintenance programming phase while the Phase 2 cohort does not. This is another explicit point of comparison attempting to measure the value of maintenance programming over and above graduation from the core curriculum.

The evaluation measurements for Chora in Phase 3 will be situated within the measurements from the Phase 2 impact evaluation, presented here¹⁰ for background:

| Evaluation category | Evaluation item | Baseline | Endline | Gain score |
|----------------------|--|----------|---------|---------------------|
| Elder knowledge | Afghan statutory law | 47% | 63% | +16% |
| | <i>Shari'ah</i> | 69% | 25% | -43% ¹¹ |
| Disputant perception | Procedural justice index ¹² | 4.04 | 4.28 | +0.25 |
| | Subversion of decision index ¹³ | 1.44 | 3.28 | +1.83 ¹⁴ |
| | Justice of outcome index | 4.14 | 4.38 | +0.24 |

Phase 2 comparison districts adopted as Phase 3 program districts

Of the 10 Phase 2 comparison districts, four were adopted as program districts for Phase 3. Of these four, two districts were selected to serve as treatment districts in the Phase 3 evaluation: Shahidi Hassas (Uruzgan), and Panjwayi (Kandahar). For Phase 2 comparison districts that are adopted as program districts in Phase 3, the evaluation design starts with the following timeline:

| Districts | Phase 2 | | | Phase 3 | | |
|--------------------------|----------------------|--------------|----------------------|---------------------|-------------------------|---------------------|
| | Jan 2012 | Jan-May 2012 | June 2012 | Dec '12 - April '13 | May – Sep 2012 | Oct 2013 |
| Shahidi Hassas, Panjwayi | O | | O | X | X | O |
| | Baseline measurement | -- | Baseline measurement | Core programming | Maintenance programming | Endline measurement |

⁹ As part of ongoing improvement as well as in direct response to the Phase 2 Impact Evaluation findings, RLS-I has conducted a systematic review of its training material, brought in a national consultant to advise on pedagogy, conducted new rounds of training of trainers to both existing and aspiring RLS-I trainers, and instituted a content monitoring system for all learning workshops.

¹⁰ For the sake of brevity and providing immediate context, these values present only the treatment district.

¹¹ This differential is considered to be an artifact of different data collection methodologies. See the Phase 2 evaluation for detail.

¹² Values based on a 5-point scale.

¹³ As the Subversion index is a measure of corruption affecting informal justice, negative values indicate a successful RLS-I intervention (i.e., a decrease in perceptions of corruption).

¹⁴ See footnote 11

In this design, the double baseline measurements allow each district to serve as its own counterfactual. The double baseline measurements from Phase 2 provide an estimate of district development without treatment, while the treatment and measurement from Phase 3 will provide an estimate of district development with treatment. The two district trendlines may also be compared with each other for additional context.

While this design may be used on its own, it suffers from the necessary duration between sequential measurement of the counterfactual and then measurement of the treatment effect. Any new maturation unrelated to the treatment that occurs in the interim between measurements will be missed by the evaluation. For this reason, it is also advisable to compare the treatment effect against a comparison group at the same point in time where such is feasible. A comparison district in Kandahar province will provide some basis for comparison, but will not be sufficiently powered as to allow robust inference with any confidence. The Phase 2 evaluation measurements for Shahidi Hassas and Panjwayi are as follows:

| Shahidi Hassas | Evaluation item | Baseline | Endline | Gain score |
|----------------------|------------------------|----------|---------|------------|
| Elder knowledge | Afghan statutory law | 44% | 60% | +16% |
| | Shari'ah | 71% | 42% | -28%* |
| Disputant perception | Procedural justice | 3.80 | 3.94 | +0.14 |
| | Subversion of decision | 1.53 | 1.79 | +0.26* |
| | Justice of outcome | 4.05 | 4.23 | +0.18 |
| Panjwayi | Evaluation item | Baseline | Endline | Gain score |
| Elder knowledge | Afghan statutory | 54% | 53% | -1% |
| | Shari'ah | 65% | 39% | -26%* |
| Disputant perception | Procedural justice | 4.23 | 4.21 | -0.02 |
| | Subversion of decision | 1.76 | 2.15 | +0.39* |
| | Justice of outcome | 4.38 | 4.34 | -0.04 |

*See footnote 11

Evaluation in new Phase 3 districts

Of 12-14 new Phase 3 districts, three have been selected to follow the same progression as the Phase 2 treatment districts:

| Province | District | Status | Nov-Dec 2012 | Dec '12 - April '13 | May – Sep 2013 | Nov 2013 |
|----------|---------------|------------|----------------------|---------------------|---------------------|---------------------|
| | | | O | X | X | O |
| Kandahar | Zhari | Treatment | Baseline measurement | Core program | Maintenance program | Endline measurement |
| Logar | Mohammad Agha | | | | | |
| Kunar | Chawkay | | | | | |
| Kandahar | Shah Wali Kot | Comparison | | | | |
| Kunar | Narang | | | | | |
| Logar | Khoshi | | | | | |

As with the previous design, the first treatment round consists of the core curriculum in the first program cycle (November-April). The second treatment round consists of the follow-on maintenance programming for the first tranche of Phase 3 districts during the second program cycle (June – November) designed to consolidate gains and push the district towards graduation. Given limited resources for data collection, RLS-I Phase 3 will take a measurement after both treatment rounds have been completed. This offers a picture of district progress over time and the persistence of any treatment effect, but will not measure the direct effect of the first treatment round of the core curriculum of learning workshops. Budget allowing, measurements of elder knowledge, practice, and attitude may be taken, but surveys of disputant assessment will not.

Measurement

Variation in treatment and group

The Phase 2 impact evaluation was designed to test for differences between the treatment and comparison groups. However, the analysis was extended to include two additional types of indicators to identify heterogeneous treatment effects: dose-response treatment and network effects. Dose-response analysis was conducted by using the number of activities an elder attended as the treatment indicator of interest. Analysis of network effects was made possible by using the number of elders passing through the core program as the treatment indicator of interest. In both cases, it was found that the number of activities attended and the size of the district cohort passing through the RLS-I core program were associated with improvements in both elder knowledge and disputant perception.

Highlighting the number of activities attended by an RLS-I participant also exposes whether any significant differences emerge between the regular cohort of elders, who would (ideally) attend an introductory and capstone networking meeting and six learning workshops, and the State-TDR working group, consisting of approximately 30 elders per district who also attend (with district government actors) a series of issue-based discussion sessions on topics such as formal-informal linkages, *baad*, forced marriage, and the best qualities of a *jirga* and *jirgamar*. Preliminary analysis from the Phase 2 impact evaluation suggested that membership in the State-TDR working group had no effect on elder knowledge or disputant assessment. One possible explanation is that members of the State-TDR working group tended to be of higher status in their district, more knowledgeable of Afghan statutory law and *Shari'ah*, and already well-respected in their communities. This will be tested in the Phase 3 evaluation.

There are therefore three treatment variables of interest for the Phase 3 evaluation. The first is simply the binary variable denoting membership in the treatment rather than comparison group. The second is a dose-response variable examining the relationship between the extent of exposure to RLS-I activities and change in elder knowledge or disputant assessment. The third is a variable detecting network effects – whether the participation of additional elders affects a given elder's gain in knowledge or in the assessment of disputants.¹⁵

¹⁵ The theoretical framework allowing robust inference of treatment effect includes what is called the Stable Unit Treatment Value Assumption (SUTVA), which holds that the participation of any one treated unit has no effect on the performance of any other treated unit. For RLS-I, SUTVA does not hold. The Phase 2 impact evaluation showed that there are in fact peer effects of elder participation that influence another elder's gain in knowledge or in the assessment of

The Phase 3 evaluation not only includes additional treatment variables, but also an additional comparison/treatment group designed to detect spillover effects. This spillover group is a small sample of approximately 20 elders for each treated district who are not invited to RLS-I activities. Members of the spillover group will typically reside in villages adjacent to the villages of elders participating in RLS-I, though they may also reside in the same villages as participating elders. In either case, the intent is to identify (a) whether participating elders are sharing their learning and advocating for behavioral change when they return to their home communities, and (b) whether there is any change in village-level adjudication as a result of such sharing and advocacy.

The variation in treatment variables and group membership is summarized as follows:

| Indicator | Indicator type | Group | Group type |
|---------------------------------|---|------------|---|
| Treatment | Binary value distinguishing treatment from comparison group | Comparison | To provide a counterfactual measurement of what would happen to the target group in the absence of treatment |
| # of activities attended | Dose-response value identifying the extent of participation | Treatment | Measurement of change over time among RLS-I participants attending network meetings and workshops (regular group), or discussion sessions in addition to network meetings and workshops (State-TDR working group) |
| # of elders per district cohort | A measure of network effects – whether participation of an elder's peers affects outcome measures | Spillover | Measurement of change over time of elders who are proximate to RLS-I participants, but not participants themselves |

Measurement of RLS-I impact

Factor analysis of RLS-I impact measures

The original measurements of disputant assessment of process and outcome were adapted from a methodology of measuring the costs and access to pathways of justice established by the Tilburg Institute for Interdisciplinary Studies in Civil Law and Conflict Resolution Systems ([TISCO](#)). The *TISCO Measuring Access to Justice Handbook* establishes ten dimensions of measurement according to a 5-point Likert scale capturing the extent to which the disputant believes a given statement to be true. Each dimension consists of a series of assessment items, and each dimension may be considered a pathway to justice.

For the Phase 2 impact evaluation, RLS-I adopted assessment items from the TISCO handbook, adapted existing items to better fit the environment, or created new items in keeping with the TISCO methodology but inspired by considerations specific to the local justice in Afghanistan. As detailed in the Phase 2 evaluation documents, the assessment items were organized according to four theorized indices: the process by which the dispute was resolved (procedural justice), the forms of influence-peddling or bribe solicitation either from the adjudicators themselves or from local powerbrokers advancing their own interests in the case (subversion of decision, or a generalized corruption index), the local dynamics leading to selection of a non-state forum for resolution (freedom of forum), and the overall justice of the outcome.

disputants. Failure of the SUTVA condition does not prevent robust inference as to a treatment effect, but does complicate the ability to generalize such an effect across the population of those treated. For a discussion, see Angrist, Imbens, and Rubin, *Identification of Causal Effects Using Instrumental Variables*, Journal of the American Statistical Association, 1996.

The Phase 2 impact evaluation provided the opportunity to empirically test the validity of these theorized measures through factor analysis. Factor analysis examines a given body of data and identifies specific patterns, or factors, among groups of items. The results may be used to reduce multiple variables into the identified factors, or to test whether the observed factors fit the original theorized model.

In this case, while the Phase 2 impact evaluation identified four indices to measure the RLS-I treatment effect, analysis of the patterns of responses to the assessment items identified only three factors. Response data clustered around the factors of procedural justice, subversion of decision, and justice of the outcome. The three assessment items for freedom of forum were interspersed among the factors for procedural justice and justice of the outcome, but were more closely associated with justice of the outcome. Furthermore, some items theorized as belonging to procedural justice were more closely associated with justice of the outcome.

As a result, the primary impact measurements for the Phase 3 evaluation will consist of procedural justice, subversion of decision, and justice of the outcome. The new arrangement of assessment items is as follows, with starred items denoting those that migrated from a different index from the Phase 2 evaluation.

| Procedural justice | Subversion of decision | Justice of outcome |
|--|---|---|
| Decision makers consulted all relevant parties/witnesses | Decision makers solicited payment to affect outcome of case | Agreed with decision |
| | | Dispute fully resolved |
| Disputant able to communicate feelings and opinions about case | Decision makers unwillingly influenced by outside factors | Overall process was fair* |
| | | Decision allowed reconciliation |
| Disputant able to communicate facts of case | Decision makers considered which party more powerful | If I am faced with a dispute in future, I would choose this body to resolve it* |
| Decision makers sought consensus | | My rights respected |
| | | My arguments given equal consideration with opposite party* |
| Case given due consideration by decision makers | Decision makers sought own gain in adjudicating dispute | I preferred that this body decided my case* |
| | | I submitted to the decision making authority of this body by my own will* |

It remains possible that baseline responses from the Phase 3 evaluation will offer different factors of disputant assessment items. If so, this will be discussed in the baseline report. See [Annex 3](#) for a listing of all assessment items with their associated factor scores.

Modes of inference

Given the lack of random assignment to treatment or comparison group (see [Nonrandom treatment assignment](#) for detail), it is likely that the treatment and comparison groups will differ on both observed and unobserved characteristics. If any of those characteristics help determine the RLS-I treatment effect, these determinants will be missed by the evaluation and there will be bias introduced into the estimates of program impact.

For observational studies in which treatment status is not randomized, there are several methods for producing “as-if” randomization that approximates the results of experimental data. These

include balancing observable background characteristics, controlling for factors that also affect program impact in addition to treatment, and using contextual variables that provide sources of random variation in the programming environment.

Balancing observable background characteristics

For the Phase 2 impact evaluation, the balance of background characteristics across treatment and comparison elders is presented below. Bolded values represent substantive difference between treatment and comparison.

| Elder variables | Treatment | Comparison | Standardized difference | p-value |
|--|-----------|------------|-------------------------|--------------|
| Age | 50.1 | 52.7 | 0.22 | 0.001 |
| Sources of income | 2.14 | 2.23 | 0.10 | 0.257 |
| Household assets | 5.54 | 5.43 | 0.07 | 0.476 |
| Literate | 69% | 55% | 0.29 | 0.002 |
| Level of education | 2.82 | 2.33 | 0.34 | 0.000 |
| Up to 6 th grade education | 56% | 44% | | |
| Number of outside trainings | 1.49 | 1.65 | 0.22 | 0.167 |
| Disputes mediated in last three months | 6.00 | 5.72 | 0.07 | 0.332 |
| Level of village development (1-29) | 12.05 | 10.30 | 0.38 | 0.000 |
| Transit time to district center | 26.99 | 29.16 | 0.06 | 0.535 |

There are differences between the treatment and comparison groups in terms of age (2.6 years difference), literacy (14% difference), and level of education (12% difference for up to 6th grade education). These variables should be controlled for in order to balance the treatment and comparison groups and enable a proper comparison.

For the Phase 2 impact evaluation, the balance of background characteristics across treatment and comparison disputants was as follows:

| Disputant variables | Treatment | Comparison | Standardized difference | p-value |
|---------------------------------------|-----------|------------|-------------------------|-------------|
| Age | 40.16 | 41.54 | 0.13 | .031 |
| Female | 10% | 4% | 0.24 | .000 |
| Sources of income | 2.10 | 2.10 | 0.00 | .994 |
| Assets owned | 5.13 | 5.03 | 0.06 | .340 |
| Literate | 0.44 | 0.35 | 0.19 | .001 |
| Level of education | 1.94 | 1.63 | 0.26 | .000 |
| Up to 6 th grade education | 77% | 68% | -- | -- |

There are differences between the treatment and comparison groups in terms of age (1.5 years difference), proportion of female respondents (6% difference), literacy (9% difference) and level of education (9% difference for up to 6th grade education). These variables should be controlled for in order to balance the treatment and comparison groups and enable a proper comparison.

Using contextual variables that provide sources of random variation

While balancing treatment and comparison groups on background characteristics helps ensure that the groups are sufficiently alike, the use of contextual variables may also help equalize treatment and

comparison groups based on common attitudes and beliefs relating to their broader view on GIRoA legitimacy, religiosity, and conditions on the ground. For this study, contextual variables include respondent perception of security conditions in their village, *manteqa*, and district, perceptions of trust between citizens and government officials, the extent to which GIRoA is perceived as properly Islamic, and the level of agreement that Afghan statutory law is applied first and foremost by government courts.

| Disputant variables | Treatment | Comparison | Standardized difference | p-value |
|------------------------------|-----------|------------|-------------------------|---------|
| Security | 2.99 | 2.88 | 0.18 | .003 |
| Trust | 2.99 | 2.87 | 0.13 | .031 |
| GIRoA Islamic | 3.39 | 3.80 | 0.44 | .000 |
| Primacy of government courts | 3.48 | 3.05 | 0.35 | .000 |

Clear differences emerge between treatment and comparison on the above contextual variables. Disputants from the treatment group feel safer, perceive higher trust between citizens and government, are less likely to consider GIRoA Islamic, but more likely to support the idea that government courts apply Afghan statutory law first and foremost.

While differences in background characteristics from the section immediately previous were merely demographic, differences in contextual variables suggest a more substantive difference between treatment and comparison groups that may interfere with estimating a treatment effect. One interpretation of these differences could be that elders from the treatment group are more closely linked with state justice (through, for example, the *malikan* system linking a village headman to the district government), and that citizens who are also more prone to turn to government or more formalized structures for support are bringing their disputes to these elders for mediation. Under this interpretation, the treatment and comparison groups represent different segments of rural Afghan society, and it is difficult to ascertain which group may be a better representation of Afghan society in general, or whether there is a cleavage between Afghans associated with government and those who are not. Further investigation of this issue is pending additional qualitative review of the data.¹⁶

Controlling for other factors related to program impacts

In addition to controlling for background differences between treatment and comparison groups, one may also control for variables that have a mediating effect on the treatment effect of interest. For disputants, these are variables examining the dynamics of the resolution process: the time from dispute occurrence to resolution, whether or not a bond was collected to bind the decision, other costs associated with the dispute (whether procedural as in hosting elders in one's home, or as part of the resolution), and the disputants' perception of the sources of law used to render a decision (Afghan statutory law, *Shariah*, or customary law).

¹⁶ Statistical theory identifies two primary sources of bias in the absence of random assignment of treatment: initial differences between treatment and comparison, and differential treatment effects by group due to self-selection. In the example offered above, participants who self-select into RLS-I (in the sense that they may have lobbied for selection or were chosen purposefully based on certain characteristics) may be more aligned with state structures relative to comparison elders, while the act of data collection in comparison districts would not elicit such self-selection. This could result in RLS-I having a different effect for each group and thus confounding the attempt to measure impact. For an exploration, see Winship and Morgan, *The Estimation of Causal Effects from Observational Data*, Annual Review of Sociology, Volume 25 (1999).

| Disputant variables | Treatment | Comparison | Standardized difference | p-value |
|---------------------|-----------|------------|-------------------------|-------------|
| Duration | 119.57 | 53.69 | 0.16 | .036 |
| Bond | 0.17 | 0.13 | 0.10 | .877 |
| Cost | \$1,313 | \$352 | 0.46 | .011 |
| Afghan law | 4.02 | 4.00 | 0.02 | .717 |
| Shariah law | 4.23 | 4.24 | 0.01 | .844 |
| Customary law | 4.32 | 4.20 | 0.14 | .018 |

The above table indicates that dispute duration, cost, and extent of customary jurisprudence show significant differences between treatment and comparison that should be controlled for so as to enable a proper comparison.

The more important reason for including mediating variables is that they may prove significant in helping predict the outcome variables, and therefore must be controlled for in order to generate a clean estimate of the effect of RLS-I. If such variables are (a) not controlled for, and (b) correlated with the assignment of RLS-I treatment, then the estimates of treatment effect will be biased. This *omitted variable bias* is the primary source of distortion when working with observational data. It is crucial that as many variables affecting both the assignment of treatment¹⁷ and the outcome variables be included in the analysis in order to generate valid estimates of the effect of RLS-I.

The effect of including control variables to generate cleaner estimates of program impact is illustrated below.

| Index (5-point scale) | Initial estimates | Estimates with control variables | Percentage of variation explained |
|------------------------|-------------------|----------------------------------|-----------------------------------|
| Procedural justice | -0.16 | -0.26 | 34% |
| Subversion of decision | -0.16 | 0.12 | 29% |
| Justice of the outcome | -0.23 | -0.30 | 26% |

For procedural justice and justice of the outcome, including control variables increases the magnitude of the estimate. In the case of subversion of decision, the direction of the estimate reverses from a weak negative to a weak positive effect. In these cases, the inclusion of controls helped account for extraneous variation based on dispute dynamics unrelated to RLS-I.

The Phase 2 evaluation data provides insight into what variables help predict the outcome variables. See Annex I for a list of variables that were found to be significant in predicting any of the three indices that serve as the impact measurements for RLS-I.

Data types and analysis

The core measurement of this evaluation is that of difference-in-differences. First, the baseline measurement is subtracted from the measurement at endline for the treatment group, and again for the comparison group. Then the comparison group's difference is subtracted from that of the

¹⁷ When treatment status is not randomized, there is assumed to be observed or unobserved factors of self-selection into treatment status. "Assignment of treatment" refers to that set of observed or unobserved factors presumed to exist in observational data, i.e., whatever factors affected the assignment of treatment to an individual that did not come from randomization and are therefore liable to distort the estimates of treatment effect.

treatment group, to arrive at the estimate of the treatment effect. Using the mathematical symbol denoting change in a variable, this measurement is presented as:

$$\Delta TreatmentGroup_{Post-Pre} - \Delta ComparisonGroup_{Post-Pre}$$

Mechanically, this measurement is calculated as follows:

| Impact Measure | Baseline (T) | Baseline (C) | Endline (T) | Endline (C) | Difference (T) | Difference (C) | Treatment effect |
|----------------|--------------|--------------|-------------|-------------|----------------|----------------|------------------|
| Item | A | B | C | D | C-A | D-B | (C-A) – (D-B) |

Under random assignment, the d-i-d measure should produce a consistent and unbiased estimate of the RLS-I treatment effect. With observational data, estimates could be biased due to initial differences between treatment and comparison, as well as differential treatment effects for treatment and comparison due to self-selection into one or another group. The section immediately previous introduced three strategies for generating a valid estimate of the RLS-I treatment effect in the absence of random assignment of treatment: balancing observable background characteristics, controlling for factors that also affect program impact in addition to treatment, and using contextual variables that introduce sources of random variation in the programming environment.

Each strategy has a specific method of data presentation and calculation based on three types of comparative data. Measuring change requires some level of equivalence between baseline and endline respondents. Equivalence may be found by comparing the same individuals, different individuals within groups so long as each individual meets the criteria for group membership, or individuals selected from groups on the basis of statistical similarity. In the case of the RLS-I impact evaluation, all three of these models are utilized and explored below.

Pooled cross-sectional data

The primary method of data analysis will be performed on what is called pooled cross-section data. Independent samples of program participants and disputants are taken at baseline and endline and pooled together in a single dataset, with outcome measures computed as the difference in group means between the treatment and comparison groups, and across the movement in time from baseline to endline. With pooled cross-sectional data, analysis typically takes place at the level of an entire group or sub-membership within groups. In regression format, the d-i-d measurement for pooled cross-sectional data is as follows:

$$y = \beta_0 + \delta_0 \text{endline} + \beta_1 \text{treatment} + \delta_1 \text{endline} \cdot \text{treatment} + \gamma_n X_n + u$$

In this format, δ_0 reflects the secular change over time that is unrelated to treatment, β_1 reflects the change across the treatment and comparison group at endline, and X_n represents the set of explanatory variables that help control for differences between treatment and comparison as well as mediating variables related to the outcomes of interest. The estimate of the RLS-I treatment effect is through δ_1 , the interaction of treatment and time. Any factor that remains outside the analysis is captured by the variable u for what is unobserved or not included in the analysis. If there are

unobserved factors that affect both the outcome variables and the assignment of treatment, there will be bias introduced into the estimates of the RLS-I treatment effect.

Including estimates of the spillover effect involves adding new interaction terms to the regression equation, but follows the same pattern as above. The Phase 3 evaluation baseline report will go into greater detail when presenting the data.

Matched sample (panel) data

Pooled cross-section data results from separate and independent samples at baseline and endline. As a result, there is natural variation between each sample that makes it more difficult to detect a program impact. To reduce such “noise” in the data, the same respondents can be included in both the baseline and endline sample in what is called a panel design. The baseline and endline groups are equivalent in that they are the same people, interviewed twice over time. Panel measurements provide the most power to detect a statistically significant treatment effect relative to other types of measurements.

In addition to the manual constructions above, d-i-d measurements may follow a linear regression format. This often facilitates analysis and also allows the inclusion of additional explanatory variables. The d-i-d model using the same participants at baseline and endline may be expressed in regression notation as:

$$\Delta Y_i = \beta_0 + \beta_1 \text{treatment} + \gamma_n X_n + u$$

With ΔY_i denoting the change in a respondent’s outcome score Y for a given unit of analysis i , and $\text{treatment} = 1$ if the given unit of analysis is subject to RLS-I programming, or $\text{treatment} = 0$ if the unit of analysis is part of the comparison group. Thus the outcome variable Y is the change from baseline to endline, with β_0 signifying the change score of the outcome variable in the comparison group, and β_1 the change score of the outcome variable in the treatment group. Additional explanatory variables that help control for other variables related to program impact, or correct for imbalances between treatment and comparison group characteristics, are represented by the set of γ_n coefficients on X_n variables.

For the RLS-I evaluation, a panel data set is the objective for both the treatment and comparison groups. Failure to locate or interview all baseline respondents will result in a smaller sample, but still large enough to provide robust inference. Pooled cross-section measurements will provide a corroboration check on the panel data.

Statistically-matched data

When evaluations employ observational data, propensity score matching (PSM) has demonstrated good results in approximating the results from true experimental data in which treatment status is randomized. Propensity scoring is based on the theorem that if an evaluation measure is independent of a participant’s treatment status given a set of characteristics, then those characteristics of the respondent can be used to match a treatment participant with a comparison participant and approximate random assignment of treatment status (within that set of identified characteristics). PSM is thus a method of statistical matching that produces “as if” randomization of treatment status that allows a causal interpretation of the treatment effect.

The crucial underpinning of propensity matching is that the characteristics used to generate the match are assumed to include most or all forces influencing the assignment of treatment. If there are dynamics generating treatment assignment that are not captured by the variables used to generate the propensity match, PSM will break down as a valid estimate of the treatment effect. The reader should be alert to this risk when considering any estimates of program impact based on propensity matching.

To generate PSM measurements, first the matching characteristics are chosen and each data case assigned a probability of being in the treatment group (the propensity score) based on the chosen characteristics. Each case from the treatment group is then matched to another case from the treatment group with identical or near-identical propensity scores, with the same matching process applied to cases within the comparison group. Change scores on impact evaluation measures are then computed to produce the first difference. The second difference is achieved by differencing the propensity-matched change scores across treatment and comparison, again matching by the propensity score. This measurement is most conveniently presented in regression format, as follows:

$$\Delta Y_i = \beta_0 + \beta_1 \text{treatment} + \delta_1 \text{propensity} + \gamma_n W_n + u$$

In this format, the propensity variable serves to provide the second difference in the d-i-d score. It signifies that the treatment effect is generated by holding the propensity score fixed across treatment and comparison groups. Therefore, for whatever respondent characteristics were used to generate the propensity score, all evaluation measurements are conducted with treatment and comparison respondents who match on those characteristics. In this study, the expected balancing/matching characteristics are region, age, literacy/level of education, and socio-economic status.¹⁸

The propensity-matched measurements provide a final corroboration of the pooled cross-section and matched sample measurements. The final evaluation report for Phase 3 will present all three measurements of impact. For valid estimates of the RLS-I treatment effect, there should be agreement across all three specifications.¹⁹

Simultaneous equations/multilevel modeling

A final, though speculative, type of measurement of the RLS-I treatment effect will be through the use of simultaneous equations. In this setting, the analysis seeks to partition the RLS-I treatment effect according to its theory of change:



¹⁸ Socio-economic status will be measured through the Grameen Foundation's [Progress out of Poverty Index](#), which is itself a composite of survey items from the [Afghanistan National Risk and Vulnerability Assessment \(2007-2008\)](#). The Phase 3 evaluation baseline report will include details.

¹⁹ Measurement agreement across different regression specifications is advocated by Edward Leamer as a check against "specification searching" for findings supporting a desired result. See [Let's Take the Con out of Econometrics](#), American Economic Review, March 1983. Note also that agreement across specifications is necessary but not sufficient for demonstrating a valid estimate of a treatment effect using observational data.

While previous discussion centered on the direct measurement of disputant perception, one may also examine two separate regression analyses, each with its own set of explanatory variables. The first regression examines the determinants of elder knowledge, attitude and behavior as a first-level results measurement. Then, this equation is inserted into the regression for disputant assessment as its own set of explanatory factors.

Consider the following equation attempting to estimate the RLS-I treatment effect on an elder's provision of informal dispute mediation services, as measured by the perception of the parties to such disputes:

$$\text{Adjudication} = \beta_{10} + \alpha_1 \text{ElderKnowledge} + \beta_{11} \text{CaseDynamics} + \beta_{12} \text{VillageDynamics} + \beta_{13} \text{DistrictDynamics}$$

In words, the supply of informal adjudication services is a function of the knowledge of elders, the specific dynamics of a given dispute, and background factors specific to a village, *manteqa*, or district. Note that the treatment variable is absent, as RLS-I operates directly through the mechanism of elder capacity building. Thus, the determinants of elder knowledge are estimated separately:

$$\text{ElderKnowledge} = \beta_{20} + \beta_{21} \text{Treatment} + \beta_{22} \text{ElderDynamics} + \beta_{13} \text{VillageDynamics} + \beta_{14} \text{DistrictDynamics}$$

The treatment variable is included since RLS-I provides direct capacity building to elders. Thus, the effect of RLS-I as well as other factors estimates an elder's gain in knowledge, and then the predicted values of elder knowledge are used in the adjudication regression to help estimate the downstream effect of RLS-I upon the perception of disputants who seek the mediation services of RLS-I participants.

Note further the geographic aggregations of village and district. This suggests a multilevel model of analysis, which will be included as part of the assessment of the baseline data. The Phase 2 evaluation data suggests a significant amount of variation is captured at the level of main village or *manteqa*. See [Power analysis for correlated samples](#) for additional discussion.

The following variables are organized according to how they would be used in a simultaneous equations/multilevel model of estimating the RLS-I treatment effect:

Adjudication variables

- Level, improvement in knowledge of Afghan statutory law, *Shari'ah*
- Number of disputes mediated in previous three months
- Number of successful decisions (accepted, documented, enforced)

Background variables affecting adjudication

- Age
- Education
- Positions in society (Elder, *mullah*, *malik*, *mawlawi*)
- Positions in society
- Sources of income
- Citizen-government trust
- Perception of security
- Distance to district center

Case dynamics

- Case type
- Bond
- Cost
- Duration
- Source of law

Background variables affecting knowledge, attitude, behavior

- Socioeconomic characteristics of the individual
- Socioeconomic characteristics of the village
- Time/distance to district center
- Extent of state justice presence and operation

Annex 2 provides a summary of such variables organized according to a theoretical supply and demand model of informal dispute resolution. It is a framework of supply and demand for any good (informal dispute mediation services, in the case of RLS-I) that motivates a simultaneous equations model.

Power analyses

Statistical power is a feature of study design to ensure that the results will have a reasonable chance of detecting a desired level of effect. For example, if the study targets a 10% gain in elder knowledge, power is the probability that the results of the study will in fact consider the 10% gain to be a statistically significant finding. By convention, studies usually seek to achieve not lower than 80% power by design. Under-powered studies may fail to detect an actual program effect, possibly leading to the erroneous conclusion that a program is ineffective. It is generally considered better to desist from launching a study, rather than to proceed with a study that is under-powered.

Power analysis for independent samples

The Phase 2 impact evaluation was sufficiently powered to detect a moderate change in scores from baseline to endline. Given its reduced scope and sample size relative to Phase 2, the Phase 3 evaluation will not be as strongly powered. To illustrate, the following table presents power analyses for independent samples across baseline and endline for elder knowledge scores. “Independent samples” denotes that the group of elders who participate in RLS-I are different at baseline and endline, and will therefore have a number of idiosyncratic differences in scores that are unrelated to the effect of participation. Larger sample sizes are needed to help distinguish the idiosyncratic “noise” from the “signal” of participation in RLS-I that is the treatment effect of interest. Based on a targeted sample of 60 elders per district in three treatment and three comparison districts, the total sample size is 360 (180 elders per group).

| Power analysis for elder knowledge – independent samples | | | | |
|---|----------------|-------------------|--------------------|--------------|
| Baseline | Endline | Gain score | Sample size | Power |
| 50 | 55 | 5 | 360 | 16% |
| 50 | 60 | 10 | 360 | 48% |
| 50 | 65 | 15 | 360 | 82% |



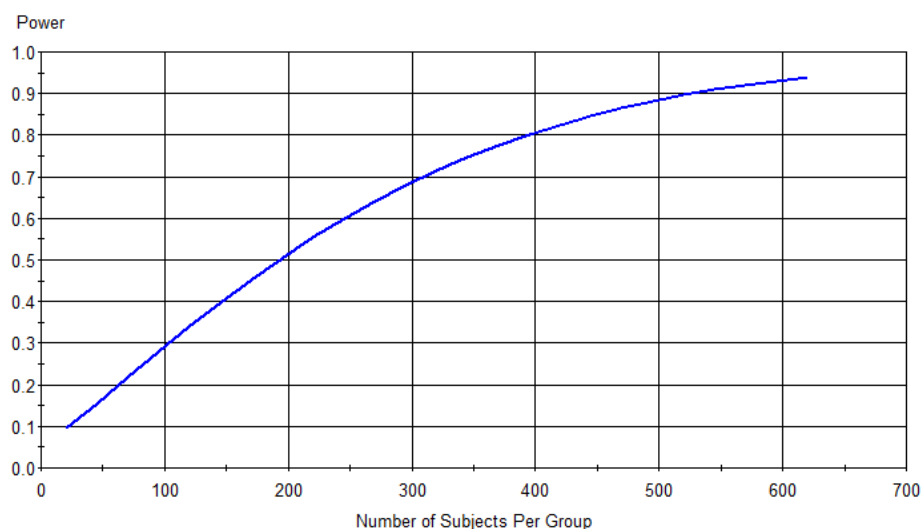
The Phase 3 evaluation study is insufficiently powered to detect a gain in elder knowledge of 5% (16% power), weakly-powered to detect a gain in elder knowledge of 10% (48% power), and well-powered to detect a gain in elder knowledge of 15% (84% power). Put differently, the Phase 3 evaluation study has a 23% chance of considering a knowledge gain score of 5% to be statistically significant, a 69% chance of considering a knowledge gain score of 10% to be statistically significant, and an 82% chance of considering a knowledge gain score of 15% to be statistically significant. Note that the Phase 3 evaluation is sufficiently powered to detect a change score of 20% - the target value in the RLS-I PMP – as statistically significant.

Power analysis for disputant assessment is more problematic. The RLS-I PMP target for disputant assessment is an improvement of 5%, or 0.25 on a 5-point scale. Based on the Phase 2 impact evaluation data, this corresponds to an effect size (change expressed in standard deviation units) of 0.17 - 0.27. The Phase 3 evaluation is sufficiently powered to detect changes at the upper end of this range, but not the lower end of the range.

| Power analysis for disputant assessment – independent samples | | | | | |
|---|---------|------------|-------------|-------------|-------|
| Baseline | Endline | Gain score | Effect size | Sample size | Power |
| 3.5 | 3.7 | 0.2 | 0.13 | 500 | 32% |
| 3.5 | 3.8 | 0.3 | 0.20 | 500 | 61% |
| 3.5 | 3.9 | 0.4 | 0.27 | 500 | 85% |

For independent samples, the Phase 3 evaluation is insufficiently powered to detect a change of 0.2 on the 5-point assessment scale, moderately powered to detect a change of 0.3 on the 5-point assessment scale, and well-powered to detect a change of 0.4 on the 5-point assessment scale. The relationship between statistical power and sample size for the gain score of 0.3 (corresponding to an effect size of 0.20) is illustrated as follows:

Power as a Function of Sample Size



Alpha = 0.050, Tails = 2, Mean 1 = 3.500, Mean 2 = 3.800, SD = 1.500

In the event that the Phase 3 evaluation yields change scores close to the target of 0.25 but the scores are not statistically significant, additional tests may be employed. For example, if a change score of 0.25 for all treatment districts were random, one would expect to see fluctuation centered

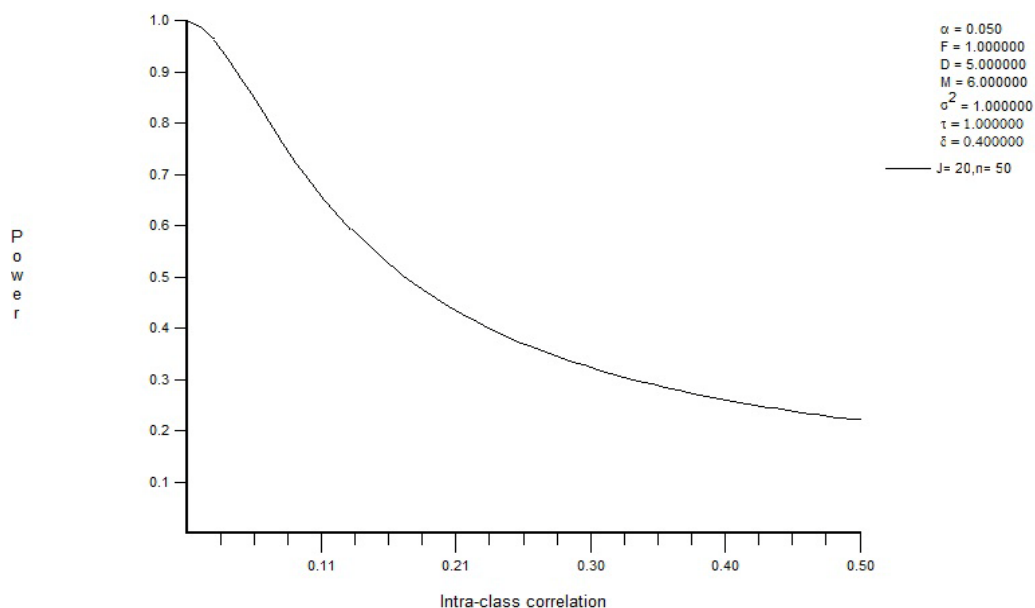
at zero for arbitrary divisions of the treatment area – for example, divisions into districts or village clusters. But if a majority of districts or village clusters showed positive gain and a minority showed zero or negative change, this could be additional evidence of an actual program result.²⁰

Power analysis for correlated samples

The previous analysis has supposed fully random samples between two independent groups. Under such conditions, study power for the Phase 3 evaluation is weaker compared to the Phase 2 evaluation, but still sufficient.

However, in reality the provision of informal justice services has displayed a degree of uniformity within the spheres of influence of more well-known elders who typically command a patronage network at the *manteqa* or district level. This implies that the assessment scores provided by disputants would reflect this uniformity, and this is in fact shown in the Phase 2 evaluation data. For example, two thirds of the variance in the procedural justice index can be explained by geographic levels of aggregation. Clustering of variance at discrete levels of analysis may be measured by the intra-class correlation (ICC). For example, responses clustered at the *manteqa* level account for 36% of the variance, corresponding to an ICC of 0.36. Clustering at the district accounts for 14% of the variance, for an ICC of 0.14. Finally, clustering at the province accounts for 16% of the variance, for an ICC of 0.16. Such homogeneity in response data drastically reduces power as the intra-class correlation increases, as shown in the following graph.

Graph: Power as a function of correlated response data (intra-class correlation)



Preliminary estimates suggest that with an ICC of 0.36 suggested by the Phase 2 evaluation data, power of the Phase 3 evaluation remains under 20% using the parameters established by the PMP target of 5% improvement in disputant assessment scores. RLS-I will follow this issue closely when it examines the baseline data set.

²⁰ The nonparametric sign test, which examines the likelihood of a series of positive or negative values deviating from chance occurrence (akin to establishing how many times a coin flip must come up heads to conclude that it is a weighted coin) could help establish whether a majority of positive values could be expected as chance occurrence.

Conclusion

This report has provided extensive review of the RLS-I Phase 3 evaluation hypotheses, design, and plans for measurement and analysis. These analyses supplement the Phase 3 PMP and, regardless of whether the analyses demonstrate an RLS-I program effect, strongly contribute to “Learning for Effectiveness”, one of the two primary objectives established in the USAID Evaluation Policy. Additionally, the exercise continues to build an impact monitoring system that can, with continued application, robustly track changes in attitudes and practices relating to informal dispute resolution and serve as a critical measure of development effectiveness and factors leading to irreversible transition to full Afghan control of its own security and governance by 2014. Finally, such a monitoring system responds directly to the stated GIRA objective of observing and documenting the practice of informal justice “so as to ensure compliance with human rights standards, Islamic values, and the law of the land.”²¹

Products to be delivered

The RLS-I Phase 3 evaluation is properly considered an adjunct to the program M&E system. In addition to the quarterly performance reporting and this inception report, RLS-I will submit baseline, midline, and final evaluation reports corresponding to the three data collection waves at baseline, midline, and endline. In addition, RLS-I may prepare secondary analyses and/or short policy briefs from the evaluation data set, time allowing or as requested by USAID. The schedule of data collection and reporting is as follows:

| Product | Description | Delivery |
|---------------------------------------|--|--|
| Performance Monitoring Plan | Establishes performance indicator definitions, targets, and plan for data collection and reporting | December 2012 |
| Evaluation Inception Report | Establishes evaluation hypotheses, methodology, and plan for analysis | February 2013 |
| Quarterly performance reports | Presentation of cumulative progress against targets and analysis of variance | Jan 31 (2013), Apr 30, July 31, Oct 31, Dec 31 |
| Evaluation Baseline Report | Presentation of baseline data | April 2013 |
| Evaluation Midline Report | Presentation of midline data – primarily knowledge tests of elders | August 2013 |
| Evaluation Final Report ²² | Presentation of evaluation measurements against baseline | December 2014 |

²¹ See page 2 of the Phase 3 PMP, or the Afghanistan National Justice Sector Strategy.

²² This report is expected to include analyses from the evaluation of the RLS-I outreach program, which is conducted separately and not discussed here. Whether the outreach evaluation is reported separately or as part of the evaluation final report, it will include a review of methodology.

Annexes

Annex I: Effects of explanatory variables on the RLS-I treatment effect

| Variable | Mean value | Procedural justice | Subversion of decision | Justice of the outcome |
|---------------------|------------|-------------------------------|-------------------------------|-------------------------------|
| | | % change at binary/mean value | % change at binary/mean value | % change at binary/mean value |
| Treatment | 0.21 | -4.7% | 5.7% | -7.9% |
| South | 0.41 | -7% | 54% | -9.3% |
| North | 0.19 | - | 49% | -2.9% |
| Female | 0.08 | -10.1% | 8.1% | -9.5% |
| Age | 41 | -4.1% | 16% | -4.1% |
| Literate | 0.38 | - | -7.3% | - |
| Level of education | 1.75 | -0.9% | 6.1% | - |
| Security | 2.96 | - | -27% | - |
| Log(Duration) | 3.81 | 1% / 1.2% | 1% / -2.1% | 1% / 1.3% |
| Log(cost amount) | 5.47 | 1% / -1.9% | 1% / 3.8% | 1% / -3% |
| Afghan law | 4.07 | +5.7% | +17.9% | - |
| Shari'ah law | 4.26 | +18% | -34% | +22% |
| Customary law | 4.26 | +18% | - | +14% |
| % Tashkil fulfilled | 0.72 | - | 9.6% | - |
| Caseload | 89 | -2% | - | -1.8% |
| District court | 0.76 | - | -11% | - |
| Village development | 11.2 | +6.7% | - | - |

Annex 2: Informal justice supply / demand factors

| Factors affecting informal dispute adjudication | | |
|---|---|--|
| Elder characteristics (supply side) | Disputant characteristics (demand side) | Situational characteristics (could affect both supply and demand) |
| Case type/subtype | Case type / subtype | War-affected Refugee/returnee population |
| Position(s) in community Education Occupation Sources of income Poverty index | Education Occupation Sources of income Poverty index | Level of economic development (village, <i>manteqa</i> , district) |
| Source of law Afghan statutory law, <i>Shari'ah</i> knowledge | Source of jurisprudence | Extent of state justice presence Extent of Taliban justice |
| Age | Age (Youth) | % youth population |
| Ethnicity | Ethnicity | Majority ethnic (in-group out-group effects on case dynamics) Degree of homogeneity/fractionalization |
| Tribe | Tribe | Majority Tribe (in-group/out-group effects on case dynamics) Degree of homogeneity/fractionalization |
| How often mediates | Who trusts to mediate a dispute | Extent of local capture of TDR (powerbrokers, imposed elders) |
| Literate | Literate | Literacy rate (village) |
| Education | Education | % secondary/elementary school educated |
| Security perceptions | Security perceptions | Incident data Local on local violence |
| | Where turn for mediation Gender | Peri-urban/rural (time/distance to district center) |
| Subversion of decision | Subversion of decision Cost/duration/bond | Influence of powerbrokers |

Annex 3: Factor analysis of disputant assessment measures

The correlations of all 18 disputant assessment items with each of the three component factors are presented here.

| Procedural justice component | Correlation with component |
|---|----------------------------|
| Decision makers consulted all relevant parties/witnesses | .695 |
| Able to communicate feelings and opinions about case | .692 |
| Able to communicate facts of case | .685 |
| Decision makers sought consensus | .675 |
| Case given due consideration by decision makers | .514 |
| Case given equal consideration compared to other side of case | .372 |
| I preferred that this body decided my case | .321 |
| My rights respected | .309 |
| I submitted to the decision making authority of this body by my own will | .284 |
| If I am faced with a dispute in future, I would choose this body to resolve it | .275 |
| Overall process was fair | .219 |
| Decision allowed reconciliation | .141 |
| Agreed with decision | .113 |
| Dispute fully resolved | .084 |
| Decision makers solicited payment to affect outcome of case | .040 |
| Decision makers unwillingly influenced by outside factors | .018 |
| Decision makers did not base deliberations on law, but also took into account which party more powerful | -.029 |
| Decision makers sought own gain in adjudicating dispute | -.100 |



| Subversion of decision component | Correlation with component |
|---|----------------------------|
| Decision makers solicited payment to affect outcome of case | .907 |
| Decision makers unwillingly influenced by outside factors | .890 |
| Decision makers did not base deliberations on law, but also took into account which party more powerful | .838 |
| Decision makers sought own gain in adjudicating dispute | .836 |
| Able to communicate facts of case | .119 |
| My rights respected | .109 |
| Case given due consideration by decision makers | -.023 |
| Dispute fully resolved | -.032 |
| Decision makers sought consensus | -.035 |
| Decision makers consulted all relevant parties/witnesses | -.043 |
| Able to communicate feelings and opinions about case | -.054 |
| Agreed with decision | -.088 |
| I submitted to the decision making authority of this body by my own will | -.100 |
| Decision allowed reconciliation | -.101 |
| I preferred that this body decided my case | -.113 |
| If I am faced with a dispute in future, I would choose this body to resolve it | -.156 |
| Overall process was fair | -.158 |
| Case given equal consideration compared to other side of case | -.164 |



| Justice of outcome component | Correlation with component |
|---|----------------------------|
| Agreed with decision | .733 |
| Dispute fully resolved | .701 |
| Overall process was fair | .624 |
| Decision allowed reconciliation | .612 |
| If I am faced with a dispute in future, I would choose this body to resolve it | .579 |
| My rights respected | .566 |
| Case given equal consideration compared to other side of case | .549 |
| I preferred that this body decided my case | .514 |
| I submitted to the decision making authority of this body by my own will | .406 |
| Case given due consideration by decision makers | .320 |
| Able to communicate facts of case | .218 |
| Decision makers sought consensus | .218 |
| Able to communicate feelings and opinions about case | .184 |
| Decision makers consulted all relevant parties/witnesses | .180 |
| Decision makers sought own gain in adjudicating dispute | -.077 |
| Decision makers solicited payment to affect outcome of case | -.079 |
| Decision makers unwillingly influenced by outside factors | -.145 |
| Decision makers did not base deliberations on law, but also took into account which party more powerful | -.157 |

Glossary

| | |
|-----------------------------------|--|
| <i>alem</i> (pl. <i>ulema</i>) | religious scholar, considered to be more knowledgeable about <i>Shari'ah</i> than most <i>mullayan</i> |
| <i>baad</i> | customary practice of resolving a dispute by giving a girl from the offender's family in marriage to a male member of the victim's family |
| <i>badal</i> | Exchange marriage performed between families or tribes to alleviate tensions or relieve the financial burden of <i>walwar</i> |
| COR | USAID/Afghanistan Contracting Officer Representative |
| CSO | civil society organization (usually but not necessarily incorporated as a legal entity) |
| DDA | District Development Assembly |
| d-i-d | Difference-in-differences. An impact evaluation measurement that includes an estimate of the counterfactual scenario of what would have happened in the absence of the USAID intervention |
| DST | District Support Team |
| GIRoA | Government of the Islamic Republic of Afghanistan |
| <i>hadith</i> | collection of scriptures detailing the actions, sayings, and tacit approvals or disapprovals of Islamic practices and beliefs of the Prophet Mohammad (PBUH), as documented by his companions and accompanied and verified by an authenticating record of the origin and lineage of each part of the collection, determining its authority as a source of Islamic law supplementing the Holy <i>Qur'an</i> |
| <i>haq-ullah</i> | a concept of <i>Shari'ah</i> that refers to the rights of society; i.e., issues that have the potential to disrupt the peace within the community and for which it is the duty of the state to issue and implement legislation (e.g., criminal law) |
| <i>haq-ul abd</i> | a concept of <i>Shari'ah</i> similar to the notion of civil law and that refers to the rights of the person; i.e., those rights that private individuals have vis-à-vis one another and that can be forfeited by the individual |
| <i>huqooq</i> | MoJ representative at the district level responsible for liaising with elders and the community to resolve civil disputes |
| IDLG | Independent Directorate of Local Governance, a sub-ministerial GIRoA body |
| <i>islah</i> | (literally, "reform") a restorative dispute resolution principle comprising the promotion of peace and social cohesion through mediation and reconciliation; in the context of registration of TDR decisions by <i>Huqooq</i> district offices, the term refers to the category in the <i>Huqooq</i> offices' record-keeping system for registering TDR decisions |
| <i>jirga</i> (pl. <i>jirgee</i>) | <i>ad hoc</i> assembly of tribal elders convened to make specific decisions or resolve a specific dispute by consensus |

| | |
|---|--|
| <i>khan</i> (pl. <i>khanan</i>) | a member of the wealthy, land-owning class, influential in the community |
| <i>machalgha</i> | a deposit required from the disputants prior to the commencement of a <i>jirga</i> to ensure compliance with its decision |
| <i>maher</i> | <i>money or goods given by a husband to a wife upon marriage and that remains the wife's property, to ensure financial security in case of divorce or the death of the husband</i> |
| <i>malik</i> (pl. <i>malikan</i>) | a tribal elder, who has been chosen as the head of the village and often liaises between the community and the government; due to this position of authority he is also approached to play a role in dispute resolution. |
| <i>manteqa</i> | an area within a district encompassing a cluster of villages that share a common characteristic such as population of the same tribal group, location within a valley, or access to a major irrigation canal. |
| <i>maraka</i> (pl. <i>marakee</i>): | Currently, often used interchangeably with the term <i>jirga</i> , especially in southern Afghanistan. Originally, used to refer to a village-level conflict resolution mechanism that included members of only one tribe or sub-tribe |
| <i>mawlawi</i> (pl. <i>mawlawiyan</i>) | highly qualified Sunni Muslim religious leader, usually with a more extensive religious education than a <i>mullah</i> |
| <i>mudir-e-huqooq</i> | <i>Huqooq</i> office director |
| <i>mullah</i> (pl. <i>mullayan</i>) | local religious leader |
| <i>nahiya</i> (pl. <i>nawahi</i>) | municipal sub-district |
| NGO | private or quasi-governmental not-for-profit organization (usually formally organized as a legal entity) |
| Platform | combined civilian-military teams at Regional Commands and PRTs that allocate resources, implement integrated programs, and assess results |
| PRT | Provincial Reconstruction Team |
| RC | Regional Command: any of the four geographic military command areas into which Afghanistan is currently divided - north (RC/N), south (RC/S), east (RC/E), and west (RC/W). The geographic areas of RC/E, RC/S, and RC/N correspond to RLS-I regions in the east, south, and north, respectively. |
| RLS-F | USAID/Afghanistan Rule of Law Stabilization Program – Formal Component |
| RLS-I | USAID/Afghanistan Rule of Law Stabilization Program – Informal Component |
| <i>Shari'ah</i> | legal precepts found in the Holy <i>Qur'an</i> and the <i>Hadith</i> ; sometimes used by non-scholars (and this report) to denote Islamic law or jurisprudence, which includes scholarly interpretations of the Holy <i>Qur'an</i> and the <i>Hadith</i> ; <i>ijma</i> ("collective reasoning" or consensus among scholars); and <i>qiyas</i> or <i>ijtihad</i> ("individual reasoning" or deduction by analogy) |

| | |
|--------------------------------------|--|
| <i>shura</i> (pl. <i>shuragani</i>) | an established council of respected community members, often registered with GIRoA, representing the interests of their community to other institutions such as GIRoA bodies and that are often involved in resolving local disputes |
| <i>spinsary</i> | (literally, feminine form of “white-headed”) respected female elder(s) involved in dispute resolution |
| TDR | traditional dispute resolution |
| ToT | Treatment on the Treated. An evaluation measurement looking only within the treatment group |
| USG | Government of the United States of America |
| <i>walwar</i> | bride price; money or goods given by a groom or his family to the head of the bride's household |