

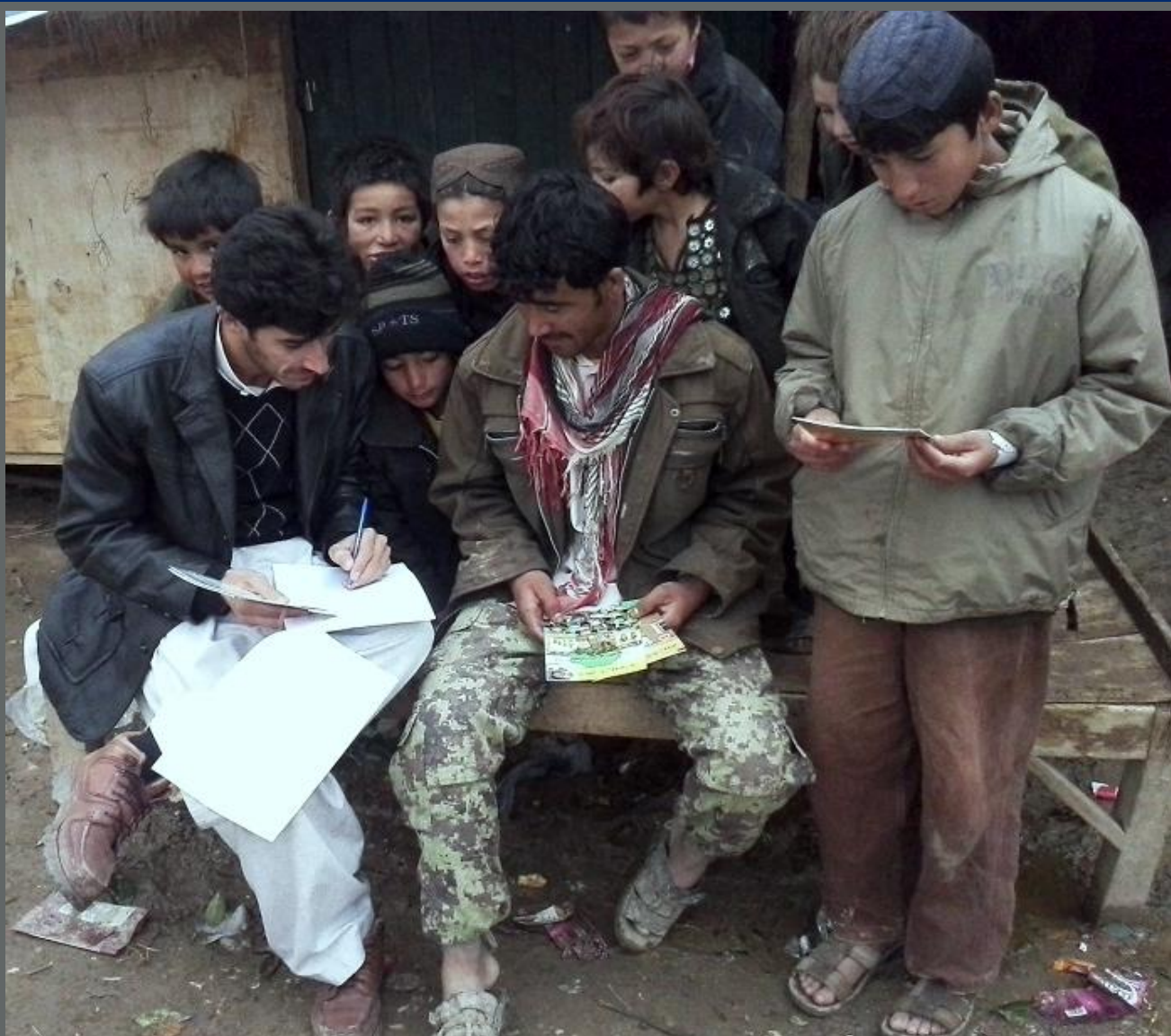


USAID
FROM THE AMERICAN PEOPLE

FINAL EVALUATION REPORT

April 2014

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



A respondent from Dahani Ghori district (Baghlan province) participates in an endline survey measuring the reach and effect of distribution of RLS-I outreach material on legal and human rights

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

Final Evaluation Report

April 2014

Rule of Law Stabilization Program – Informal Component

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

Submitted April 11, 2014 by:

Checchi and Company Consulting, Inc.

1899 L Street NW

Suite 800

Washington, DC 20036

(202) 452-9700

checchi@checchiconsulting.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.

Contents

Executive Summary	1
Introduction	7
Development problem.....	7
Summary of hypotheses.....	7
Methodology	9
Research team	9
Summary of design.....	9
Sample selection.....	10
Limitations	14
Findings.....	16
General monitoring and evaluation questions.....	16
Hypothesis 1: TDR decisions better reflect Afghan law, Islamic <i>Shari'ah</i>	18
Hypothesis 2: Disputants perceive <i>jirga</i> members as more impartial	25
Hypothesis 3: Decrease in TDR decisions harming women and children	31
Hypothesis 4: Strengthened role for women in TDR.....	34
Conclusions and lessons learned	35
Annexes	39
Annex 1: Variation in treatment and group.....	39
Annex 2: Constructing d-i-d measures	45
Annex 3: General evaluation questions, by district.....	47
Annex 4: Hypothesis 1.1 measures by district.....	48
Annex 5: Hypothesis 1.2 measures by district	48
Annex 6: Hypothesis 1.3 measures by district	48
Annex 7: Knowledge items dropped from baseline to endline	49
Annex 8: Constitutional rights knowledge items.....	50
Annex 9: Jurisdiction and criminal procedure knowledge items.....	51
Annex 10: Family law knowledge items	52
Annex 11: Inheritance knowledge items.....	53
Annex 12: Property law knowledge items	53
Annex 13: Hypothesis 1.4 measures by district.....	54
Annex 14: Disputant perception outliers and alternate results measures.....	55
Annex 15: Disputant assessment items.....	57
Annex 16: Dose-response results measures for disputant perception	58
Glossary	59

Tables

Table 1 Base specification for treatment and spillover group	10
Table 2 RLS-I Phase 3 Evaluation Districts	11
Table 3 Elder selection by district, wave	12
Table 4 Elder selection, special evaluation districts	12
Table 5 Elder selection, baseline	13
Table 6 Elder selection, endline	13
Table 7 Disputant selection by district, wave	14
Table 8 method of disputant selection	14
Table 9 General feedback on RLS-I	16
Table 10 Demand for elder services and changes in adjudication practice, D-i-D	17
Table 11 Change in elder perceptions of legal sources of dispute adjudication, D-i-D	20
Table 12 Jurisdiction, D-i-D	21
Table 13 Changes in elder reports of decision documentation and registration, D-i-D	22
Table 14 Elder knowledge gain scores	24
Table 15 Hypothesis 2 index items	27
Table 16 RLS-I impact measures, D-i-D	27
Table 17 Elder perceptions on the influence of local powerbrokers, D-i-D	28
Table 18 Outreach measures on women's role in TDR, D-i-D	32
Table 19 Consent to marriage and family disputes, D-i-D	33
Table 20 Elder reports of women's role in <i>jirga</i> , D-i-D	34
Table 21 Household reports of women's role in <i>jirga</i> , D-i-D	34
Table 22 Disputant perception by gender, D-i-D	35
Table 23 Endline spillover and participant groups by district, Elders	39
Table 24 Comparison, participant, and spillover groups by wave, Elders	40
Table 25 Endline spillover and participant groups by district, Disputants	41
Table 26 Comparison, participant and spillover groups by wave, Disputants	41
Table 27 Activities attended and RLS-I graduation by district, Elders	42
Table 28 RLS-I mediating elders identified by disputants	42
Table 29 Number of RLS-I elders meeting graduation levels of workshop attendance	43
Table 30 D-i-D measurement setup	45
Table 31 D-i-D equation manipulation	46
Table 32 Percent gain in disputant perception by number of RLS-I mediating elders, D-i-D	58
Table 33 Percent gain in disputant perception by number of "graduates", D-i-D	58
Table 34 Percent gain in disputant perception at mean value of activities attended, D-i-D	58

Figures

Figure 1 Elders reporting instance or type of positive change in adjudication practice, D-i-D	17
Figure 2 Increase in elders resolving only the civil aspects of a dispute, D-i-D	21
Figure 3 Informal decisions recorded, D-i-D	23
Figure 4 Overall knowledge gain, D-i-D	25
Figure 5 Procedural justice, D-i-D	28
Figure 6 Dose-response by number of RLS-I graduates mediating a dispute.....	30
Figure 7 Elder knowledge and disputant assessment	31
Figure 8 Base specification for treatment and spillover group	40
Figure 9 Total workshops attended across all mediating elders	43

Executive Summary

“The potential usefulness and importance of impact evaluation is well exemplified by the way impact evaluations have tested and challenged many of the key assumptions and theories of change that underpin conflict prevention and peacebuilding activities.”

- *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*
World Bank Policy Research Working Paper, 2013

This study completes over two years of research and evaluation to validate the RLS-I development hypothesis that capacity building of informal justice actors increases access to justice and strengthens stability in conflict-affected areas. To address the deficit in government justice services and reduce the incidence of social practices that Afghans themselves recognize as harmful to their communities, USAID developed the Rule of Law Stabilization Program – Informal Component (RLS-I). Implemented by Checchi and Company Consulting, RLS-I addresses the primary objectives of (1) strengthening Traditional Dispute Resolution (TDR) mechanisms, (2) enhancing linkages between the formal and informal systems, and (3) facilitating the resolution of longstanding and destabilizing disputes. These program objectives fall under broader United States Government (USG) rule of law and stabilization objectives and follow the development hypothesis that efforts at improving and strengthening TDR will increase stability in the project districts, where stability is indicated by perceptions of increased access to justice, increased confidence in TDR mechanisms, and a decrease in longstanding, potentially destabilizing disputes.

RLS-I was first piloted in 2010 in 15 key districts, and marked USAID’s initial attempt to address legal literacy and human rights concerns by targeting local mediators at the village, *manteqa*¹, and district level. RLS-I subsequently expanded to an additional 33 districts over two program phases, and added an evaluation component to test the development hypothesis in anticipation of potential scale-up. The evaluation incorporated comparison districts for counterfactual analysis, surveyed RLS-I’s key beneficiary groups (elders, disputants, *spinsaries*², and citizens) before and after programming, included a second comparison group of respondents within treatment districts who may have benefitted from interaction with program participants.

The Phase 3 evaluation covered RLS-I programming over the period of October 2012 – January 2014. The main component involved surveying 114 elders from three districts in Logar, Kunar, and Kandahar provinces who participated in the program and 208 elders from comparison districts who did not participate in the program, along with 144 disputants in districts whose cases were adjudicated by RLS-I elders and 290 disputants whose cases were adjudicated by non RLS-I elders. Qualitative research was done to understand the effects of RLS-I programming with *spinsaries*.

¹ An area larger than a village but smaller than a district, typically consisting of one or a few clusters of villages and often demarcated by natural geographic boundaries.

² Female elders who participate in training and organized into informal support groups to assist female disputants and advocate for stronger women’s role in TDR.

While the RLS-I impact evaluation was designed to provide credible estimates of program impact, there are many limitations to note. Due to the nature of the program, neither districts nor individuals could be randomized to treatment or control groups. The evaluation uses the “difference-in-difference” (D-i-D) measurement, an estimator that helps isolate the real impact of the program, so long as the treatment and control groups are sufficiently alike. Unfortunately, the treatment and comparison groups’ scores on several measures were much different than expected under a random sampling design, leaving the possibility that estimates of treatment effect are biased. A limited number of districts (due in part to security constraints) from which to choose comparable comparison groups contributed to this limitation.

The study team could only collect baseline and endline data for 12 RLS-I elders in Zhari, due to the fact that the elders changed after baseline data collection had taken place. This calls into question the validity of results for Zhari. The evaluation team found data quality to be questionable in some cases. For each indicator, the research team assessed validity based on data quality and whether the measure fit with RLS-I’s implementation experience and/or prior theory predicted by the RLS-I development hypothesis. Questions of data quality prevented the use of district-level measures. As the research partner pointed out during the study, some data quality problems may have resulted from respondent fatigue due to the length of the data collection instruments. Other problems noted included some questions or answer choices which were not sufficiently clear for respondents or interviewers. Where these were noted, they have been removed from analysis. Given a lack of randomization, the small number of districts, moderate sample sizes within each district, and data quality issues noted above, it is best to consider the results of this study as internally valid, but the reader should proceed with caution in attempting to apply any results to new contexts.

The following key findings of the Phase 3 evaluation build on Phase 2 evaluation findings covering the period of October 2011 – September 2012. There are both continuities and differences across each evaluation round, both of which provide insight into the programmatic setting and trajectory of RLS-I outcomes and impacts. The primary lessons provided by this research effort demonstrate the link between elder capacity building and improved disputant perception predicted by the RLS-I development hypothesis, show that regional differences in program setting are associated with differences in outcomes, and suggest a possible J-curve³ impact trajectory where RLS-I may raise community (evaluation respondents) expectations before program participants have fully adopted RLS-I learning content and disputants witness positive changes in elders’ adjudication practice.

RLS-I is validated by its participants

Ninety-five percent of program participants surveyed reported that RLS-I activities were of practical benefit to them. Over 80% of elders consulted RLS-I learning and outreach material at least occasionally and shared the information with someone outside of their immediate household. Over 70% of surveyed elders reported applying some aspect of RLS-I training in their home communities, and that RLS-I legal

³ For projects like RLS-I that aim to increase participation of marginalized groups, enhance women’s empowerment and improve the accessibility of formal and informal legal systems, the most likely shape of such project’s impact trajectory is a J-curve. A “J-curve” impact trajectory occurs where initial results show worse outcomes before they get better – for example, by raising citizen expectations prior to elders revising their adjudication practice, or by exposing elders to new knowledge that conflicts with what they thought they knew prior to discarding old knowledge and accepting new knowledge as their own.

training and outreach material helped them solve some problem or answer a question on legal rights and protections.

RLS-I elders self-report positive changes in dispute resolution

RLS-I elders were 30% more likely to report a positive change in how disputes were resolved in their communities compared to one year ago, and 24% more likely to successfully provide a concrete example of what positive change had occurred, relative to comparison group elders. See [general monitoring and evaluation questions](#).

RLS-I elders are in higher demand for mediation services

Afghans requested the mediation services of RLS-I elders 15% more than comparison group elders in the previous 3-6 months.

RLS-I elders recognize the jurisdiction of the state and the legally-acceptable scope for non-state dispute resolution

RLS-I elders reporting that their dispute adjudication resolved the criminal aspects of a case fell by 11%, while those reporting that their dispute adjudication resolved only the civil aspects of a case increased by 38%, relative to comparison group elders. Additionally, the proportion of elders refusing to answer queries on jurisdiction (suggesting avoidance behavior for adjudication elders knew to intrude upon the jurisdiction of the state – more typical from the South) fell by more than half – from 43% to 20%. This is further indication that elders are more willing to address the idea of accepting limits to informal dispute resolution, and by extension more willing to refer criminal disputes to the state and collaborate with formal justice actors in resolving the civil aspects of the case. [See Hypothesis 1.2](#).

RLS-I elders record their decisions in writing and register them with a government entity more often than elders who do not participate in RLS-I

RLS-I elders record their decisions and register these decisions with a government entity 26% and 13% more, respectively, than comparison group elders, who also saw increases in documentation and registration. This increase in formality of informal decisions contributes to their legality, longevity, and visibility to state actors. [See Hypothesis 1.3](#).

RLS-I elders gain and retain knowledge that is practical and relevant in helping them resolve disputes...

RLS-I elders demonstrate strong knowledge gains of 10-40% in topics such as family, inheritance, and rights of the suspect, detained, and accused. Family and inheritance topics are of immediate and practical relevance to elders in resolving disputes, while knowledge of Afghans' legal rights in criminal matters helps citizens in their interaction and discourse with district authorities.

... but RLS-I elders struggle to retain abstract legal knowledge such as *de jure* rights that may not be observed in their communities or practiced by their district government

On measures of more abstract constitutional rights such as freedom of assembly or gender equality under the law, there is a mild gain in more secure districts and more educated elders, but no gain or even a decline in such knowledge in insecure districts with less educated elders. This is best explained as a consequence of a lack of practical relevance, failure to see such rights demonstrated in their

communities or practiced by their district governments, and a more critical reception to any knowledge content relating to an Afghan's sense of identity and religiosity. See [Hypothesis 1.4](#).

The success of RLS-I is contingent upon an elder's capacity to learn

RLS-I operates at the margins of basic human capacity, with overall literacy rates at one-third to one-half for men, but much lower literacy in the South region and among women. Predictably, elders who were at least partially literate and completed primary school education fared best on RLS-I knowledge tests. An elder who scored at the 75th percentile or higher at baseline exceeded 21% gain scores at endline, while an elder who scored no higher than the 25th percentile at baseline gained 4-6% at endline. While RLS-I legal materials are already highly refined (as indicated by strong knowledge gains among more educated participants), the finding indicates the potential value of moving ahead with previously considered (but delayed due to cultural sensitivities) supplements for under-educated and illiterate participants (e.g., additional illustrated supplements, broadcasts, and community theatre) to teach and evaluate the least educated participants.

Disputants seeking the mediation of RLS-I elders are more satisfied with their dispute resolution than disputants who did not seek RLS-I elders

RLS-I elders scored 11% and 8% higher on disputant assessments of procedural fairness and justice of the outcome of their TDR processes, respectively, compared to disputant assessments of comparison group elders. See [Hypothesis 2](#).

More RLS-I elders involved in a dispute resolution means higher disputant satisfaction

For every RLS-I graduate who helps mediate a dispute, disputant satisfaction is predicted to increase 4-7% relative to disputants who seek the mediation services of elders who did not participate in RLS-I. Disputes adjudicated by three RLS-I graduates scored 15-25% higher (with some measures as high as 30%) on justice measures compared to disputes with no RLS-I graduates among the mediators. RLS-I data indicate that there are typically 7-10 elders involved in resolving a dispute, which allows for estimates of the necessary number of RLS-I graduates needed on a *jirga* to predict a given increase in access to legal and fair justice. See [Dose-response treatment variables](#).

Higher elder knowledge means higher disputant satisfaction

The 12-14% average gain in elder knowledge was associated with an increase in disputant assessment of procedural fairness of 9%, an increase of overall justice of the outcome of 6%, and a decrease in disputant assessment of undue outside influence (manipulation) of 4%. These results suggest that a primary mechanism behind access to justice gains resides in the learning content from RLS-I workshops. See [Elder knowledge and disputant assessment](#).

RLS-I raises community awareness of illegal and un-Islamic practices that harm women and girls

RLS-I included awareness raising and advocacy against harmful practices at every level of programming – from workshops for male and female elders, to outreach material to households, to radio and television programming throughout a district or region. Household surveys during Phase 2 showed a 14% increase in respondents affirming that giving away girls in marriage (*baad*) is both illegal and un-Islamic. This result was not replicated in the Phase 3 household survey, but households were 7% more likely to support

alternatives to *baad*, and 5% more likely to affirm that *baad* is not an effective solution to a dispute. See [Hypothesis 3](#).

Citizens in RLS-I districts are more likely to report a change in women’s roles as disputants, mediators, and decision makers, as well as support stronger women’s roles in general. RLS-I elders report no change in women’s roles in justice.

Households in RLS-I districts reported a 17% increase in women taking their disputes directly to a decision making body, relative to households in comparison districts. Households were also 19% more likely to support the idea of women serving as dispute resolvers, and 24% more likely to support the idea of women serving as dispute resolvers for other women. It is possible that awareness of RLS-I *spinsary* groups contributed to this last result. See [Hypothesis 4](#).

Female disputants assess RLS-I elders more critically than male disputants

Female disputants report negative assessments of procedural fairness and justice of the outcome even as males report positive assessments on the same measures. However, women also perceive lower levels of undue outside influence in dispute adjudication.

These findings generated by two years of impact measurement and experimentation represent bold new practice on the part of USAID in applying evaluation best practice, exploring new practice, and continually discovering what works and what does not – not just for development, but also for development evaluation.

Conclusions and lessons learned

This research effort has empirically demonstrated the link between elder capacity building and improved disputant perception, thereby validating a key aspect of RLS-I’s theory of change. This study has also shown improved disputant perceptions in two other ways – by measuring directly and by showing the positive impact of each additional RLS-I mediator in a dispute resolution. In all three ways, RLS-I is shown to improve disputant assessment of TDR processes and outcomes as much as 30% relative to disputants whose cases are not mediated by RLS-I elders.

On the other hand, the impact evaluation helped to identify program limitations, including negligible gains in more abstract constitutional knowledge and very modest gains by male TDR practitioners in acceptance of women’s role in dispute resolution. Rather than disengaging with negative or modest results, it is recommended that programming continue to build on success and allow time needed to consolidate gains. Detailed programmatic recommendations can be found in the RLS-I Phase 3 Final Technical Report.

The rigorous RLS-I evaluation design answers the USAID 2011 Evaluation Policy’s call to use counterfactual scenarios where possible. As an in-house evaluation initiative, it reflects the interest of RLS-I to continuously learn and improve by applying evaluation best practices, taking risks to explore new approaches, and continually discover what works and what does not in the area of informal justice. The effort was complicated by data quality issues – a nearly universal problem of surveying in the challenging context of Afghanistan. In addition, lengthy survey instruments contributed to additional data

quality problems and modest sample sizes further limited the insight the study could provide. Therefore, RLS-I offers the following recommendations for further enhancing future impact evaluation efforts:

- Expand sample size to improve confidence in the data (which may require focusing and reducing the number of study questions).
- Shortening the survey tools to be implementable within standard survey timeframes (which would free resources for expanding sample sizes).
- Conduct a thorough pilot test and allow sufficient time to review and rephrase any questions that are found to cause confusion.

Introduction

Development problem

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 are nonexistent or weak in many districts, and where present often lack the capacity for application or enforcement of criminal penalties. Informal justice providers, meanwhile, often rely on local customary law that is neither consistent with *Shari'ah* nor Afghan law, sometimes resulting in unjust, un-Islamic, and illegal decisions. Finally, though most elders are regarded as honest and unbiased, corruption, tribal and socio-economic discrimination, and the influence of local powerbrokers undermines confidence in local justice. These factors reduce citizen access to justice and are recognized as 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). These objectives fall under USAID/Afghanistan's Democracy and Governance Assistance Objective (AO), improved performance and accountability of governance; Intermediate Result (IR) 1.1, increased public confidence in the rule of law system; and Sub-IR 1.1.4, strengthened traditional dispute resolution (TDR) and justice in contested areas. The program objectives follow the development hypothesis that efforts at improving and strengthening the TDR system will increase stability in the project districts where stability is indicated by perceptions of increased access to justice, increased confidence in TDR mechanisms, and a decrease in longstanding, intractable disputes.

In addition, this program follows closely the USAID Administrator's guidance on Afghan Assistance for 2011. The guidance emphasizes that USAID's assistance should create incentives for the peaceful resolution of conflict and away from violence and insurgency. In governance, core investments include expanding access to justice and dispute resolution by working with formal and traditional entities at the provincial and district level, and supporting pluralistic and inclusive governance by state and local institutions that addresses drivers of conflict and sources of instability.

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 access to justice and confidence in TDR mechanisms for both men and women. The Phase 3 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 law, *Shari'ah*, and human rights norms
2. The intervention will result in TDR decisions and *shuraljirga* 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*
- 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 law and *Shari'ah*, as presented in RLS-I legal education 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 impact evaluation findings, the evaluation focused on Hypotheses 1 and 2. Hypotheses 3 and 4 were measured separately through the interview of RLS-I *spinsary* groups and household surveys, but data quality issues do not allow use of the data for inferential purposes. 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?

While evaluation findings do not address these questions directly, they remain relevant in the interpretation and discussion of findings that follow.

Methodology

Research team

The RLS-I research and evaluation efforts are led by the RLS-I Impact Evaluation Specialist, with ongoing collaboration and critical review and feedback from the RLS-I program advisors, senior Afghan staff, and other RLS-I senior leadership.⁵ Following an open tender and review of some 15 proposals, RLS-I subcontracted with Strategic Social (S2) to collect data and provide consulting services. Strategic Social had previous experience in conducting surveys and providing general monitoring and evaluation services throughout Afghanistan. Based on its experience with the evaluation data collection, ad hoc evaluation data collection in other RLS-I districts, and occasional monitoring of RLS-I activities, Strategic Social provided a number of valuable observations and recommendations relating to the conduct of RLS-I evaluation and its methodology.

Summary of design

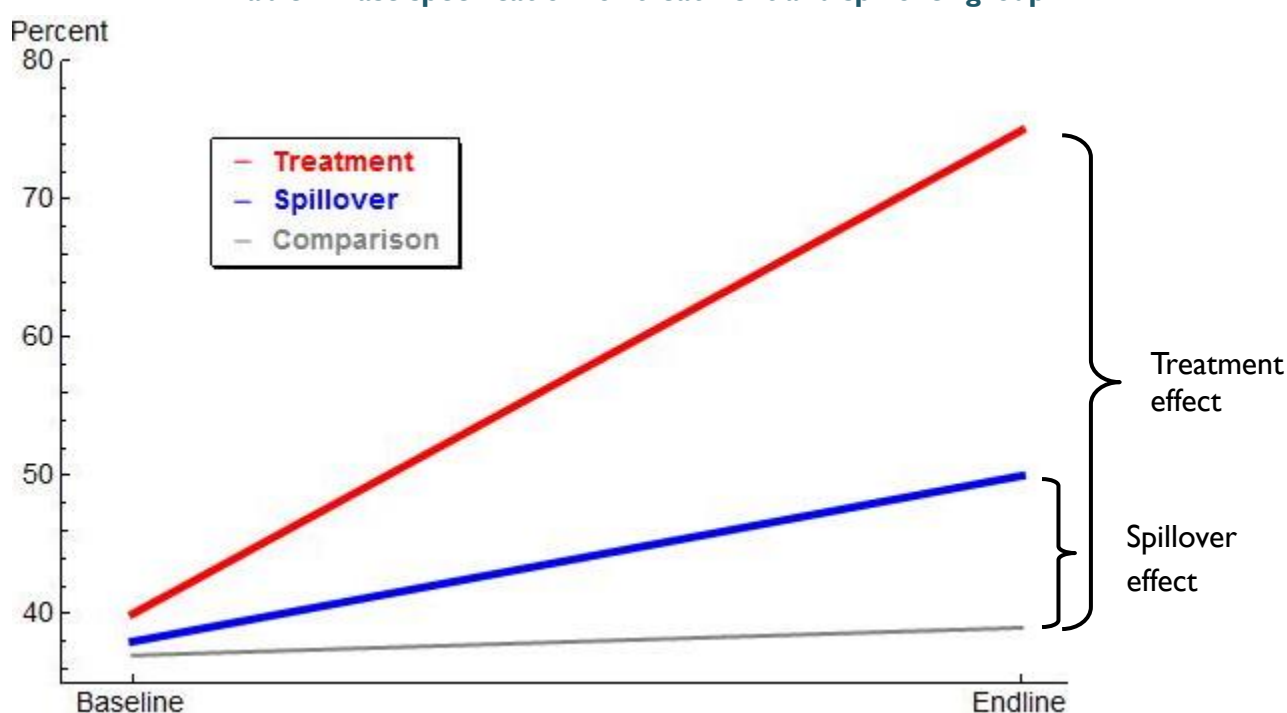
The RLS-I evaluation was designed as a quasi-experimental, mixed methods study uniting data from quantitative and qualitative data streams both to evaluate impact and to describe factors that contribute to impact. Elders and disputants were queried from both treatment districts and a sample of comparison

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 militia commanders or members of the government *malikan* system or members of development committees such as the Community Development Councils (CDCs).

⁵ RLS-I also includes a qualitative research and rapid assessment component, which is reported through district and regional assessments. These assessments are done in-house by RLS-I staff and triangulated with the research and evaluation data presented here.

districts in a longitudinal panel design. Treatment and comparison groups were compared through a “difference-in-differences” (d-i-d) design. Scores on knowledge and attitudes of informal justice providers, dispute adjudication practiced by these providers, and disputant case assessments were contrasted from baseline to endline. By including a comparison group and testing both groups before and after the intervention period, d-i-d methods helped control for unobserved characteristics that might otherwise bias outcome measurements. An additional design feature for Phase 3 was the use of a second comparison group within the treatment district. This “spillover” group was intended to show the effect of RLS-I programming upon elders who did not attend RLS-I activities but may have interacted with RLS-I participants. The base specification for RLS-I impact measurement is illustrated graphically as follows:

Table 1 Base specification for treatment and spillover group



The grey line indicates changes over time that are not related to RLS-I. The blue line represents gains among elders who experienced positive peer effects after interacting with RLS-I elders. The red line represents the direct effect of RLS-I upon its participants.

[Annex 1](#) provides a fuller discussion of evaluation measurements that include the spillover group.

Sample selection

Selection of districts

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. Insecurity in many potential comparison districts reduced the extent to which treatment (program) districts and comparison districts were comparable, as would be ideal. Under Phase 3, there were repeated measurements from Phase 2 and Phase 3 elders in Chora (Uruzgan) to investigate questions of critical mass, two comparison districts from Phase 2 became treatment districts in Phase 3 and measured again at endline, and six districts follow the standard d-i-d design from the Phase 2 impact evaluation. The identified districts, their role in the Phase 3 evaluation, and schedule of data collection is as follows.

Table 2 RLS-I Phase 3 Evaluation Districts

RLS-I Phase 3 evaluation districts						
Region	Province	District	Status	Data collection schedule		
				Baseline (Oct 2012)	Midline (April-May 2013)	Endline (Sep-Oct 2013)
South	Uruzgan	Chora	Phase 2 continued treatment	Elders and disputants	Elders and disputants	Elders and disputants
South	Uruzgan	Shahidi Hassas	Phase 2 comparison district adopted into Phase 3 treatment	-	-	Elders and disputants
South	Kandahar	Panjway	Phase 2 comparison district adopted into Phase 3 treatment	-	-	Elders and disputants
South	Kandahar	Zhari	Phase 3 treatment	Elders and disputants	-	Elders and disputants
East	Logar	Mohammad Agha	Phase 3 treatment	Elders and disputants	-	Elders and disputants
East	Kunar	Chawkay	Phase 3 treatment	Elders and disputants	-	Elders and disputants
South	Kandahar	Shah Wali Kot	Phase 3 comparison	Elders and disputants	-	Elders and disputants
East	Logar	Khoshi	Phase 3 comparison	Elders and disputants	-	Elders and disputants
East	Kunar	Narang	Phase 3 comparison	Elders and disputants	-	Elders and disputants

This study is primarily concerned with findings in new Phase 3 districts and their corresponding comparison districts. Measurements from Panjway and Shahidi Hassas provide additional context as Phase 2 comparison districts adopted Phase 3 treatment, while Chora represents a special case of an extended program cycle and additional cohort of elders. Measures from Panjway, Shahidi Hassas, and Chora will be referred to as special evaluation districts that play an occasional role in adding additional depth or context to this study.

Selection of elders

To select program participants in new Phase 3 districts RLS-I program staff gathered demographic facts about district characteristics and collected various government and NGO lists to begin identifying participants representative of the ethnic, tribal, geographic, and population characteristics of the district.

District lists included rosters of registered *maliks* (village headmen and liaisons between government and the village), *mullahs* (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*. RLS-I culled a selection of male elders from these lists in consultation with and support from the district government and identified others through village-level research, confirming each invitee's local authority as trusted and respected dispute resolver. RLS-I program staff interviewed the selected participants to gather background information and confirm their commitment to participate.

Once a representative group of approximately 120 program participants for each district had been identified, 60 elders were randomly selected for the baseline assessment.

The survey research partner was also responsible for selecting an additional 10-20 elders who were not chosen to participate in RLS-I activities but may interact with RLS-I participants. These elders comprised the spillover group intended to provide an estimate of any benefit to non-program participants in treatment districts. Where these elders were not identified through district lists, enumerators identified elders by querying villagers on who normally helps mediate disputes, and, second, by asking villagers whom they would trust to help mediate a dispute. Like the program participant selection process, the two-step process was intended to identify elders who are trusted by villagers, but who may not necessarily be an officially recognized, or even traditional, leader of the village.

For the Phase 3 impact evaluation, elder sampling by district is as follows:

Table 3 Elder selection by district, wave

Province	District	Status	Baseline	Endline	Overall
Kunar	Chawkay	Treatment	75	73	148
	Narang	Comparison	74	67	141
Logar	Mohammad Aga	Treatment	76	67	143
	Khoshi	Comparison	67	70	137
Kandahar	Zhari	Treatment	66	68	134
	Shah Wali Kot	Comparison	76	71	147
Overall			434	416	850

The table below shows the elder sampling for the special evaluation districts that draw from the Phase 2 evaluation.

Table 4 Elder selection, special evaluation districts

Province	District	Jan 2012	Jun 2012	Dec 2013	Jun 2013	Sep 2013
Kandahar	Panjway	20	21	-	-	70
	Shahidi Hassas	20	19	-	-	75
Uruzgan	Chora (Phase 2 cohort)	36	34	-	70	-
	Chora (Phase 3 cohort)	-	-	74	-	65

The following table shows the different sources for elder identification.

Table 5 Elder selection, baseline

Method of elder identification	Count	Percent
Program participant list	150	28%
District Development Assembly list	40	8%
IDLG <i>shura</i> list	89	17%
District governor identified	67	13%
District village and <i>malik</i> list	5	1%
<i>Mullah</i> list	4	1%
Who villagers trusted to resolve disputes	121	23%
Elders who normally resolved disputes	35	7%
Other / unidentified	16	3%

At endline, enumerators sampled first from program participants who were interviewed at baseline, second from a more general list of RLS-I program participants, and finally through direct field work.

Table 6 Elder selection, endline

Method of elder identification	Count	Percent
From list of elders interviewed at baseline	327	52%
From general list of RLS-I participants	150	24%
Identified in field	116	18%
Other / unidentified	41	6%

Selection of disputants

While the impact evaluation aimed to include 60-80 elders per district, the quota sample for disputants was 80-100. The majority of disputes were identified during the elder interviews. After the elders were interviewed and identified what disputes they helped mediate in the past several months, they were asked to refer parties to the disputes they helped resolve. Interviews of opposing parties to the same dispute were also possible. As a secondary identification method, elders were asked if they could refer disputants they were aware of even if the referring elder had not played any role in resolving the dispute. An additional method of purposive sampling was accomplished through random walks in the community. Enumerators queried citizens whether they had recently resolved a dispute at public centers such as the mosque, bazaar, transport depot, etc. The final identification method was by snowball sampling, in which an interviewed disputant was queried whether they in turn knew of and could refer another disputant in the village.

For the Phase 3 impact evaluation, disputant sampling by district is as follows:

Table 7 Disputant selection by district, wave

Province	District	Status	Baseline	Endline	Overall
Kunar	Chawkai	Treatment	78	70	148
	Narang	Comparison	76	110	186
Logar	Mohammad Aga	Treatment	74	90	164
	Khoshi	Comparison	75	90	165
Kandahar	Zhari	Treatment	70	90	160
	Shah Wali Kot	Comparison	83	90	173
Overall			456	540	996

The table below shows the breakdown of disputant selection methodologies:

Table 8 method of disputant selection

Method of disputant identification	Count	Percent
Elder – involved in resolving dispute	540	56%
Elder – knew of dispute, but not involved in resolution	167	17%
Purposive sampling	87	9%
Snowball sampling	110	17%

Compared to the baseline data collection, endline selection of disputants increased in purposive and snowball sampling, and decreased in the direct referrals from elders.

This section introduced the basic sampling methods used to generate results measures. [Annex 1](#) continues the methodological discussion for spillover groups and dose-response treatment variables, while [Annex 2](#) establishes the mechanics of generating impact measurements.

Limitations

While the RLS-I impact evaluation was designed to provide credible estimates of program impact, there are still threats to the validity of any claim. Due to the nature of the program, neither districts nor individuals could be randomized to treatment or control groups. Under random assignment, one can compare treatment and control groups to understand the effect of the intervention. However, in an observational study such as this one, a simple comparison of results is often misleading. This study uses the “difference-in-difference” (D-i-D) measurement, which simply subtracts the change over time in one group with the change over time of another. The D-i-D estimator can help to isolate the real impact of treatment, so long as the treatment and control groups are sufficiently alike, both in terms of how they would respond to treatment and how their scores at baseline compare.⁶ As described in the Phase 3

⁶ See [Using Randomization in Development Economics Research](#) for more discussion.

Baseline Evaluation Report, the treatment and comparison groups' scores on several measures were much different than expected under a random sampling design. If those differences were also related to the assignment of treatment (i.e. if treatment districts were systematically lower performers at baseline or vice versa) and/or how treatment affects the beneficiary (i.e. if treatment districts were inherently more receptive to RLS-I intervention than comparison districts or vice versa), then the estimates of treatment effect will be biased. The addition of control variables to account for known differences in the two groups can help improve the validity of the measure, but only as far as the variables are known to the researcher and measured during the study.

The spillover group was introduced to explore whether RLS-I gains extended to elders who were not directly involved in programming. However, the very small size of the spillover group hindered the evaluation team's ability to confidently assess the effect of programming beyond direct participants. Sample size also was a limitation in the case of Zhari district, where the study team could only collect baseline and endline data for 12 RLS-I elders due to the fact that district officials insisted upon choosing a new cohort of elders after baseline data collection had taken place. This calls into question any validity in results measures for Zhari, and is assessed on a case by case basis. For example, elder knowledge scores for Zhari are considered valid because there is complementary data from Panjway district that is similar. On the other hand, data on informal dispute documentation and registration, as well as disputant perception of undue influence in Zhari, are not considered valid given the wild swings in values and the lack of any coherent pattern in data.

The data quality issues for measurements from Zhari elders holds more generally for district-level measures. In some cases district-level measures are considered valid, while in other cases they are not. The research team assessed the validity based on data quality for any given measure, and whether the measure fit with RLS-I's implementation experience and/or prior theory predicted by the RLS-I development hypothesis. As the research partner pointed out during the study, some data quality problems may have resulted from respondent fatigue due to the length of the data collection instruments. Other problems noted included some questions or answer choices which were not sufficiently clear for respondents or interviewers. Where these were noted, they have been removed from analysis.

Given a lack of randomization, the small number of districts, moderate sample sizes within each district, and data quality issues noted above, it is best to consider the results of this study as internally valid, but the reader should proceed with caution in attempting to apply any results to new contexts.⁷ Results should generalize to the universe of RLS-I graduates, of which there were nearly 2,000 in 21 program districts, with the caveat that the south was underrepresented.

⁷ A good discussion of applying evaluation results to new contexts is found in [Evidence-Based Policy](#).

Findings

General monitoring and evaluation questions

RLS-I is validated by its participants

Participant satisfaction with RLS-I

The endline data collection asked a series of general evaluation questions as to RLS-I's usefulness and how it is applied in Afghan communities. The findings indicate that RLS-I participants overwhelmingly found the program valuable, applying RLS-I learning in their home communities, sharing what they learned with others, and solving concrete problems or answering specific problems using RLS-I material and teachings. The following table summarizes:

Table 9 General feedback on RLS-I

Feedback item	Overall
RLS-I activities useful	95%
Applied RLS-I training in home community	73%
Received handouts at RLS-I workshops	93%
Handouts useful	96%
Consult handouts at least occasionally	82%
Handouts helped answer a question or solve a legal problem	71%
Shared handouts outside of immediate household	84%
Would continue to attend RLS-I activities without external support	48%

One should not take these overwhelmingly positive statistics at face value, as they will include a healthy dose of acquiescence bias from participants interested to have RLS-I continue in their communities regardless of its overall usefulness. However, it remains evident that RLS-I is quite well-received and helpful to Afghan communities. Participants at RLS-I workshops continue to use written materials from RLS-I workshops, share those materials outside their home, and are able to answer questions or solve legal problems through the materials. Many participants also find RLS-I activities useful enough that they would attend them without external support, and inability or unwillingness to do so appears driven by the security environment. Elder willingness to attend activities even without external support has been attested to anecdotally throughout Phase 3, with more reports coming from the East region and Uruzgan province.⁸

[Annex 3](#) disaggregates participant feedback by district.

⁸ RLS-I targets its workshops to a cohort of 120 of males and 80-120 females. However, other elders or *spinsaries* frequently show up to RLS-I workshops and are content to observe the workshop even when RLS-I cannot provide lunch or transportation due to planning and budget constraints.

Elder self-reports on behavior change and disputant demand

RLS-I elders self-report changes in behavior and community practice and higher demand for their mediation services

Elders were queried about their own dispute resolution practices and whether anything was different about their own practices or dispute resolution in general in their communities, compared to 6-12 months ago. Elders who answered that practices had changed were then asked to enunciate a specific instance or type of change that occurred. Results are as follows:

Table 10 Demand for elder services and changes in adjudication practice, D-i-D

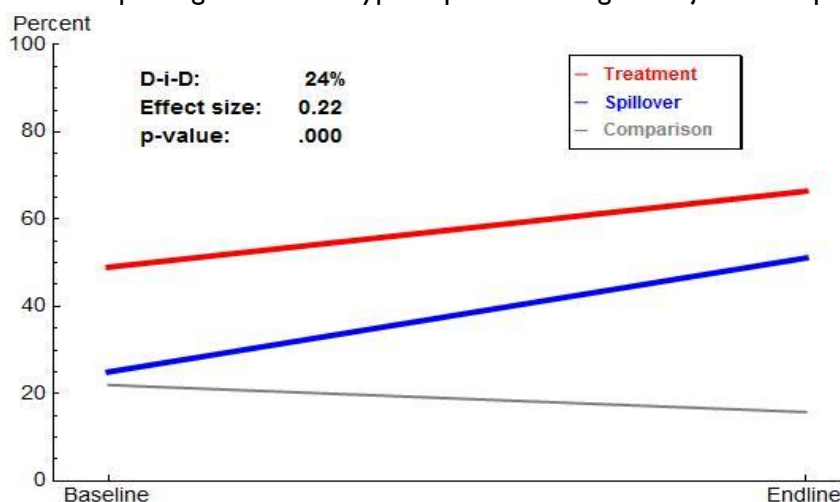
Measure	Treatment gain	Comparison gain	Spillover gain	D-i-D	Spillover D-i-D*
Was asked to mediate five or more disputes in the previous 3-6 months	20%	5.6%	24%	15%	18%
Dispute resolution practice had changed in previous 6-12 months	29%	-1%	26%	30%	27%
If change reported, could enunciate specific instance or type of change that occurred	17%	-6.2%	26%	24%	32%

*Spillover D-i-D represents the difference in change over time between the comparison and spillover groups.

The increase in reports of changed adjudication practice was 30% higher among RLS-I elders. As expected, requiring that the elders successfully enunciate a specific instance or type of change in their communities decreased scores across the board, but more so among RLS-I elders (-12%) than comparison group elders (-6%). This differential of 6% may be considered an estimate of the social desirability bias of RLS-I elders eager to report positive changes in order to validate their own participation and maintain the possibility of continued programming. The result using validated data is shown graphically below.

At the same time, elder reports indicate that disputants more often requested the mediation services of RLS-I elders relative to comparison group elders.

Figure 1 Elders reporting instance or type of positive change in adjudication practice, D-i-D



Hypothesis I: TDR decisions better reflect Afghan law, Islamic Shari’ah

“Those people who attended the [RLS-I] workshops also attend the jirgas with us. We share our ideas for resolving the dispute, but the people who attended the workshops always know better than us.”

- Elder from spillover group in treatment district of Chawkai (Kunar province)

Hypothesis I pertains most immediately to the primary RLS-I objective of strengthening informal justice in Afghanistan through direct capacity building of elders. The evaluation Inception Report identifies the specific measurements by hypothesis. Those measurements are reproduced here as sub-hypotheses:

Hypothesis I	The intervention will result in TDR decisions that better reflect Afghan statutory law, Islamic <i>Shari’ah</i> , and human rights norms
Hypothesis I.1	The intervention will result in elders who perceive their decision-making as having firmer grounding in Afghan law and Islamic <i>Shari’ah</i> , and less grounding in local customary law
Hypothesis I.2	The intervention will result in an increase in elders perceiving their decision making to be resolving the civil aspects of a case, and a decrease in perceiving decisions as resolving criminal aspects of a case
Hypothesis I.3	The intervention will result in an increase in legality and enforceability of informal decisions through documentation and registration with a government entity
Hypothesis I.4	The intervention will result in an increase in elder knowledge of Afghan law and Islamic <i>Shari’ah</i> ⁹

Evaluation results are positive on all sub-hypotheses, with three of four sub-hypotheses considered validated and the fourth showing variability too excessive to be considered robust. RLS-I elders admit that community norms may sometimes conflict with Afghan law or Islamic *Shari’ah*, recognize the lines of legal demarcation between formal and informal justice actors, record and register informal decisions more often, and retain knowledge that is relevant and practical to their daily lives.

Elders perceive their decision-making as grounded in Afghan law, Islamic Shari’ah

“After these workshops all jirga decisions are based on Shari’ah and Afghanistan constitution. Before we did not know anything that we know now.”

- RLS-I elder from Chawkai district (Kunar)

⁹ Hypotheses I.1 and I.3 are measured across both elders and disputants, while Hypotheses I.2 and I.4 are measured only by elders.

Graduating elders report applying more Afghan law and Shari’ah and less community norms in their decision making and more often agree that community norms may sometimes conflict with Afghan law and Shari’ah.

Hypothesis 1.1 is measured by asking elders the sources of law applied by their jirga when resolving a dispute. This is seconded by disputant perceptions of the legal sources behind the resolution of their specific outcome. The selected indicators did not appear strong enough to make a decisive conclusion about application of legal sources, however, since it is suspected that respondents may have reported a mixture of actual practice and what they knew to be “best practice”. However, the indicators do confirm that respondents had an increased sense of how the process should work. Furthermore, excessive variation in the data suggests that it should be viewed cautiously. Further qualitative research is needed to probe the thought processes behind elder responses to Hypothesis 1.1 questions and properly contextualize the results. Nevertheless, it is instructive to chart movements in these indicators to gauge whether elders cite suggested practice, even if the indicators are not sufficient to demonstrate changes in practice.

On balance, measures for Hypothesis 1.1 do support the RLS-I development hypothesis in that graduating elders perceive themselves to be applying Afghan law and Islamic *Shari’ah* to a greater extent and community norms to a lesser extent in their decision making. Elders further perceive, as a result of RLS-I, that community norms may sometimes conflict with Afghan law and Islamic *Shari’ah*. However, the data appear too noisy to draw solid conclusions and are generally not recommended for use except as cursory diagnostic measures, with the possible exception of elders admitting that community norms may sometimes conflict with Afghan law or Islamic *Shari’ah*.

Evaluation findings for Hypothesis 1.1 follow. Gain scores for each group are expressed as raw values along the 5-point rating scale presented in the survey interview, while d-i-d and spillover measures are presented as percentage increases from baseline values. Bolded values represent findings of substantive significance.¹⁰

¹⁰ In this study, substantive significance refers to findings that are considered meaningful for the RLS-I programming environment. In contrast, statistically significant findings may be able to determine that the result is not likely due to chance, but still may not reflect a meaningful result. Where effect sizes are given, substantive significance may be considered to have been reached at 0.2.

Table 11 Change in elder perceptions of legal sources of dispute adjudication, D-i-D

Attitudinal item	Treatment gain	Comparison gain	Spillover gain	D-i-D	Spillover D-i-D
Jirgas apply Afghan law	0.77	0.46	0.05	8%	-18%
Jirgas apply <i>Shari'ah</i> law	0.48	0.37	-0.42	3%	-30%
Jirgas apply community norms	0.19	0.78	-0.67	-17%	-55%
Community norms conflict with <i>Shari'ah</i>	0.30	0.03	0.37	12%	6%
Community norms conflict with Afghan law	0.58	0.00	0.93	25%	32%
Afghan constitution expresses <i>Shari'ah</i>	0.06	0.19	-0.77	1%	-36%

One check on consistency of results is to note whether district-level results agree with each other, or move in contrary directions – indicating either noise in the data or different dynamics specific to each district. Based on this criterion, the most consistent measures for Hypothesis 1.2 are that community norms may sometimes conflict with Afghan law or Islamic *Shari'ah*.

See [Annex 4](#) for scores disaggregated by district.

Elders perceive their decision making to be resolving the civil aspects of a case

“In places where our government is stable we only try to maintain peace; we try to decrease violence between people. But in areas where there is no government, jirga has full authority and we ask for agreement papers.”

- Chawkai elder on his role in resolving disputes in relation to the state

RLS-I elders recognize the jurisdiction of the state and the legally-acceptable scope for non-state dispute resolution. Along with the modest decline in elders perceiving their decision making as resolving the criminal aspects of a case, there is a strong increase in elders perceiving their decision making as resolving only the civil aspects of a case.

A critical RLS-I objective is to impress upon elders an understanding of and appreciation for the respective boundaries for legal action by formal and informal justice actors. Put simply, elders may help resolve civil disputes and the civil aspects of criminal disputes, but may not apply any sort of punishment and must leave all aspects of criminal prosecution and punishment to the state.¹¹ Government jurisdiction over criminal prosecution and penalty corresponds to the *Shari'ah* concept of *haq-ullah*, or the rights of God that the state, and only the state, is charged with satisfying. The reconciliation of parties, in which forgiveness by the victim is offered in return for the appropriate restitution from the offending party, corresponds to the *Shari'ah* concept of *haq-ul abd*, or the civil rights that should be satisfied when an offense has occurred.¹² Results for Hypothesis 1.2 show a respectable decrease in elders perceiving themselves as resolving the criminal aspects of a case, and a sharp increase in elders perceiving themselves as resolving only the civil aspects of a case.

¹¹ District actors often engage elders for support and fact-finding during the investigation of crimes prior to prosecution.

¹² See the baseline evaluation report for an exploration of elder views on the scope of their jurisdiction in resolving disputes.

Evaluation findings for Hypothesis 1.2 follow. The elder survey posed direct questions asking whether elders saw themselves as resolving the criminal aspects of disputes such as acts of violence or serious crime, or whether they saw themselves as resolving the civil aspects of such disputes, or a mix of both. The impact measures are as follows:

Table 12 Jurisdiction, D-i-D

Elder jurisdiction	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover
Elders resolve <i>haq-ullah</i> (criminal aspects)	-6.6%	14%	4.5%	-11.1%	3.7%
Elders resolve <i>haq-ul abd</i> (civil aspects)	36%	25%	-2.6%	38%	1.6%
Elders resolve both	-15%	38%	23%	-38%	2.3%

There is a modest decline in elders perceiving their decision making as resolving *haq-ullah*. It is also noteworthy that the refusal statistic for the question of jurisdiction fell by more than half from 43% to 20% baseline to endline, with a slightly higher proportion of RLS-I elders willing to answer the question relative to non-program elders. This is further indication that elders are more willing to address the idea of accepting limits to informal dispute resolution, and by extension are more willing to refer criminal disputes to the state and collaborate with formal justice actors in resolving the civil aspects of the case. In such cases, district courts may then take the civil reconciliation between parties into consideration when applying discretion in sentencing as allowed by law.

This result is seen graphically as follows:

Figure 2 Increase in elders resolving only the civil aspects of a dispute, D-i-D



See [Annex 5](#) for results measures by district.

A total of 28 elders took the opportunity to offer a general comment on how they perceived themselves in terms of their jurisdiction and relation to the state. A review of these comments along with comments from the baseline data collection indicates that the most relevant factor behind elders reporting that their adjudication resolved criminal aspects of a case was whether the district government was able to carry out its responsibilities for criminal prosecution. In instances where the government could not or would not carry out its responsibilities, elders were unapologetic in claiming to carry out these responsibilities by default. There was also a significant number of elders claiming to resolve both civil and criminal aspects of a case wherever the elders cooperated with the district government, and/or were deputized by the district government to handle the civil aspects of a case while district authorities handled the criminal aspects.

Increased formality of informal decisions through documentation and registration

“RLS-I decision books give a formal image to our decision since it has a standard format for all types of decisions and it helps us keep the records. Before we used to document the decision on a piece of paper and then after a short time we would lose those copies.”

- Elder comment from Rodat, Nangarhar

RLS-I elders put their decisions on record and register those records with a government entity more often than elders who do not participate in RLS-I. RLS-I elders are 26% more likely than comparison group elders to report that decisions are recorded in their community.

Hypothesis 1.3 is measured by querying elders on the general practice of writing down decisions and registering them with a government entity, and disputants as to whether their specific decision was written down, and if so whether the recorded decision was also registered with a government entity. These data are complemented by tracking of RLS-I decision book usage and assessments of formal-informal justice linkages at the district level.

Results for Hypothesis 1.3 show an increase in both incidence and extent of dispute documentation and registration, as shown in the following table.

Table 13 Changes in elder reports of decision documentation and registration, D-i-D

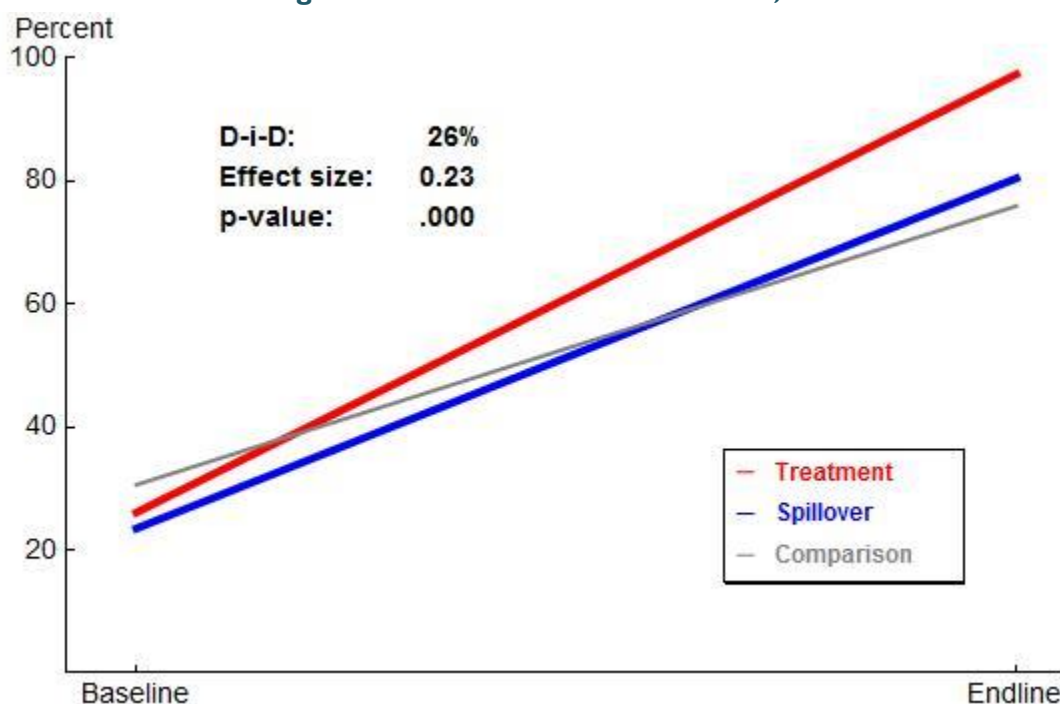
Measure	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover D-i-D
Decisions documented	71%	57%	45%	26%	12%
Decisions registered	21%	18%	7.7%	13%	10%

RLS-I promoted a significant increase in the documentation of informal justice decisions, over and above a large increase observed in the comparison group (which most likely reflects increased rates of documentation generally). Despite the strong underlying trend, RLS-I elders are 26% more likely than comparison group elders to report that decisions are recorded in their community. There is a mild increase in RLS-I elders reporting the incidence of registration .¹³

¹³ The incidence statistics likely reflect a degree of acquiescence bias.

For illustration, the d-i-d measure for decision documentation is graphed below.

Figure 3 Informal decisions recorded, D-i-D



In the above graph, note the steep slope for all groups, indicating the general increase across the board in dispute documentation. There are no appreciable differences in reported rates of documentation and registration based on whether an RLS-I elder received an official program decision book.¹⁴ There is a small increase in both documentation and registration among elders who had heard of the RLS-I decision book, but did not receive one.

See [Annex 6](#) for results measures by district.

¹⁴ The RLS-I decision book contains blank forms for dispute documentation and possible registration and is meant to increase the legality and enforceability of informal decisions to whatever extent elders and citizens find value in doing so. There are many situations where disputants and/or the greater community prefer not to record or register a decision. For example, family cases are kept private, while disputants in cases involving violence or serious crime (usually the offender) would obviously prefer that the community resolve the civil aspects of the case without alerting the district government to handle the criminal aspects.

Increased elder knowledge of Afghan law, Islamic Shari’ah

“We learned the conditions in which a woman is allowed to be separated from her husband; before we believed that a woman can never be separated from her husband.”

- A spinsary from Alishang (Laghman province)

RLS-I elders gain and retain knowledge that is practical and relevant in helping them resolve disputes, but struggle to retain abstract and *de jure* legal knowledge. RLS-I participants saw a 12.3% overall gain in knowledge scores relative to the comparison group. This was driven by strong gains of 20% in family and inheritance law and 18% in property law, but only a 4% gain in constitutional criminal law.

Hypothesis 1.4 was most directly within RLS-I’s manageable interests through application of a core curriculum of six workshops delivering 30-35 hours of lecture, small group discussion, and question and answer sessions delivered over the course of 4-5 months. It should be noted that while the evaluation questions were presented as tests of objective knowledge, respondents likely answer according to a mixture of what they understand the law to be, what they think it should be, and the extent to which they see it practiced in the community. While these different effects cannot be disentangled in the data, they do provide the added benefit of measuring some degree of actual change in the community, and not simply a change in an elder’s understanding of the law. On the other hand, the reported data may mask actual elder knowledge gains, where that knowledge covers rights not yet enjoyed in the community.

The Phase 3 impact evaluation originally consisted of 42 survey items testing elder knowledge on legal topics spanning constitutional law, family and inheritance, and property and deeds. The final set of knowledge items consists of 34 items due to a mix of formatting error and respondent feedback on how the questions were understood and answered.¹⁵ Results for the final set of knowledge items follow. As explained in [Annex I](#), results are based on program participants in treatment districts, distinct from the two separate comparison groups of non-program districts and elders in treatment districts who were not program participants (spillover group).

Table 14 Elder knowledge gain scores

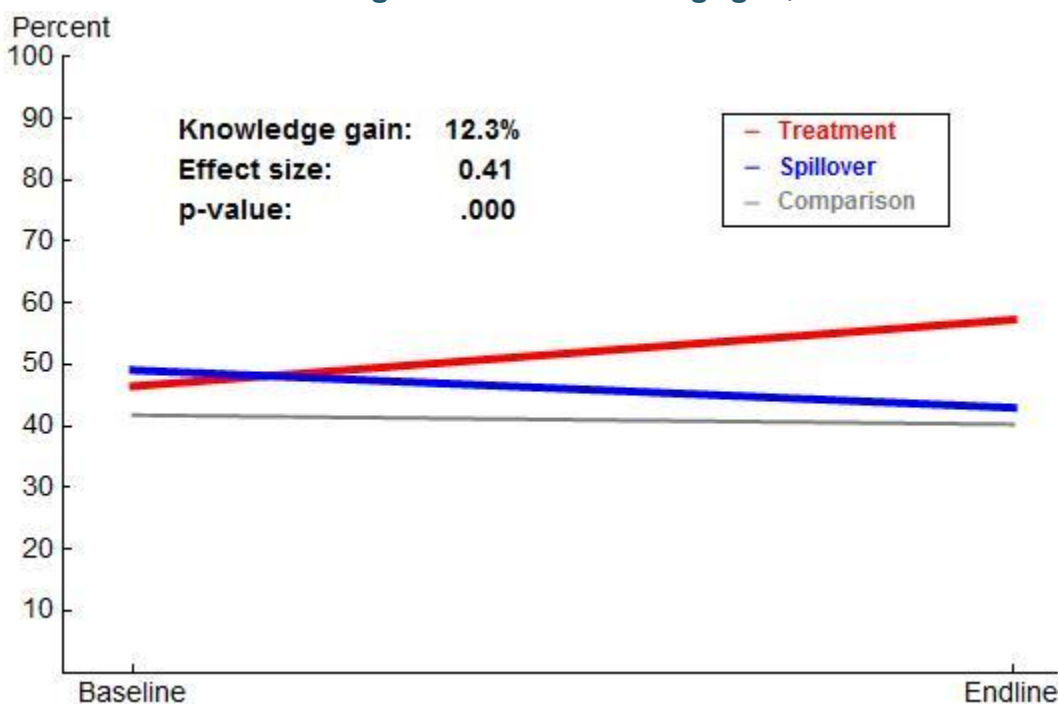
Topic	Items	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover D-i-D
Constitutional and criminal law	16	11.5%	-7.1%	7.5%	4%	-15%
Family and Inheritance	14	11.5%	-6.1%	-8.3%	20%	2.2%
Property	4	5.8%	-2.5%	-12%	18%	9.3%
Overall	34	11.3%	-6.1%	-1.8%	13.1%	-4.6%

There are strong gains in family, inheritance, and property, although those gains are partially supplied by decreases in the comparison group. The spillover effect to non-program elders in treatment districts is

¹⁵ These items are not the full set of knowledge items initially tested. See [Annex 7](#) for a list of questions dropped from the evaluation for various reasons, and [Annexes 8-12](#) for results disaggregated by individual knowledge item.

also substantial for these topics. RLS-I saw a weak gain in elder knowledge of constitutional and criminal law, though this result is not statistically different from zero and is also partially due to an increase in the comparison group score. Relative to the spillover group, program elders realized a 12.3% gain on constitutional and criminal law items.

Figure 4 Overall knowledge gain, D-i-D



See [Annex 13](#) for a breakdown of elder knowledge scores by district. An additional and significant point of context is Chora district (Uruzgan), where RLS-I repeated the program cycle with a new cohort and also selected 20 elders from the original cohort to repeat the program cycle. These experimental steps aimed to assist RLS-I in determining the extent and breadth of RLS-I programming needed for sustainable district-wide impacts. Here the elder knowledge gain was 29% - stronger even than the high-performing district of Chawkai (Kunar). This demonstrates that results are possible even in difficult environments in the South region, and may simply be a matter of applying both longer and deeper programming needed in more insecure districts and/or where local justice traditions are more entrenched.¹⁶

Hypothesis 2: Disputants perceive *jirga* members as more impartial

“The land that they were talking about was mine. It was the inheritance of my father.”

- Comment of disputant who disagreed with elders’ decision of his case

In Chawkai district (Kunar province), RLS-I activities were shown to increase disputant assessment of procedural justice and justice of the outcome of TDR processes adjudicated by RLS-I graduates, and this effect increases according to the extent of RLS-I programming. However perceptions of procedural

¹⁶ See the relevant recommendations and discussion of these points in the RLS-I Phase 3 Final Technical Report.

fairness fell in the two other treatment districts. Both disputants and elders report higher perceptions of corruption, but among RLS-I elders such perceptions are much less than comparison group elders. These results require further exploration and critical discussion in order to fit within the RLS-I development hypothesis, and suggest that in Kandahar in particular and insecure districts more generally, even a program cycle of 12 months is not of sufficient breadth or depth, and may not gain sustainable traction among beneficiaries.

While Hypothesis 1 was situated more closely within RLS-I manageable interests as outcome-level measures, Hypothesis 2 measures the downstream impact of RLS-I through the perceptions of Afghans who sought the mediation efforts of participating elders. Hypothesis 2 measures the dual objectives of strengthened access to justice for citizens and improved stability within communities. The primary measurements for Hypothesis 2 are three index values tracking fairness in the procedural aspects of TDR decision-making, the influence of local powerbrokers or direct wrongdoing from the decision makers, and the overall justice of the TDR outcome. The three primary measures for Hypothesis 2 are reproduced here as sub-hypotheses:

Hypothesis 2	The intervention will increase the perception of disputants that <i>jirga</i> members and their decision making is impartial
Hypothesis 2.1	The intervention will increase the extent of disputants perceiving fairness in the procedural aspects of the resolution, regardless of outcome
Hypothesis 2.2	The intervention will decrease the extent of external influence over the process and outcome of an informal dispute resolution
Hypothesis 2.3	The intervention will increase the extent of disputants perceiving justice in the decision

While disputants are the primary beneficiary group of interest, for some measures there were corroborating data from elders and citizens.

Measuring access to justice

“We have more access to justice and jirga because jirga members learned from your workshops to keep equality between rich and poor, men and women. Today, women are much valued by jirga.”

- Comment from RLS-I spinsary from Goshta (Nangarhar province)

Disputants seeking the mediation of RLS-I elders are more satisfied with the process and outcome of their dispute resolution than disputants who did not seek RLS-I elders

The core impact evaluation measurements for Hypothesis 2 consist of a battery of attitudinal items on case resolution process and outcome. The attitudinal items 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 five-point Likert scale capturing the extent to which the disputant believes a given statement to be true.

RLS-I adapted the TISCO methodology by establishing three dimensions specific to USAID Rule of Law objectives and evaluation hypotheses: Procedural Justice, Subversion of Decision, and Justice of Outcome. Each dimension is an index value of a set of attitudinal items relating to the specific dynamics of the respondent's dispute and its resolution. These items are as follows:

Table 15 Hypothesis 2 index items

Procedural justice	Subversion of decision	Justice of outcome
Freely submitted to the decision making authority	Decision makers considered which party more powerful	Decision makers sought consensus amongst themselves
Disputant able to communicate facts of case		My rights respected
Disputant able to communicate feelings and opinions about case	Decision makers unwillingly influenced by outside factors	Decision allowed reconciliation
Case given due consideration by decision makers		Dispute fully resolved
My arguments given equal consideration with opposite party	Decision makers solicited payment to affect outcome of case	Agreed with decision
Decision makers sought consensus from community		Overall process was fair
Decision makers consulted all relevant parties/witnesses	Decision makers sought own gain in adjudicating dispute	Would choose <i>jirga</i> again if faced problem in future
I preferred that this body decided my case		

Following the assessment scale in the *TISCO Measuring Access to Justice Handbook*, each item was evaluated along a five-point Likert scale: To no extent (1), To little extent (2), To some extent (3), To great extent (4), and Completely (5). Respondents express their perceptions of how well these aspects of justice performed in their own cases.¹⁷

[Annex I](#) establishes the study samples of disputants. Results are as follows, measured along the 5-point response scale:

Table 16 RLS-I impact measures, D-i-D

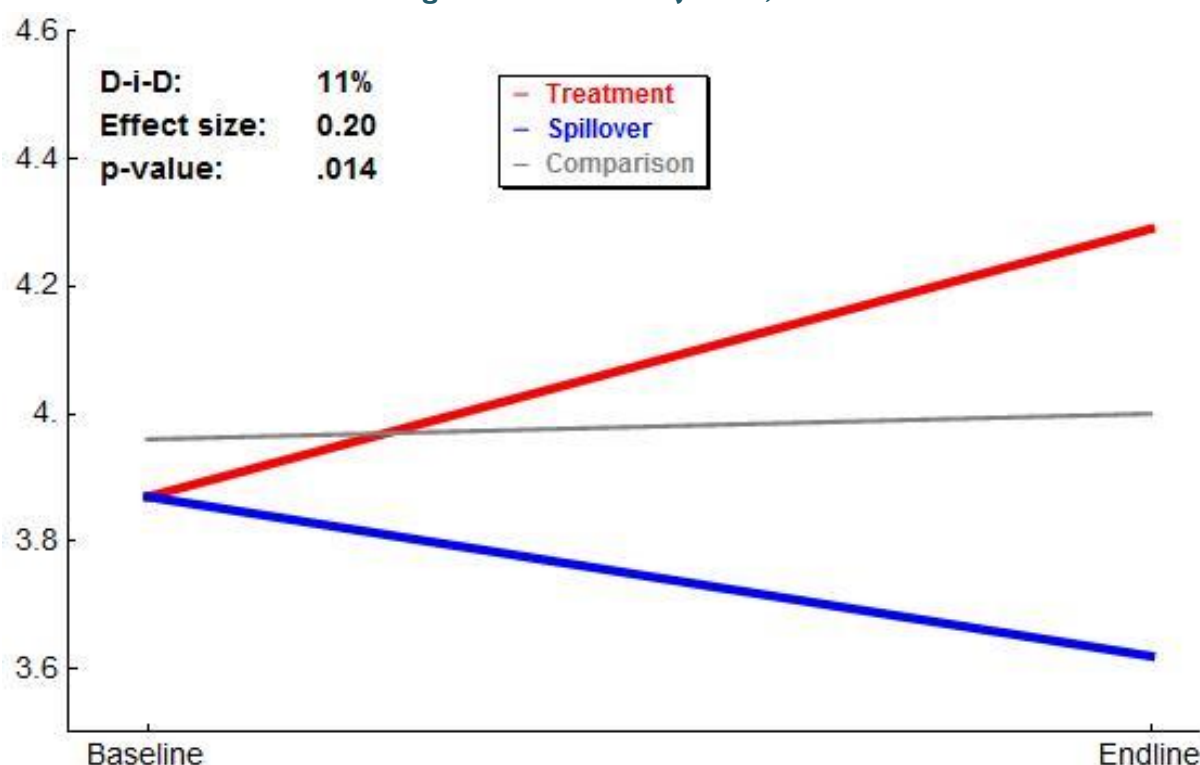
Index	Treatment gain	Comparison gain	Spillover	D-i-D	Effect size	p-value ¹⁸
Procedural Justice	0.42	0.04	-0.29	0.39	0.21	.014
Subversion of Decision	0.34	0.02	0.70	0.33	0.13	.293
Justice of Outcome	0.24	-0.06	-0.39	0.30	0.16	.037

There are respectable improvements in procedural justice and justice of the outcome, and an increase in the corruption measure that is not statistically significant. The procedural justice result is shown graphically, and expressed as a percent gain.

¹⁷ Of the nineteen questionnaire items initially included among these indices, four items were removed after examination of the endline data set. See [Annex 14 for discussion, and Annex 15 for d-i-d scores by individual index item](#).

¹⁸ Standard errors are clustered by village

Figure 5 Procedural justice, D-i-D



The spillover group measures for disputants may be safely disregarded as an artifact of the results measurement, and not having any programmatic relevance with the possible exception of cases where the spillover group serves as a better comparison group than disputants in a neighboring district.

Impartiality – the elder and disputant’s perspectives

Results measures for Hypotheses 2.1-2.3 established a general increase in disputant perception of undue influence, either from within the *jirga* or from local powerbrokers attempting to interfere with the resolution. This trend is corroborated by RLS-I elders, who also report an increase in undue influence, though at a lower level than comparison districts. The table below summarizes.

Table 17 Elder perceptions on the influence of local powerbrokers, D-i-D

Survey item	Unit of measure	Treatment gain	Comparison gain	D-i-D	Spillover D-i-D
Outsiders attempt to influence outcome based on their own interests or connections	Percent	-1.9%	14.8%	-17%	-0.7%
How often outsiders attempt to influence	1-4	0.13	-0.08	0.21	0.55
How often outsiders succeed in influencing resolution	1-4	0.23	1.31	-1.07	-0.76
Jirga members consider party more powerful	1-5	0.30	0.81	-0.51	0.25
Jirga members unwillingly influenced	1-5	-0.23	0.60	-0.83	0.23

Relative to comparison districts, measures of undue influence in informal justice dispute adjudication fall in four out of five measures in RLS-I districts, with one measure showing an increase in RLS-I districts that is not statistically significant. . RLS-I elders perceive a greater extent of undue influence (including a self-assessment measure that elders take into account the relative social standing of disputants in their decision making) on three of five measures, which is not nearly as high as the assessment from comparison group elders. There is a general trend of increasing extent of undue influence in both groups, with a much lower increase among RLS-I elders.

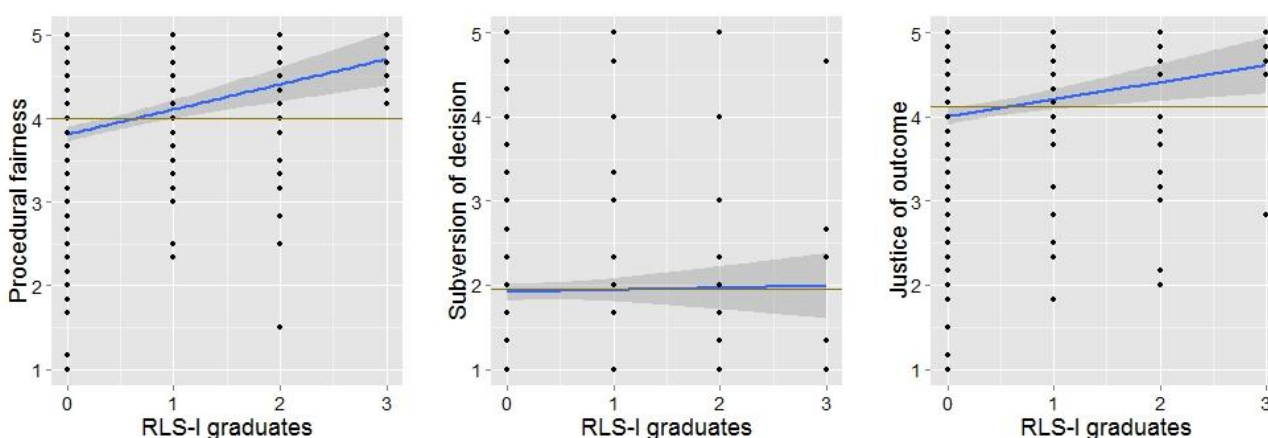
This leads to two possible, but somewhat speculative, explanations for increased perception of undue influence among Afghans who sought the mediation services of RLS-I elders. First is that there is a general degradation of the implementation environment (which, for example, allows powerbrokers to manipulate TDR processes and decisions), and this degradation just happens to be perceived more in RLS-I districts than in comparison districts. Another explanation is that the effect of RLS-I is to raise disputant expectations as they become more aware of best practice and their legal rights as a result of RLS-I public outreach and direct outreach to *spinsaries*. Disputants will become aware of undue influence in dispute adjudication and observe it in their disputes before elders actually change their practice. This would be reflected in the data as an increase in corruption, which in actuality is a sort of results statement in that the effect of RLS-I was to make disputants aware of power relations and imbalances in the resolution of their dispute, or generally make disputants more demanding of informal justice.

Dose-response treatment variables

As more RLS-I elders adjudicate, disputants report higher satisfaction. Disputes adjudicated by three RLS-I graduates scored 15-25% higher (with some measures as high as 30%) on justice measures compared to disputes with no RLS-I graduates among the mediators.

To assess dose-response relationships for disputants, the evaluation team examined disputant satisfaction in relation to the number of mediating elders, the number of “graduates” as defined by requisite levels of workshop attendance (one graduate being equivalent to six workshops attended by any participant), and the number of workshops attended. See [Annex I](#) for discussion. A graphical representation of the dose-response variable follows, using “graduates” as the dose variable.

Figure 6 Dose-response by number of RLS-I graduates mediating a dispute



The horizontal gold line represents the comparison group at endline. A single RLS-I graduate serving on a jirga predicts a slight uptick in disputant assessment of procedural fairness. There is a mild to moderate increase as two and then three RLS-I graduates serve on the jirga. For every RLS-I graduate serving as a mediator, disputant assessment of procedural fairness and justice of the outcome increases by 4-7%.

There is no relationship between RLS-I graduates and perceptions of undue influence. Recall that the previous section offered two possible, non-exclusive explanations for why perceptions of undue influence might increase among Afghans who sought the mediation of RLS-I elders. The absence of a relationship between number of RLS-I graduates on a jirga suggests that it is not the case that RLS-I is raising expectations prior to elders adjusting their adjudication practice. Rather, it was the degrading implementation environment during the period of performance that increased perceptions of undue influence.

See [Annex I6](#) for the actual values of each dose-response specification.

Elder knowledge and disputant assessment

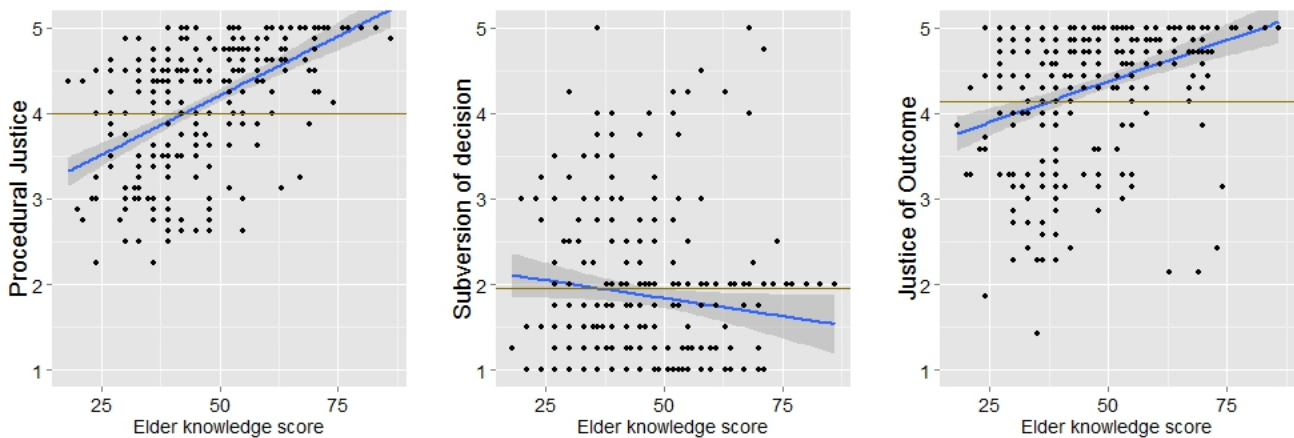
“We learned that bride price is forbidden according to Shari’ah and Afghan law, and we are encouraged to know about women’s access to formal justice if the jirga violates the rights of a woman.”

- Spinsary from Nazyan (Nangarhar)

As RLS-I elders gain knowledge, disputants report higher satisfaction. At the average knowledge gain score of 12-14%, disputant perception of procedural fairness is predicted to increase by nine percent, justice of the outcome six percent, and undue outside influence to fall by four percent.

In addition to examining the relation between the number of RLS-I graduates and disputant assessment, it is also possible to identify whether there is a direct relation between elder knowledge and disputant assessment. The following graphs illustrate.

Figure 7 Elder knowledge and disputant assessment



The horizontal gold line represents the comparison group at endline. Note that elder knowledge at approximately 50% or higher represents an improvement over comparison group elders. Phase 3 elders improved their knowledge from 46% at baseline to 57% at endline. At the average knowledge gain score of 12-14%, disputant assessment of procedural fairness is predicted to increase by nine percent, justice of the outcome to increase by six percent, and undue influence to fall by four percent. This is further validation of the development hypothesis, and also suggests that a primary mechanism behind improvements in disputant assessment are driven by the learning content of RLS-I activities.

Note also that the average knowledge gain among RLS-I elders has the desired negative relationship with disputant assessment of undue outside influence, even though it was previously reported that overall disputant assessment of undue influence increased. This suggests that at a certain threshold of knowledge transfer, RLS-I does help reduce disputant perception of undue influence, but that in the Phase 3 evaluation sufficient knowledge had not been transferred. Stated differently, the effect of the changing environment was stronger in increasing disputant perception of undue influence than was the effect of RLS-I programming in reducing disputant assessment of undue influence.

Hypothesis 3: Decrease in TDR decisions harming women and children

“My uncle engaged his son without asking his consent, and he then started shouting and begging his father to stop it because he is not happy. But his parents forced him. Now, the girl is suffering and her husband is not valuing her as a wife.”

- Spinsary comment on the freedom to choose one’s spouse in Afghan society

Households receiving RLS-I outreach material were 7% more likely to support alternatives to *baad* and 4.6% more likely to affirm that *baad* is not an effective solution to disputes – both of these attitudinal items are part of RLS-I messaging.

Hypothesis 3 posits that RLS-I will reduce harmful TDR practices in Afghan society. Unfortunately, it is not feasible to estimate the incidence of such practices given their sensitivity and rarity, and the limited resources of RLS-I to carry out national surveys or community censuses. RLS-I does query citizens and *spinsaries* as to cases of *baad* and forced marriage; however, such estimates are not credible for inferential purposes due to a variety of data quality concerns. The evaluation team explored a number of relevant secondary measures, displayed in Table 18. D-i-D measures are expressed as percentage gains, while the treatment and comparison gain scores are in units of the original 5-point scale.

Table 18 Outreach measures on women's role in TDR, D-i-D

Measure	Treatment gain	Comparison gain	D-i-D
Better to find alternatives to <i>baad</i>	0.57	0.31	6.9%
<i>Baad</i> is against Sharia	0.54	0.78	-6.4%
<i>Baad</i> against Afghan law	0.43	0.34	2.4%
Families practice <i>baad</i> due to economic situation	0.28	-0.53	24%
<i>Baad</i> will always be a part of society	-0.03	-0.02	-0.7%
If State justice improves, <i>baad</i> will decrease	0.35	0.08	9.2%
<i>Baad</i> is not an effective way to prevent disputes	-0.06	0.06	4.6%

Households receiving RLS-I outreach material were 7% more likely to support alternatives to *baad* and 4.6% more likely to affirm that *baad* is not an effective solution to disputes – both of these attitudinal items are part of RLS-I messaging. On more diagnostic rather than evaluative items, households were 24% more likely to recognize an economic link with the practice of *baad*, and 9% more likely to link effective state justice with reduction in *baad*.

In the absence of incidence measures, qualitative research is more valuable and revealing for assessment purposes, even if such research cannot answer whether prevalence of a given practice is increasing or decreasing. For Phase 3, RLS-I received 2-3 dozen reports of RLS-I elders and *spinsaries* successfully advocating against the practice of *baad*, either by stressing the availability of alternative payments to achieve resolution, stressing the un-Islamic and illegal nature of the activity, or some combination.

There were also 2-3 reports of unsuccessful advocacy, including one intercession against a judgment rendered by anti-government actors that imperiled the life of an RLS-I *spinsary*. Generally speaking, RLS-I elders and *spinsaries* could find success advocating against *baad* when the dispute was between two families, especially if there were any familial connection between the disputing families and the RLS-I elder or *spinsary*. However, when disputes escalated to the attention of tribal elders and the issue was considered to involve entire clans or tribes, the use of *baad* as a means of resolution was much more difficult to combat. Both elders and *spinsaries* from treatment and comparison districts are in unanimous agreement that *baad* is an outdated and harmful practice that is still practiced but also in decline as general education levels improves and extreme poverty decrease.

Another powerful opportunity for advocacy concerned the practice of *badal*, or the mutual exchange of girls between two families or tribes. RLS-I *spinsaries* reported several cases where one spousal couple from the exchange was unhappy in the marriage and eventually divorced, leading family members of the divorced bride to return the bride from the other spousal couple regardless of the state of that

marriage or their son's wishes. This magnified the rancor in a community as one couple divorced out of unhappiness while the other couple was unhappy that they too had to divorce. At these times, RLS-I *spinsaries* took the opportunity to educate the community on the negative consequences of giving away girls in marriage, and again stress its un-Islamic and illegal nature.

The most heartening examples of avoiding *baad* involved cases where a couple had been engaged to each other as children, and upon reaching maturity were to formalize the engagement in a marriage contract. At such times, both RLS-I *elders* and *spinsaries* took the opportunity to advocate against the practice and remind the father that under both Afghan law and Hanafi jurisprudence (one body of Islamic law) the daughter must consent to the marriage. In several cases such marriages were cancelled despite the childhood engagement, and in other cases the daughter consented to marriage or requested a delay in order to pursue further education before making a decision. In this way child engagements that had been made years previous could still be redressed at the time of marriage.

Consent to marriage and domestic violence

"We learned about the right of a woman to consent to her marriage. This caused us to ask our daughters' consent when we come to a decision about their marriage."

- *Spinsary* from Goshta (Nangarhar)

While the data on the incidence of harmful practices is not trusted, there is (marginally) more reliable data that attempts to draw a link between children's consent to their marriage partners and later disputes between spouses or their families. In the following table of results, the first two measures are diagnostic, while the last two measures demonstrate RLS-I impact.

Table 19 Consent to marriage and family disputes, D-i-D

Measure	Treatment gain	Comparison gain	Spillover gain	D-i-D	Spillover D-i-D
Is it possible that parents' marriage arrangements could be against the wishes of the children?	2.9%	-6.5%	25%	9.4%	32%
If a marriage was against a child's wishes, might this lead to more disputes between the spouses or families?	3.3%	-4.5%	-6.8%	7.8%	-2.4%
Are sons in your community able to choose their marriage partner?	19%	8.2%	4%	11%	-4.1%
Are daughters in your community able to choose their marriage partner?	11%	4.6%	37%	6.3%	33%

There is an encouraging increase in consent to marriage in both treatment and comparison districts, with even stronger gains in the RLS-I districts. However, the data are noisy and the results are not statistically significant. While chance cannot be ruled out as an explanation for these results, they are

supported by anecdotal reporting, with both male and female elders regularly reporting that consent to marriage under Afghan law and Islamic *Shari'ah* was taught to and accepted by their communities.¹⁹

Hypothesis 4: Strengthened role for women in TDR

Citizens in RLS-I districts are more likely to report a change in women's roles as disputants, mediators, and decision makers, as well as support stronger women's roles in general. RLS-I elders report no change.

Survey measures such as women serving as mediators for *village-level* disputes, however, show no movement from baseline to endline, per the following table.

Table 20 Elder reports of women's role in *jirga*, D-i-D

Survey item	Treatment gain	Comparison gain	D-i-D	Effect size	p-value
Jirgas consult all relevant parties, including women	0.15	0.12	0.03	0.01	.858
If a female involved, her testimony would be solicited	-0.12	-0.23	0.11	0.04	.510
In cases involving women, elders would consult other women	0.34	0.46	-0.12	0.04	.452

The household survey presents a more encouraging picture, though the household survey data is considered to be of uncertain validity.²⁰

Table 21 Household reports of women's role in *jirga*, D-i-D

Measure	Scale	Treatment difference	Comparison difference	D-i-D
Women personally present the case	%	-1%	-18%	17%
Women present case to the wife of dispute resolver	%	-3%	0%	-3%
A family member represents the female disputant	%	4%	18%	-14%
Wives of Mohammad played role in resolving disputes	1-5	0.24	0.57	-8%
My community would not accept women as dispute resolvers	1-5	0.15	0.11	0%
If I had a dispute, I would not accept a woman as a dispute resolver	1-5	0.10	0.12	0%
I support women helping resolve disputes in the community	1-5	0.29	-0.19	19%
I support women helping resolve disputes of other women	1-5	0.50	-0.11	24%

There is a 17% gain in women presenting their case directly before *jirga*, as well as a 19% gain in households supporting the proposition that women could serve as dispute resolvers. The strongest gain of 24% is in support of women helping other women resolve disputes. It is possible that citizen awareness of RLS-I *spinsary* groups helped contribute to this result.

¹⁹ When additional controls are added, the d-i-d measure for consent to marriage for sons declines to zero while consent to marriage for daughters increases to 15-20%.

²⁰ A few household measures for Hypothesis 4 were discarded after data quality assessment. The data presented here are considered to be of moderate to good quality.

Gender justice gap among disputants

The Phase 2 evaluation documented a 15-30% “gender justice gap” between male and female disputants. A similar gap is found in the Phase 3 evaluation. While these measures looked generally at the experience of female disputants in both the treatment and comparison groups, the following table presents RLS-I impacts for disputants disaggregated by gender.

Table 22 Disputant perception by gender, D-i-D

Measure	Treatment gain	Comparison gain	D-i-D (male)	D-i-D (female)
Procedural fairness	13%	1.3%	12%	-14%
Subversion of decision	17%	-5%	22%	-11%
Justice of outcome	-7.5%	-1%	8.4%	-16%

Female disputants show declines in assessment of elder adjudication on measures of procedural fairness and justice of the outcome, but also a decline in perception of undue influence. These results are statistically significant, leading to a finding that RLS-I raises expectations of Afghan women such that they assess their dispute resolution experience lower than at baseline.

Conclusions and lessons learned

“My son, Israel, is an educated boy. I carried RLS-I handouts to him. During his off days, he makes short notes from handouts in white flip charts and presents these at the Mosque congregations. [He focuses on] the inheritance shares in the handouts... and he also encourages each one individually to give the rightful portion to their daughters, sisters, wives and other family members.”

- Spinsary from Goshta (Nangarhar)

Program evaluation in conflict-affected environments

This study completes over two years of research and evaluation to validate the RLS-I development hypothesis that capacity building of informal justice actors increases access to justice and strengthens stability in conflict-affected areas. This research effort has shown that regional differences in program setting coincide with differences in outcomes; has empirically demonstrated the link between elder capacity building and improved disputant perception predicted by the RLS-I development hypothesis; and has suggested a hypothetical J-curve impact trajectory where RLS-I may be disruptive and initially lead to a period of zero or even negative results until beneficiary participants and communities fully accept RLS-I learning content and disputants witness positive changes in elders’ adjudication practice.

The fact that building elders’ capacity has now been shown to lead to improved disputant assessment demonstrates that RLS-I improves access to justice. This study examines disputant assessment in three separate ways – by measuring directly, by considering the number of RLS-I mediators to a dispute resolution, and by correlating elder knowledge with disputant assessment.²¹ In all three ways, RLS-I is shown to improve disputant assessment as much as 30% relative to disputants whose cases are not mediated by RLS-I elders. RLS-I programming experience suggests that disputant assessment is most

²¹ Other specifications were used in the analysis and also agreed with the specifications discussed in the text.

positive where elders have some level of education, implementation is not threatened by insecurity, and where there is a district government that is at least partially functioning. This explanation fits the pattern of varied results by district, with Chawkai (Kunar) showing strong results, Mohammad Agha (Logar) showing mixed results, and Zhari (Kandahar) showing no results or possibly even declines in desired outcomes.

Regional differences in outcomes are primarily driven by the extent of insecurity in a community and the level of capacity of program participants. Such outcomes should also be considered in a more general context of consistent retreats in security and increases in perceptions of undue influence of local powerbrokers or anti-government actors during the period of performance. Other relevant factors include the timing of measurement relative to the proposed trajectories of RLS-I impacts and the historical strength of informal justice as a locally-adaptive institution. In Kunar province in eastern Afghanistan, for example, RLS-I shows strong effects, a result that likely generalizes to many districts of Nangarhar province and other relatively secure districts east of Kabul. In areas such as Kandahar, on the other hand, results are more measured or, in some cases, exhibit worse outcomes than at baseline, which likely reflects a “J-curve” impact trajectory in which initial results show worse outcomes before they get better – for example, by raising citizen expectations prior to elders revising their adjudication practice, or by exposing elders to new knowledge that conflicts with what they thought they knew prior to discarding old knowledge and accepting new knowledge as their own.²²

More speculatively, negative outcomes may also reflect attempts by program participants to use RLS-I programming in a manner that strengthens informal justice practices and stability in a community at the expense of local state-building and women’s and children’s rights.²³ This is suggested both by qualitative monitoring reports and evaluation data, where elders from Zhari district, and to a lesser extent Mohammad Agha elders, show contrary outcomes on jurisdictional boundaries between formal and informal.²⁴

Obviously such reactions, should they come into more direct evidence, must be contested and overcome in the event of future programming. But seeing results of legal literacy programming in Kandahar may primarily be a matter of applying the appropriate program cycle, with a 12-month program cycle in eastern Afghanistan often adequate, but perhaps both longer and deeper programming needed in more insecure districts and/or where local justice traditions are more entrenched. In neighboring Uruzgan province, for example, extended programming didn’t seem to affect disputant perception but did boost elder knowledge scores by 29% - stronger than the high-performing district of Chawkai (Kunar), while Kandahar elders demonstrated a range of lackluster results from significant losses to slight gains.

The donor response to negative or absent outcomes should certainly not be disengagement and ceding the informal local self-governance and dispute resolution space to elders, ungoverned militias, or anti-government elements. Rather, more prolonged exposure to learning content, including additional simplified material geared toward less educated participants, is recommended, per the RLS-I Final

²² When a *Shari’ah* professor from Nangarhar University visited Kandahar for guest lectures, his immediate comment was that “Kandahar elders need to learn that they don’t know anything.”

Technical Report. In eastern Afghanistan, elders and state actors alike report that the effect of RLS-I is to expand and build the capacity of a network of elders who cooperate with district governments to extend the rule of law to more distant villages – connecting local dispute resolution to the state and building GIRoA legitimacy at the local level while empowering elders to continue their time-honored roles in resolving community problems. In Kandahar, such effects seem to be premature, especially in the context of corruption issues endemic to all levels of governance throughout the province. But the case of Uruzgan, where the program cycle was repeated over the course of two years, shows that results are possible. Similarly, more peri-urban and secure districts in Kandahar such as Arghandab and Dand – districts where elders have more education and a stronger connection to the state – are considered successful outcomes for RLS-I. Success in Kandahar’s Zhari district should therefore be considered nascent rather than beyond the reach of a legal literacy program.

It should not be forgotten that elders in all districts personally validate the development hypothesis and implementation approaches of RLS-I. Elders consistently report that they would like to maintain their historical roles in local self-governance and dispute resolution, but also state a desire for strong and effective government with which they would be glad to cooperate. Security and governance issues prevent such a result in Kandahar presently, making legal awareness programming all the more crucial to increase legality of decisions, mitigate and prevent harms, and prepare local elders for eventual incorporation into state structures.

Furthering, and discovering, evaluation best practice

The RLS-I Phase 2 and 3 impact evaluations represent bold new practice on the part of USAID. The rigorous RLS-I evaluation design answers the USAID 2011 Evaluation Policy’s call for robust estimate of counterfactual scenarios. One shortcoming is that the evaluation was conducted by an internal sub-contractor in Phase 2, and conducted fully in-house in Phase 3 while best practice dictates that impact evaluation be conducted by external experts.

However, in-house impact evaluation has the advantage that the implementing partners who undertake them will have full situational awareness of the context, evolution and “mission creep” that affect all programs over time, and (hopefully) an immediate grasp of what most needs to be measured and why. External evaluators may not appreciate the context and may be more interested in theoretical constructs to inform an academic debate rather than the nuts and bolts of program implementation.

Furthermore, the question of in-house versus external impact evaluation speaks directly to the question of whether the ultimate intent of impact evaluation is to make judgments about overall performance, promote learning about development effectiveness, or both (and if both, then to know the weight assigned between these competing objectives). Development scholars such as [Lant Pritchett](#) have pointed out that implementing organizations will find it difficult to regard external evaluations as anything but a referendum on overall program performance that will ultimately decide future funding. Obviously, a commitment to learning that entails the risk of highlighting failure – or simply failure to show results – is more difficult amidst a suspicion that evaluation findings will decide future funding.

This tension highlights the pressing need for USAID to establish a programmatic space where its implementing partners may feel safe in undertaking internal evaluation initiatives that can promote learning to improve implementation while addressing larger development research and policy questions

to assist USAID managers and leaders. One possibility is to incorporate impact-level measurements into program performance monitoring plans (PMPs), as occurred with RLS-I Phase 3. However, this too has complications, especially in fragile and/or fluid environments. Performance indicator targets are linked to contractual performance and by extension future funding prospects. Additionally, impact-level measures that incorporate counterfactual estimates are not suitable PMP indicators.²⁵ Target-setting for impacts is also more difficult than for output or outcome measures. For example, a 5% target increase from one program cycle to the next is normal according to a theory of continuous improvement or similar “six sigma” implementation metric. But a 5% increase in an impact level indicator can be a significant shift indicating durable social change. It is difficult to theorize or target such changes *a priori*, and the ability to statistically detect such a change can require significant funding for a study to reach the required level of power. It is not clear that a program PMP can accommodate the additional considerations that come with impact measurement.

The experience of RLS-I Phase 3 suggests that incorporating impact-level measurements into the program PMP is not sufficient to create the recommended “safe space” for implementers to commit to experiment, learn, and improve in a manner that informs broader development theory and practice. Rather, program M&E systems need to be viewed in part as applied research and experimentation in addition to monitoring outputs and identifying outcomes.²⁶

The RLS-I impact evaluation is an example of USAID’s willingness to “[crawl the design space](#)” in applying evaluation best practice, explore new practice, and continually discover what works and what does not – not just for development, but also development evaluation.

More pragmatically, data quality issues, while emblematic of data collection in the challenging context of Afghanistan, also stemmed in part from moderate to small sample sizes. This was particularly prominent with the spillover group data variation, which hindered RLS-I’s ability to confidently assess the scope of RLS-I’s dispersion beyond direct program participants. At the same time, a lengthy interview guide may have contributed to data-corrupting respondent fatigue.

In some cases, questions were unclear to respondents and subsequently removed from the analysis, as noted throughout the report. Therefore, RLS-I recommends:

- Expanding sample sizes to improve confidence in the data (which may require focusing and reducing the number of study questions).
- Shortening the survey tools to be implementable within standard survey timeframes (which would free resources for expanding sample sizes).
- Conduct a thorough pilot test and allow sufficient time to review and rephrase any questions that are found to cause confusion.

²⁵ In the case of RLS-I, impact measures in the program PMP are the more standard before and after measures that do not incorporate counterfactual estimates, while the same measures in the RLS-I impact evaluation do incorporate counterfactual estimates.

²⁶ For a guiding reference see [It’s All About MeE](#), which introduces the concept of structured experiential learning as an additional function for program M&E systems.

Annexes

Annex I: Variation in treatment and group

Variation in group

The evaluation Inception Report establishes two separate groups of interest within treatment districts: elders attending RLS-I activities (participant group) and elders who are proximate to RLS-I participants (either within the same village or in neighboring villages) but who do not attend RLS-I activities. These elders are referred to as the spillover group, and are selected through direct field work. The following table breaks down elders in treatment districts between the participant group and the spillover group at endline.

Table 23 Endline spillover and participant groups by district, Elders

Province	Treatment District	Spillover group	Participant group	Total	% spillover group
Logar	Mohammad Agha	15	52	67	22%
Kunar	Chawkai	23	50	73	32%
Kandahar	Zhari	56	12	68	82%
Overall		94	114	208	45%

The original target for a district spillover group was 10-15% of the total treatment district sample, but increased to 37% in the baseline data and 45% in the merged data including both baseline and endline. This increase was mainly due to the enumerators' difficulty in finding elders from the RLS-I enrolment list, in which case the enumerators identified elders not on the RLS-I enrolment list through direct field work. Another complication that occurred during mobilization phase was that after a cohort of participants were enrolled in Zhari district (Kandahar province), state actors later recanted and insisted upon choosing a new cohort. The stated reason was that the initial enrolment had relied upon outdated lists with many elders who were no longer official members of local bodies such as CDCs and IDLG shura. A new cohort was enrolled after the baseline data collection took place, leaving only 12 elders from the original selection to participate in RLS-I activities.

The total group sizes across districts and including the comparison group is as follows:

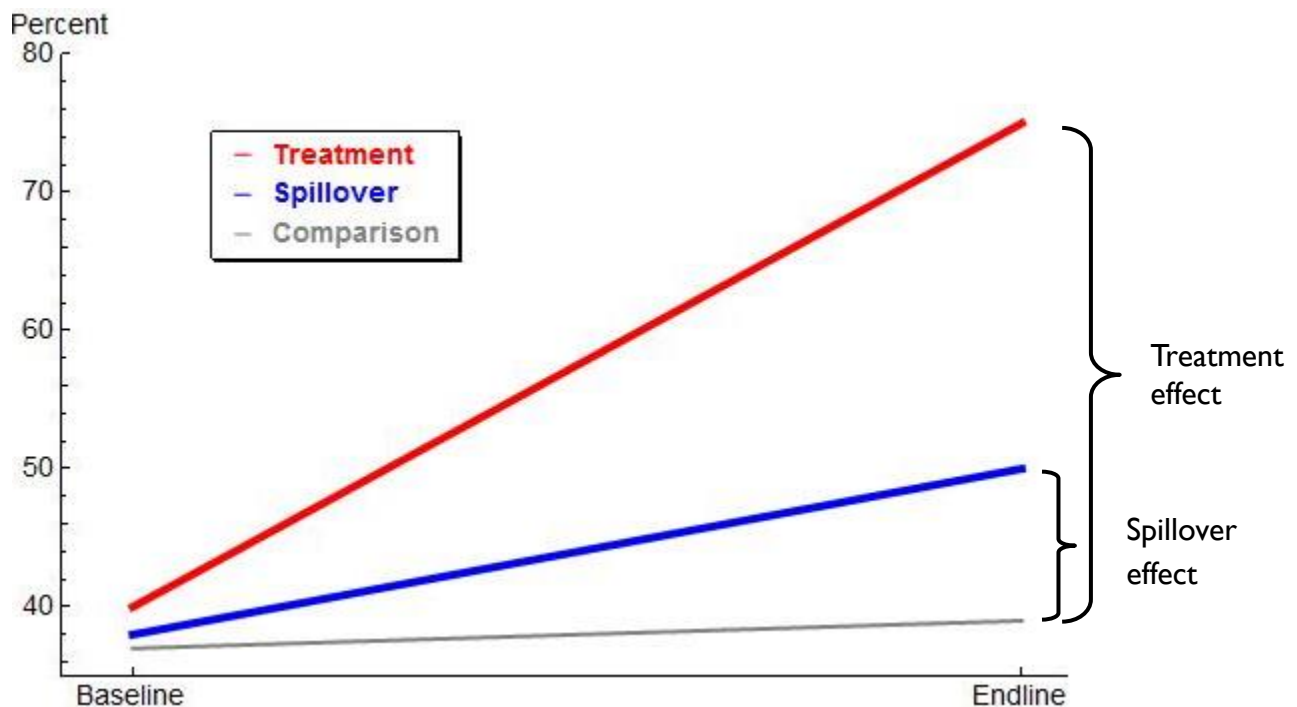
Table 24 Comparison, participant, and spillover groups by wave, Elders

Group	Baseline	Endline
Comparison	217	208
Treatment (Participant)	182	114
Treatment (Spillover)	35	94
Overall	434	416

This breakdown of groups across time allows for presenting the RLS-I evaluation measurements. The base specification for impact evaluation measurements takes into account the trend lines of both the comparison districts and the spillover group within the treatment district. The spillover group is essentially a second comparison group, but designed for the purpose of identifying any peer effects that RLS-I elders may have upon elders who may interact with RLS-I elders but did not attend RLS-I activities. The base hypothesis of this specification is therefore that if there are positive peer effects between RLS-I participants and those they interact with, the spillover group should also see gains in RLS-I results measures, but not to the same degree.

The base specification for RLS-I impact measurement is illustrated graphically as follows:

Figure 8 Base specification for treatment and spillover group



The grey line indicates secular changes over time that are not related to RLS-I. The blue line represents gains among elders who experienced positive peer effects after interacting with RLS-I elders. The red line represents the direct effect of RLS-I upon its participants.

The identification of disputants within either a participant or spillover group is more difficult, as it depends on self-reports of disputants or a successful linking of a disputant interview with the referring elder. As mentioned in the previous section, 56% of disputes were identified by an interviewed elder who was also a mediator for the dispute, and 17% of disputes were identified by an interviewed elder who knew of the dispute without playing a mediating role. This offers the opportunity for linking elders' background characteristics and performance data with disputant perception, and was the mechanism that enabled the Phase 2 evaluation to show a linear relationship between the extent of RLS-I programming and subsequent gains in elder knowledge and disputant perception.

The Phase 3 evaluation continues the practice of linking disputant data with elder characteristics and performance, and also adds another method of linking. Disputants are asked to name the three most active mediators of their dispute, and these names are then linked to RLS-I program participants and elders interviewed in comparison districts. Disputants in treatment districts who did not identify any RLS-I elders as the mediators of their disputes are designated as belonging to the Spillover group.²⁷ The breakdown is as follows:

Table 25 Endline spillover and participant groups by district, Disputants

Province	Treatment District	Spillover group	Participant group	Total	% spillover group
Logar	Mohammad Agha	35	55	90	39%
Kunar	Chawkai	7	63	70	10%
Kandahar	Zhari	64	26	90	71%
Overall		106	144	250	42%

Note that the spillover group is skewed towards Zhari and Mohammad Agha and against Chawkai. The reader should be alert to the possibility that dynamics captured by the Spillover designation may also carry regional differences such as the security environment or extent of state presence. The total group sizes across districts and including the comparison group are as follows:

Table 26 Comparison, participant and spillover groups by wave, Disputants

Group	Baseline	Endline
Comparison	234	290
Treatment (Participant)	222	144
Treatment (Spillover)	--	106
Overall	456	540

This breakdown of groups across time allows for presenting the RLS-I evaluation measurements according to Figure 1 above.

²⁷ In an additional 24 cases, the disputant identified an RLS-I elder, but that elder still did not attend any workshops. These cases are also designated as belonging to the Spillover group.

Variation in treatment

While the addition of a spillover group should help separate the direct RLS-I treatment effect upon its participants and the indirect effect upon those who interact with RLS-I participants, the designation of treatment remains a binary yes/no variable that does not take into account the extent a given elder may have participated in RLS-I. Creating variables that reflect the extent of an elder's participation in RLS-I is part of a dose-response analysis, and provides yet another tactic for isolating program impact upon participants and in beneficiary communities. It is also useful to look at treatment effects for an elder who graduates from the core curriculum of six workshops. For elders, the breakdown of activities attended and graduation from RLS-I is as follows:

Table 27 Activities attended and RLS-I graduation by district, Elders

Province	Treatment District	Workshops	Activities	Graduates
Logar	Mohammad Agha	5.1	10	29
Kunar	Chawkai	5.5	9.6	32
Kandahar	Zhari	3.9	8.6	3
Overall		5.2	9.7	64

Since the spillover group will capture all variation among elders who did not attend any workshops, the treatment variable will implicitly capture RLS-I elder attendance at the mean value of 5.2 workshops or 9.7 activities. The addition of a variable capturing the value-add for a graduate of the core curriculum of six workshops has the advantage of presenting the RLS-I treatment effect in terms of graduated elders, which helps generalize evaluation findings to the universe of RLS-I graduates (over 1,700 during Phase 3) and aids in the discussion of questions of critical mass needed before a district can be considered graduated. It is also possible to define the treatment variable as the number of workshops attended. Elder knowledge gains are slightly higher in such a specification, but it is still considered more convenient to discuss results in terms of graduated elders.

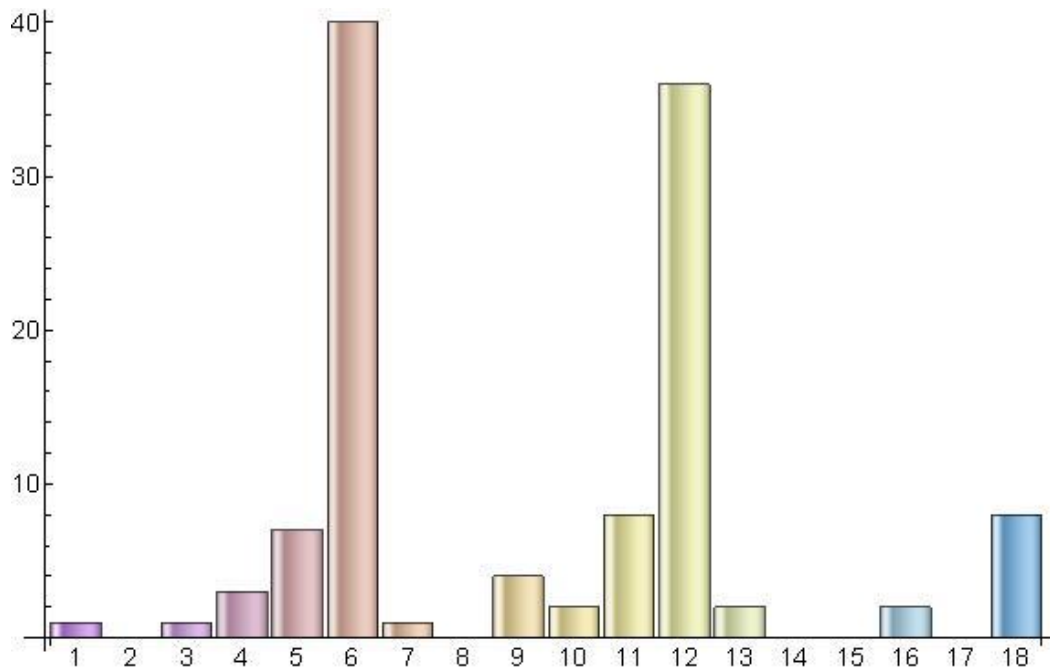
For disputants, dose-response treatment variables can be constructed in three ways. First is the number of RLS-I elders the disputant identifies as a mediator for their dispute, with a range from zero to three. The following table breaks down the number of elders that disputants identified as mediators along with the average attendance for the total number of elders identified for each dispute:

Table 28 RLS-I mediating elders identified by disputants

RLS-I elders	Count	Workshops attended	Activities attended
0	106	--	--
1	37	5.9	11.3
2	46	10.2	18.5
3	32	12.3	25.4
1.8	--	9.4	18

Note that it requires three mediating elders to reach 12 workshops attended, or the equivalent of two graduating elders. Three mediating elders cannot approximate the effect of having three RLS-I graduates help mediate the dispute. The next dose-response variable addresses this by defining a graduate not by the elder, but by attendance at six workshops attended across the sum of all mediating elders. For clarity, note the distribution of the sum of workshops attended:

Figure 9 Total workshops attended across all mediating elders



The horizontal axis denotes the total number of workshops attended across however many mediating elders were identified by the disputant, while the vertical axis denotes the number of disputes in the sample for a given number of total workshop attendance. The spikes in the number of disputes with six, 12, and 18 workshops attended correspond to the RLS-I objective of graduating elders from its core curriculum. These spikes will most often correspond to one, two, and three mediating elders, respectively, but will not always do so. The following table shows the exact distribution of these spikes across the number of mediating elders:

Table 29 Number of RLS-I elders meeting graduation levels of workshop attendance

RLS-I elders	Workshops attended		
	6	12	18
1	27	0	0
2	9	21	1
3	4	15	7
"Graduates"	40	36	8

There is not an exact mapping of elders to graduation levels of attendance, but graduation levels of attendance are reached according to a variety of combinations of the number of graduating elders.

Therefore, a variable for the number of graduating elders can be constructed solely from the total number of workshops attended, regardless of the actual number of mediating elders of a dispute, or the attendance record of each mediating elder. For example, in nine disputes, two mediating elders combined for six workshops attended, or the equivalent of a single graduating elder. In one dispute, two mediating elders combined for 18 workshops attended, or the equivalent of three graduating elders. The “Graduate” variable accepts any combination of mediating elders whose attendance reaches six, 12, or 18 workshops. This dose-response variable essentially asks about the extent of RLS-I learning content brought to bear upon a given dispute, and ignores the issue of how many elders were required to bring it about.²⁸

The final dose-response variable is simply to use the sum of workshops attended across all RLS-I mediating elders, with a range from one to 18. This has the advantage of using the entire distribution of attendance, but the number of elders mediating or the number of graduates are still considered better alternatives in terms of generalizing the results to the entire realm of RLS-I implementation.

The treatment groups and variables introduced in this section form the basis for subsequent evaluation findings. The following annex reviews the regression notation for these specifications.

²⁸ This discussion also ignores the possibility that some combinations of six, 12, or 18 workshops across all mediating elders will include extra repetitions of some workshops while missing others. While this is possible in cases where the attendance of different elders combine to form a single “graduation”, it is considered unlikely that this feature of the data renders this specification uninformative.

Annex 2: Constructing d-i-d measures

The core measurement of this evaluation is that of difference-in-differences with the addition of a spillover group. First, change scores from baseline to endline are generated for all three groups. Then the comparison group's gain score is subtracted from the gain scores of both the treatment and spillover groups to arrive at the estimate of the treatment effect for each. Mechanically, this measurement is calculated as follows:

Table 30 D-i-D measurement setup

Impact Measure	Treatment gain	Comparison gain	Spillover gain	Spillover	D-i-D
Item	Post - Pre	Post - Pre	Post - Pre	Spillover gain – Comparison gain	Treatment gain – Comparison gain

Under random assignment, the d-i-d with spillover group 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. Overcoming the inherent bias in observational data requires balancing pre-treatment characteristics of participants and controlling for factors that also affect program impact in addition to treatment, and even these tactics will not eliminate sources of bias that are not observed in the data.

In regression format, the d-i-d measurement 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, and also depends on whether the spillover group includes a longitudinal component. For disputants, identification of membership in the spillover group was only possible at endline; as a result, the baseline values in treatment districts pool the values of both treatment and spillover disputants at baseline and then diverge at endline. This is expressed in regression format as follows:

$$y = \beta_0 + \delta_0 \text{endline} + \beta_1 \text{treatment} + \delta_1 \text{endline} \cdot \text{treatment} + \delta_2 \text{Spillover} + \gamma_n X_n + u$$

The specification is identical to the base d-i-d measurement, with the exception of another endline measurement showing the divergence of the spillover disputant measures from the treatment disputant measures.

For elders, some (though not all) members of the spillover group were interviewed at baseline, allowing for estimation of the trend lines for all three groups. The specification with spillover group for elders is as follows:

$$y = \beta_0 + \delta_0 \text{endline} + \beta_1 \text{treatment} + \delta_1 \text{endline} \cdot \text{treatment} + \delta_2 \text{Spillover} + \delta_3 \text{endline} \cdot \text{Spillover} + u$$

Here there are two variables capturing the spillover effect – one at baseline and one at endline. Manipulation of the variables in this equation yields the same measurements as found in the table at the beginning of this section. The following table shows how combinations of variables yield the d-i-d measures:

Table 31 D-i-D equation manipulation

Manipulation	Resulting value
Constant	Comparison baseline
Constant + endline	Comparison endline
Constant + treatment	Treatment baseline
Constant + endline + treatment + endline*treatment	Treatment endline
Constant + treatment + Spillover	Spillover baseline
Constant + endline + treatment + endline*treatment + Spillover + endline*Spillover	Spillover endline

In addition to being computationally more convenient, expressing results measures in regression format allows the inclusion of control variables that help overcome the bias in observational data, whether through balancing pre-treatment characteristics or adjusting for other variables that are related to both treatment status and the outcome measure. While inclusion of such controls seldom overcomes the inherent bias in observational data, it can provide more credible estimates of program impact than simple comparison of means, even when the comparison uses a d-i-d framework.²⁹

²⁹ Using regression to generate comparison of group means also has disadvantages. Because regression is sensitive to the conditional variance of a given variable, the coefficients of a regression model will not always produce identical values as those generated by manually comparing the group means. In this study most outcome measures are identical whether generated manually or through regression, but there are some cases where regression coefficients and group mean values differ by 1-3%. These slight deviations do not substantively alter any of the results.

Annex 3: General evaluation questions, by district

The following table presents [general evaluation questions](#) disaggregated by district.

Evaluation item	Logar	Kunar	Uruzgan		Kandahar		Overall
	Mohammad Agha	Chawkai	Chora	Shahidi Hassas	Zhari	Panjway	
RLS-I activities useful	95%	95%	98%	100%	92%	91%	95%
Applied RLS-I training in home community	75%	73%	100%	100%	62%	40%	73%
Received handouts at RLS-I workshops	100%	85%	100%	100%	97%	88%	93%
Handouts useful	98%	96%	100%	100%	86%	98%	96%
Consult handouts at least occasionally	91%	74%	98%	100%	67%	65%	82%
Handouts helped answer a question or solve a legal problem	58%	67%	98%	100%	54%	39%	71%
Shared handouts outside of immediate household	80%	96%	97%	95%	65%	72%	84%
Would continue to attend RLS-I activities without external support	91%	98%	11%	1%	27%	41%	48%

Note the high variance across districts for elders stating whether they could attend RLS-I workshops without external support. While the Uruzgan districts of Chora and Shahidi Hassas have been among the more dedicated elders requesting RLS-I activities, these same elders also report that movement to the district center is not feasible without help. More generally, elders from the South region and more insecure areas report that they would not attend RLS-I activities without support.³⁰

³⁰ It is difficult to surmise whether this reflects a genuine willingness to attend activities without support, but cannot due to security, against elders' unwillingness to participate without external support. In Uruzgan, elders have been the more dedicated and vocal in requesting RLS-I activities, and RLS-I staff consider their motivation to be one of genuine benefit from learning. Elders from Kandahar however, have been reported to be also interested in external support for its own sake rather than for the sake of gaining legal literacy.

Annex 4: Hypothesis 1.1 measures by district

The following table presents only the final d-i-d and spillover measures for [Hypothesis 1.1](#), disaggregated by district. District measures are expressed as percent gains relative to baseline values of the entire comparison group consisting of Narang, Khoshi, and Shah Wali Kot districts.

Measure	Chawkai		Mohammad Agha		Zhari	
	D-i-D	Spillover	D-i-D	Spillover	D-i-D	Spillover
Jirgas apply Afghan law	16%	-11%	3.4%	-10%	-3.8%	-23%
Jirgas apply <i>Shari'ah</i> law	14%	-2.4%	3.3%	-24%	-46%	-43%
Jirgas apply community norms	-8.6%	-13%	-15%	-39%	-58%	-76%
Community norms conflict with <i>Shari'ah</i>	16%	22%	2.4%	3%	42%	0.9%
Community norms conflict with Afghan law	13%	37%	29%	35%	66%	29%
Afghan constitution expresses <i>Shari'ah</i>	20%	2%	-21%	-21%	6.4%	-53%

Annex 5: Hypothesis 1.2 measures by district

The following table presents only the final d-i-d and spillover measures for [Hypothesis 1.2](#), disaggregated by district.

Measure	Chawkai		Mohammad Agha		Zhari		Chora
	D-i-D	Spillover	D-i-D	Spillover	D-i-D	Spillover	D-i-D
Elders resolve criminal aspects	-17%	-10%	-9.5%	-10%	7.8%	13%	-16%
Elders resolve civil aspects	78%	50%	8.4%	-9.8%	0%	-14%	14%
Elders resolve both	-54%	-37%	-33%	-6.1%	2.7%	20%	-11%

Annex 6: Hypothesis 1.3 measures by district

The following table presents only the final d-i-d and spillover measures for [Hypothesis 1.3](#), disaggregated by district.

Measure	Chawkai		Mohammad Agha		Zhari	
	D-i-D	Spillover	D-i-D	Spillover	D-i-D	Spillover
Documented	29%	31%	25%	25%	20%	0%
Percent documented	1.4%	-13%	18%	-1.5%	14%	11%
Registered	25%	51%	20%	42%	-63%	-14%
Percent registered	2.4%	-13%	9.7%	-12%	50%	2.7%

Annex 7: Knowledge items dropped from baseline to endline

The following knowledge items for [Hypothesis 1.4](#) were initially included but subsequently dropped from the analysis:

Learning item	Legal topic	Cause for dropping item
Elders may negotiate release of a convicted person	Jurisdiction	Formatting error in translation of baseline interview
Crime is a personal act (police may not conduct proxy arrest)	Criminal law	Formatting error in translation of baseline interview
Government jurisdiction over criminal cases according to Islamic <i>Shari'ah</i>	Jurisdiction	Formatting error in translation of baseline interview
Defense lawyer code of conduct (defense lawyer may not lie before court as part of defense)	Constitutional rights	Question too long to hold in working memory of respondent
A woman escaping her home due to abuse or unhappiness is not breaking the law	Family	Too sensitive to consistently cover in workshops
Portion of inheritance shares for daughters	Inheritance	Coding error in endline survey
Order of precedence in assigning inheritance shares	Inheritance	Question wording too abstract for elders to understand
Ownership claim after rehabilitating unowned land according to <i>Shari'ah</i>	Property	Conflicted with government claim over all lands not owned privately

The remaining 34 items aggregated by workshop topic is as follows:

Topic	Items
Constitutional rights	5
Jurisdiction	3
Criminal procedure	8
Aggregate: Constitutional and criminal law	16
Family	8
Inheritance	6
Aggregate: Family and inheritance	14
Property	4
Aggregate: Overall knowledge	34

Annex 8: Constitutional rights knowledge items

Topic	Item	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover
Constitutional Rights 1	Gender equality	16.4%	-13.7%	17.4%	-1.0%	-31.1%
Constitutional Rights 2	Freedom of speech	18.4%	-0.2%	34.4%	-15.9%	-34.6%
Constitutional Rights 3	Freedom of assembly	-4.7%	-17.8%	1.4%	-6.1%	-19.2%
Constitutional Rights 4	Presumption of innocence	-9.9%	-22.1%	0.7%	-10.6%	-22.8%
Constitutional Rights 5	Freedom from torture / forced confessions	22.2%	-1.3%	-25.1%	47.3%	23.8%

Annex 9: Jurisdiction and criminal procedure knowledge items

Topic	Item	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover
Jurisdiction 1	Elders may negotiate release of suspect	-5.1%	-14.8%	20.9%	-26.0%	-35.7%
Jurisdiction 2	Elders may apply punishment	-7.4%	-43.8%	3.5%	-10.9%	-47.3%
Jurisdiction 3	Only Afghan courts may prosecute or punish	26.5%	-10.9%	1.4%	25.1%	-12.3%
Topic	Item	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover
Criminal Procedure 1	Coerced confessions not admissible in court	-15.2%	-18.1%	-2.0%	-13.3%	-16.1%
Criminal Procedure 2	All defendants entitled to legal defense	36.5%	5.2%	21.1%	15.5%	-15.9%
Criminal Procedure 3	Detained entitled to visitation	-7.3%	4.1%	16.7%	-24.0%	-12.7%
Criminal Procedure 4	Police have 72 hours to hold a suspect without charge	25.2%	-2.0%	1.5%	23.7%	-3.5%
Criminal Procedure 5	Range of penalties for petty crimes	27.2%	17.1%	13.6%	13.6%	3.5%
Criminal Procedure 6	Range of penalties for misdemeanors	13.5%	2.9%	1.6%	12.0%	1.3%
Criminal Procedure 7	Range of penalties for felonies	36.1%	8.7%	6.0%	30.1%	2.7%
Criminal Procedure 8	Proxy arrest not acceptable	77.4%	47.3%	50.0%	27.4%	-2.7%

Annex I0: Family law knowledge items

Topic	Item	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover
Family 1	Legal age of marriage for females	20.4%	-5.7%	24.6%	-4.3%	-30.3%
Family 2	Marriage leading to suffering not valid	15.0%	5.9%	-16.7%	31.7%	22.6%
Family 3	Consent to marriage	-0.2%	-20.4%	17.1%	-17.3%	-37.5%
Family 4	Maher Islamic	-3.2%	-33.3%	-16.2%	13.0%	-17.1%
Family 5	Bride price un-Islamic	4.1%	-28.7%	-13.5%	17.6%	-15.2%
Family 6	Right to contest an arranged marriage	11.6%	22.7%	28.5%	-16.9%	-5.8%
Family 7	Baad un-Islamic	20.2%	6.3%	-10.7%	30.9%	17.0%
Family 8	Baad illegal	-5.9%	-13.7%	-20.3%	14.4%	6.6%

Annex 11: Inheritance knowledge items

Topic	Item	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover
Inheritance 1	Wife survives husband without children	26.6%	11.9%	-9.7%	36.4%	21.7%
Inheritance 2	Wife survives husband with children	14.8%	-3.7%	-22.5%	37.3%	18.7%
Inheritance 3	Husband survives wife with brother	5.2%	-5.0%	-18.9%	24.1%	13.9%
Inheritance 4	Husband survives wife with son	15.6%	-4.3%	-21.7%	37.3%	17.5%
Inheritance 5	Grandmother survives married child with grandchildren	21.2%	5.8%	-27.0%	48.2%	32.8%
Inheritance 6	Amount that may be bequeathed	16.0%	-23.8%	-9.2%	25.3%	-14.6%

Annex 12: Property law knowledge items

Topic	Item	Treatment gain	Spillover gain	Comparison gain	D-i-D	Spillover
Property 1	Pre-emption right of shareholder	-18.7%	-19.4%	-44.2%	25.5%	24.8%
Property 2	Pre-emption right of neighbor with right of access	36.0%	29.1%	-10.8%	46.8%	39.9%
Property 3	Responsibility of witness to a deed	1.3%	-6.5%	31.6%	-30.3%	-38.1%
Property 4	Splitting valid deed illegal	4.5%	-13.0%	-23.6%	28.1%	10.7%

Annex I3: Hypothesis I.4 measures by district

The following table presents only the final d-i-d and spillover measures for [Hypothesis I.4](#), disaggregated by district.

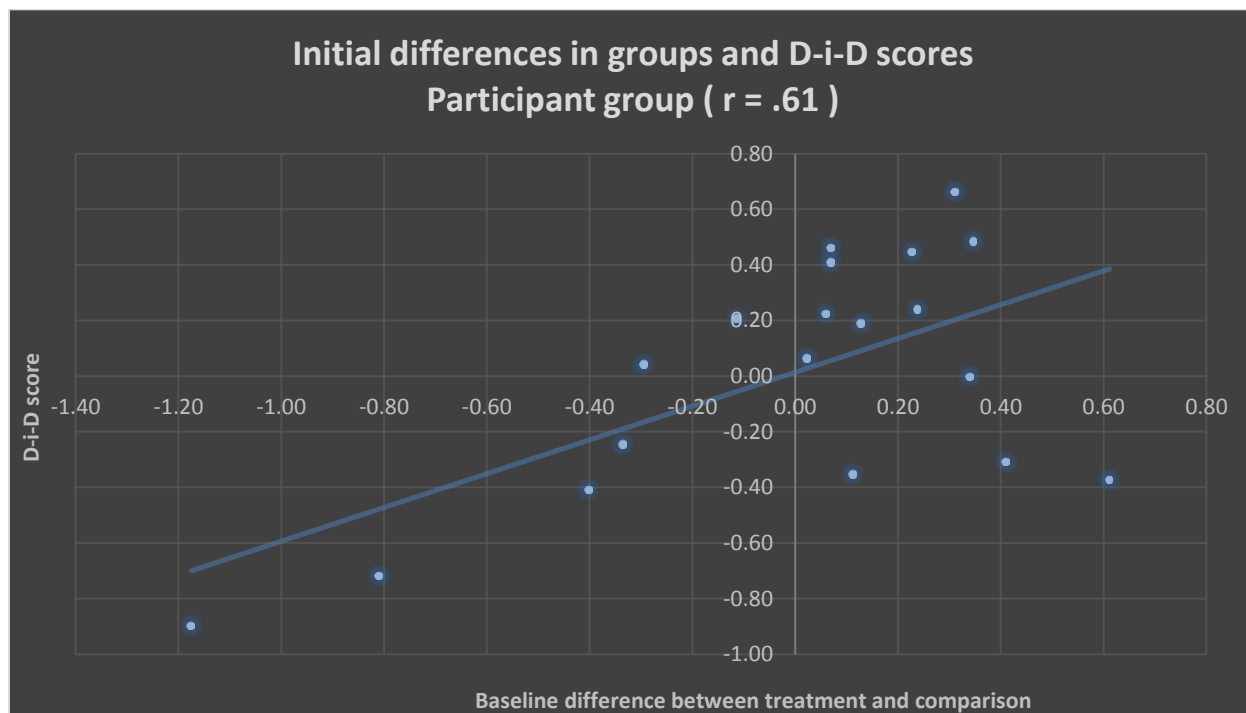
Measure	Items	Chawkai		Mohammad Agha		Zhari		Chora	
		D-i-D	Spillover	D-i-D	Spillover	D-i-D	Spillover	D-i-D	Spillover
Constitutional rights	5	14%	3.7%	-6.5%	-15%	-5.6%	-25%	23%	--
Criminal procedure	8	17%	5.7%	5.1%	2.1%	-14%	-13%	17%	--
Jurisdiction	3	22%	-3.7%	-23%	-38%	-34%	-41%	7.6%	--
Constitutional and criminal law	16	17%	3.2%	-4.3%	-11.4%	-15%	-23%	17%	--
Family	8	23%	13%	-1.1%	-12%	-8.6%	-14%	9.2%	--
Inheritance	6	41%	20%	37%	33%	0.4%	8.5%	68%	--
Family and inheritance law	14	30%	16%	15%	7.4%	-4.7%	-4.6%	34%	--
Property	4	22%	11%	13%	-1.3%	17%	11%	50%	--
All knowledge items	34	23%	9.5%	6.1%	-2.2%	-6.8%	-11%	29%	--

Note that Chora represents a treatment district of a repeated program cycle applied to a second cohort of elders, with 20 elders selected from the first cohort to also repeat the program cycle. This design does not include a spillover group for Chora district.

If the Chora result is included in the general evaluation design, the overall RLS-I knowledge gain is 17%.

Annex I4: Disputant perception outliers and alternate results measures

The following table shows the relation between the distance between [Hypothesis 2](#) impact measures at baseline on the X-axis (with greater distance an indication of incomparable groups) and d-i-d scores at endline on the Y-axis.



Note that extreme differences on the X-axis at baseline are strongly associated with large d-i-d values at endline on the Y-axis, suggesting that it is the initial differences between groups that is driving the d-i-d measure rather than any genuine dynamic. For this reason, the four most extreme outliers were removed per the following table.

Disputant perception items dropped from analysis

Removed item	Index
Decision makers sought consensus from community	Procedural justice
Decision makers consulted all relevant parties/witnesses	Procedural justice
Decision makers sought own gain in adjudicating dispute	Subversion of decision
Decision allowed reconciliation	Justice of outcome

For reference, the impact measures without removing outliers are as follows:

Participant group	Treatment gain	Comparison gain	Spillover	D-i-D
Procedural justice	0.37	0.25	-0.63	0.12
Subversion of decision	0.43	0.02	0.79	0.41
Justice of outcome	0.21	0.00	-0.42	0.21

The remaining set of 15 items used for analysis are summarized as follows, along with their reliability scores (Cronbach's Alpha):

Revised summary of items, with Cronbach's Alpha

Index	Items	Reliability
Procedural justice	6	.831
Subversion of decision	3	.731
Justice of outcome	6	.834

For the reliability scores above, values between .7 and .8 are generally recommended, though the score is also a function of the number of scale items. The reader should also be alert to the small number of items making up the corruption index. Analysis has shown that this index is not as stable as desired.

Annex 15: Disputant assessment items

Procedural fairness	Treatment gain	Comparison gain	D-i-D
I submitted to the decision making authority of this body by my own will	0.58	-0.08	0.66
Disputant able to communicate facts of case	0.56	0.35	0.21
Disputant able to communicate feelings and opinions about case	0.45	0.41	0.04
Case given due consideration by decision makers	0.34	-0.11	0.45
Case given equal consideration by decision makers	0.29	-0.20	0.49
Decision makers sought consensus among community	0.06	0.96	-0.90
Decision makers consulted all relevant parties/witnesses	0.09	0.81	-0.72
I preferred that this body decided my case	0.11	-0.12	0.23
Subversion of decision	Treatment gain	Comparison gain	D-i-D
Decision makers considered which party more powerful	-0.26	-0.26	0.00
Decision makers unwillingly influenced by outside factors	0.22	-0.08	0.31
Decision makers solicited payment to affect outcome of case	0.76	0.41	0.35
Decision makers sought own gain in adjudicating dispute	0.42	0.05	0.37
Justice of outcome	Treatment gain	Comparison gain	D-i-D
Decision makers sought consensus among themselves	0.08	0.33	-0.24
My rights respected	0.02	-0.05	0.06
Decision allowed reconciliation	-0.05	0.36	-0.41
Dispute fully resolved	0.17	-0.24	0.41
Agreed with decision	0.14	-0.10	0.24
Overall process was fair	0.14	-0.05	0.19
If I am faced with a dispute in future, I would choose this body to resolve it	0.23	-0.23	0.46

Annex 16: Dose-response results measures for disputant perception

The following discussion builds on initial analyses of [Hypothesis 2](#). As mentioned in [Annex 1](#), there are three dose-response specifications for disputants: the number of mediating elders, the number of “graduates” based on requisite levels of workshop attendance, and the total number of workshops attended per elder. Presentation of results will include the spillover group along with the range of possible effects. Measures are by percent, rather than along the 5-point scale. For the number of mediating elders, the Phase 3 disputant data averages 1.8 RLS-I elders who helped mediate a dispute, with the data spread fairly evenly across 1, 2, and 3 mediating elders.

Table 32 Percent gain in disputant perception by number of RLS-I mediating elders, D-i-D

Index	Spillover	1 RLS-I mediator	2 RLS-I mediators	3 RLS-I mediators
Procedural Justice	-9.2%	5.4%	10.8%	16.2%
Subversion of Decision	35%	6%	11.9%	17.9%
Justice of Outcome	-8.6%	3.5%	6.9%	10.4%

There are respectable gains in procedural justice and justice of the outcome. However, there is an increase in corruption perceptions. The spillover groups show decreases in the justice measures and increase in corruption.³¹

The next dose-response variable is the number of “graduates” helping to mediate a dispute, as measured by requisite attendance levels across all mediating elders.

Table 33 Percent gain in disputant perception by number of “graduates”, D-i-D

Index	Spillover	1 graduate (6 workshops)	2 graduates (12 workshops)	3 graduates (18 workshops)
Procedural Justice	-9.9%	5.8%	11.5%	17.3%
Subversion of Decision	31%	6.1%	12.2%	18.3%
Justice of Outcome	-13.4%	3.4%	6.7%	10.1%

The final dose-response variable is the number of workshops or activities attended.

Table 34 Percent gain in disputant perception at mean value of activities attended, D-i-D

Index	Workshops (mean = 9.4)	Spillover
Procedural Justice	10.2%	-8.2%
Subversion of Decision	13.4%	35%
Justice of Outcome	7.1%	-11.8%

All dose-response measurements tell the same story: mild to moderate improvement in procedural justice and justice of the outcome, and increase in corruption perceptions. The increase in disputant perceptions of corruption is addressed in the findings and conclusion.

³¹ An alternative specification leaves out a dedicated variable capturing the spillover group, and constrains its effect to be that of zero activities attended. Under this specification, RLS-I treatment effects nearly double.

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>arbaki</i>	local tribal militia
<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>
Comparison group	Group of participants that are compared to the treatment group. These participants receive the placebo or current standard treatment to provide a comparison to the treatment being studied
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>huduud</i>	specified punishments for certain crimes established in the <i>Qur'an</i>

<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>	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>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)
NRVA	Afghanistan National Risk and Vulnerability Assessment (2007-2008). A household profiling and poverty survey of over 50,000 respondents

Platform	combined civilian-military teams at Regional Commands and PRTs that allocate resources, implement integrated programs, and assess results
PPI	Progress out of Poverty Index. A poverty measurement tool pioneered by the Grameen Foundation to help microfinance organizations to measure their success in elevating individuals or households above national poverty lines.
PRT	Provincial Reconstruction Team
<i>qawm</i>	a basic unit of Afghan social structure; most often translated as tribe, but can also apply to larger or smaller groupings such as clan or ethnicity
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>shafa</i>	legal right of pre-emption; a landowner's right of first purchase of land before it is offered for public sale. Conditions for pre-emption are shared ownership of the land being sold, shared access to public resources with the land being sold (i.e., a path or water well), or shared boundaries with the land being sold
<i>Shari'ah</i>	legal precepts found in the Holy <i>Qur'an</i> and the <i>Hadith</i> ; sometimes used 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
Spillover group	elders in treatment districts who were not program participants
<i>spingary</i>	"White-beard"; respected elder of the community
<i>spinsary</i>	(literally, feminine form of "white-headed") respected female elder(s) involved in dispute resolution
TDR	traditional dispute resolution

Treatment group	elders attending RLS-I activities
USG	Government of the United States of America
<i>usul al-fiqh</i>	a body of authoritative opinions on matters of religious faith and practice; usually coincides with the different schools of Islamic jurisprudence
<i>walwar</i>	bride price; money or goods given by a groom or his family to the head of the bride's household