AgResults Evaluation Design:
Uganda Legume Seed Pilot

Submitted to:
Department for International Development
Abercrombie House, Eaglesham Road
East Kilbride, Glasgow G75 8EA

7 February 2017

Abt Associates
4550 Montgomery Ave., Suite 800N
Bethesda, MD 20814

In Partnership With:
Denise Mainville Consulting

Table of contents

List of tables

List of figures

Acronyms

DFID	Department for International Development
FOWODE	Forum for Women and Democracy
IQA	Initial qualitative assessment
IRB	Internal Review Board
ISSD	Integrated Seed System Development
KAP	Knowledge, attitudes, and practice
LSB	Local seed business
MDES	Minimum detectable effect size
MT	Metric tonne
NAADS	National Agricultural Advisory Services
NGO	Non-governmental organisation
QED	Quasi-experimental design
SCP	Structure, Conduct, Performance
USAID	U.S. Agency for International Development
WHO	World Health Organization

Foreword

AgResults is a $118 million, multi-lateral initiative promoting the development and dissemination of high-impact agricultural innovations for global food security, health, and nutrition through the design and implementation of pull mechanism pilots. It is funded by the governments of Australia, Canada, the United Kingdom, and the United States, and by the Bill & Melinda Gates Foundation, and managed through a Financial Intermediary Fund operated by the World Bank. By using pull mechanisms, AgResults goes beyond traditional aid measures to promote the adoption of innovative technologies with high-yield development impact. AgResults provides economic incentives to private sector actors to develop and ensure the uptake of innovative technologies with the potential for high development impact. It helps overcome market failures impeding the establishment of sustainable commercial markets for such technologies, or goods produced by means of them. It thereby achieves substantial and sustained development impacts, including improved food security and food safety, increased smallholder incomes, and better health and nutrition. AgResults calls upon the ingenuity and drive of the private sector to identify and execute the most effective and efficient strategies to achieve development outcomes.

The AgResults team comprises a Steering Committee, a Secretariat, a Trustee, country-specific pilot managers, and an External Evaluator. The Steering Committee oversees the implementation of AgResults and is composed of the five donor agencies and the Trustee. The Steering Committee is responsible for strategic oversight of the initiative, including endorsement of key management decisions, approval of concepts and business plans for proposed pilots, and monitoring of pilots and the initiative as a whole. The Secretariat is responsible for implementation of the initiative and reports to the Steering Committee. In order to fulfil its role effectively, the Secretariat maintains a close working relationship with the Trustee and External Evaluator. Core functions include appointing and managing pilot implementation and verification agents, developing and sourcing new pilots, and communicating results. As Trustee for AgResults, the World Bank provides an agreed set of financial intermediary services that include receiving, holding, and investing funds, and transferring them to recipients or other agencies for implementation as directed by the Secretariat on behalf of the Steering Committee.

The Steering Committee appointed Abt Associates Inc. to serve as External Evaluator for the AgResults pilots. Abt's role is to use rigorous scientific tools to determine if the pull mechanisms achieve their objectives—to measure whether they produce private sector behaviours and social outcomes different from, and better than, what would have happened in the absence of the AgResults initiative. In our role as the External Evaluator, Abt defines the overall evaluation framework for the AgResults initiative and an impact analysis strategy for each pilot. Through a local survey contractor, Abt implements and analyses field surveys based on established best practices, conducts qualitative market analyses, and communicates evaluation findings to the Steering Committee and wider audiences as needed. Our role will be vital to the AgResults learning agenda of understanding the potential of private sector involvement in the development and spread of agricultural innovation. We will report our assessment of the sustainability of the results produced in the private market once the pilot incentives are removed.

This report presents Abt's evaluation design for the Uganda pilot. Abt's team for the AgResults Uganda evaluation is headed by Betsy Ness-Edelstein, who has worked closely with Dr Tulika Narayan on overall research direction and adherence to the theoretical framework for the evaluation. Dr Denise Mainville is the qualitative lead. Miriam Kotalimye serves as our Uganda-based agricultural economist. Dr Abigail Conrad provided inputs for the evaluation approach. Dr Judy Geyer is the quantitative lead. Dr Stephen Bell provides quality assurance on quantitative aspects of the evaluation.

1. Setting for the AgResults pilot

1.1 Overview

This report presents our plans for a mixed-methods evaluation of the AgResults legume seed pilot, which aims to increase the production and sale of improved common bean and soybean seed in Uganda. We begin in Section 1 by introducing the causes and consequences of the problems underlying the missing market for improved legume seed in Uganda. We explain how the pilot seeks to address these issues, thereby catalysing the development of a sustainable private-sector-driven market for improved bean and soybean seed. We also present relevant information on the pilot and markets context that influence our evaluation design. Section 2 presents the evaluation questions and briefly describes our overall evaluation approach as well as our specific strategy for answering each question. Section 3 details the methodology, data, and analysis plan that we intend to employ for each evaluation question. Section 4 concludes with the timeline, and our plans for quality assurance, ethical standards, and risk identification and mitigation.

1.2 The "missing market" for improved legume seed in Uganda

Smallholder farmers' constrained access to improved legume seed is broadly recognized as an important factor limiting production of legumes, and the nutritional, financial, and soil-health benefits that increased production could bring to Ugandan smallholder farmers (e.g. Monitor Group 2012, MISEREOR 2012, DIMAT 2012, Ashour 2015, and Mabaya 2015). Uganda has a dynamic private-sector-driven seed sector, with numerous national-level companies competing for market share. Although the vast majority of seed companies' activities are directed to the hybrid maize seed market, many of the companies produce at least some legume seed on a commercial basis. Nonetheless, a market assessment conducted while preparing the legume seed pilot's business plan (AgResults 2014) revealed that commercial seed companies consistently produced less improved legume seed than they had demand for.

Improved legume seed is billed as such because it is higher yielding than the local varieties that most Ugandan farmers produce. Throughout this report, "improved legume seed" is used to refer to seed that was multiplied from breeder seed by seed companies or outgrower farmers and packaged and sold as improved seed. However, Uganda has rampant counterfeiting and adulteration in its agricultural inputs industry. As a result, much of the seed that farmers purchase as "improved" actually is not. Some proportion of the seed sold each season is non-improved or expired seed that has been packaged by third-party counterfeiters and sold as "improved". And some proportion of seed companies' own sales into the market is adulterated: they often mix improved seed with grain to fulfil demand.[1] This problem of adulteration and counterfeiting undermines any attempt to estimate the penetration of improved seed in Uganda based on sales figures alone.

A number of factors underlie this apparent "market failure" (i.e., the failure of a dynamic private-sector-driven market to emerge that meets existing demand for improved legume seed). Research that the evaluation team conducted in 2015 (Mainville 2015) identified the lack of an effective quality certification system for seed as the primary issue limiting the development of a market for improved legume seed. This issue, which is largely due to the extremely limited capacity of the National Seed Certification Service, leads to rampant counterfeiting and a disincentive for seed companies to maintain quality. The inability to certify the quality of improved legume seed effectively makes the legume seed market a "market for lemons". In other words, there is no cost-effective means of differentiating

[1] This was reported by multiple seed companies and other stakeholders during interviews in April 2015.

between good and poor quality seed on the market, which depresses the price that consumers are willing to pay for seed in general. This forces suppliers to move even further into low quality seed production at a cost below the depressed market price (Akerlof 1970). For the same reasons, seed companies' products also tend to lack purity and often combine improved legume seed with grain because of the lack of a market mechanism by which consumers can tell the difference. This problem is exacerbated by the Ugandan government's large-scale purchases of legume seed from lowest price bidders, which dominate the market in terms of volume while creating intense price pressures, further undermining incentives to produce quality seed.

The legume seed pilot's business plan (AgResults 2014) highlights several additional factors inhibiting seed companies' investments to expand their production of quality legume seed. First, demand is highly variable from year to year, and orders are rarely made in advance of production. This keeps seed companies from developing realistic demand forecasts on which to base their production decisions. Instead, they make conservative production decisions to mitigate the risk of producing more seed than they can sell and thus losing money. Second, financing constraints keep seed companies from expanding their production of seed. They have lower interest and ability to invest significantly in the legume seed market given the trade-off between the legume seed market and the larger, more profitable hybrid maize market.

In the context of these challenges, AgResults' legume seed pilot, which will be implemented from 2017 to 2021, will employ an innovative development approach—pull mechanisms—to catalyse the emergence of a dynamic market for improved legume seed in Uganda.

1.3 Pilot design and theory of change

The AgResults initiative aims to offer the right incentives to the right market actors to overcome the most crucial bottleneck preventing a well-functioning private sector market. In Uganda, AgResults focuses on seed companies, offering them ex-post prizes for increasing the quantity of improved (and quality-verified) legume seed they sell each year for five years. However, given the additional challenge of the "lemons market" in the Uganda case, the AgResults team concluded that a pull mechanism alone would not be sufficient to substantially grow the market for improved legume seed. To remedy this, AgResults has added a quality verification component in addition to providing a pull mechanism. This quality verification will be provided by an external partner, AgVerify, which is launching in Uganda to provide third-party agricultural input quality verification services. Companies participating in AgResults will be required to obtain AgVerify certification in addition to Uganda Seed Certification Service certification for their sales to be counted under the pilot. In this section, we describe the components of the AgResults theory of change—including this verification piece—in further detail.

1.3.1 Design of the pull mechanism

Pull mechanisms provide incentives for private sector actors to work creatively to achieve development goals. Unlike traditional approaches to development problems, which tend to rely on grants, loans, and technical assistance to "push" stakeholders and beneficiaries down pre-determined paths to desired outcomes, pull mechanisms are agnostic with respect to the specific inputs and processes that stakeholders choose to adopt. Rather, they reward achievement of pre-defined results without preference for strategies and technologies involved in achieving those results.

In the Uganda legume seed market, the pull mechanism will be available to qualifying national seed companies ("solvers", in AgResults parlance) that engage in the production and marketing of improved (and AgVerify certified) legume seed. The pull mechanism will create incentives for companies to expand their production and sales of improved legume seed by mitigating the risks of aggressively expanding production and by motivating participating seed companies to take a long-term view of their investment in the market. This

long-term outlook will be particularly important given the likelihood that much of the anticipated long-term demand for improved legume seed is currently latent. The reasons for the latency of demand are the limited awareness of many farmers about the benefits of improved seed, as well as the poor results that many seed buyers have had in the "lemons market". Seed companies will need to develop strategies to translate latent demand to effective demand, which will require sustained investment in their marketing and promotion, among other areas. The pull mechanism design in Uganda includes the following:

- **Annual prize proportionate to increases in sales of improved seed:** Cash payments proportional to increases in sales relative to the baseline period are intended to increase returns to expanding sales as well as enable re-investment in the market; for example, supporting expenditures on marketing, packaging, staffing, and certification. There is an 8% sales growth threshold to be eligible for the annual prize, and the prize is capped to incentivize growth only up to the equivalent of 20% compound annual growth rate over the life of the pilot.
- **At-cost provision of refrigerated storage:** This component is intended to reduce costs and risk of aggressively expanding production by facilitating carry-over of stocks to the following season.[2]

In addition, to counteract the "market for lemons" problem, AgResults is partnering with AgVerify to act as a third-party quality certification system. AgVerify, currently being launched, will not take the place of the Uganda National Seed Certification Service but will provide independent quality assurance for seed via field inspection, laboratory testing, and training for companies who pay to use its service. Products certified by AgVerify, including seed, fertilizer, and other agricultural inputs, will bear a label attesting to their quality. Furthermore, in the future seed companies may also attach an "e-verification" label allowing purchasers to verify that the product is genuine. Purchasers may send a short code message with a unique code found on the label to a central database, which will then return a message indicating whether the product is genuine as well as its expiration date. AgResults is subsidizing the companies' use of the AgVerify service for legume seed at first and plans to phase out that support over time. Eventually, AgVerify aims to be funded entirely by the input supply companies themselves.

The pilot's theory of change (Figure 1-1) posits that, when there is a means to verify seed quality (the AgVerify initiative), incentives that increase the rewards of marketing improved legume seed and mitigate the risk of investing in the market for improved legume seed will lead national seed companies to invest more in the legume seed market. The AgResults-supplied prizes will temporarily reduce the profitability gap between legume and maize seed. Assuming that AgVerify certification gives seed companies access to higher-paying and more discriminating (quality-conscious) markets for improved legume seed, those companies will be motivated to increase their production of improved legume seed. They will also be motivated to invest in developing their long-term position in the market; for example, through establishing effective and sustainable distribution channels for improved seeds, and marketing their own particular brand.

As awareness and availability of quality-verified improved legume seed increase, the existing unmet demand for quality legume seed will increasingly be met. Furthermore, seed companies' marketing investments and efforts to develop distribution channels, coupled with the effects of witnessing the benefits of improved legume seed among neighbours, should

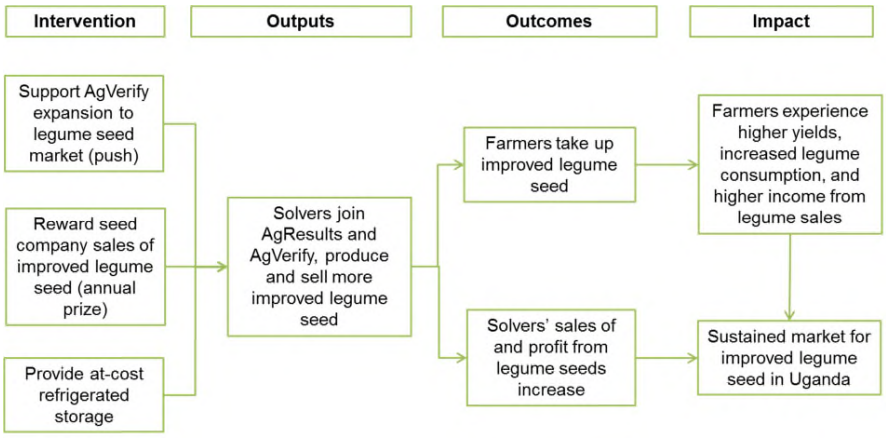
[2] This component may be described as a 'push' element of the design. The original pilot design would have had the Pilot Manager dispose of unsold seed after each season to prevent it from spoiling and being sold despite being spoilt. The provision of cold storage achieves the same goal without having to destroy seed.

lead farmers who did not previously participate in the formal legume seed market to develop demand for improved legume seed as well. As increasing numbers of farmers procure (either through their own purchases from commercial suppliers or from institutional sources of seed such as development projects) the seed, plant it, and realise its yield and profitability benefits, then a sustainable market for improved legume seed should emerge. Because legumes are self-pollinated, they can be re-planted for up to three years before they need to be refreshed; even so, the pilot's design expects that continuous introduction of new varieties will sustain farmers' demand for new legume seed.

The potential development impacts for farmers include increased yield, higher income from legume sales, and greater soil health (legumes are nitrogen-fixing crops, improving soil fertility). Under ideal conditions, yield increases could be as high as 40% (AgResults 2014), though many factors affect yield, so the actual improvements could be lower. Yield increases will likely translate to higher legume-related income.

While legumes are an affordable source of protein, we expect the nutritional benefits of the pilot to be modest. Most Ugandans already consume beans several times per week, and improved legume seed offers no gain in nutrients over local varieties. Wealthier households tend to express a preference for reducing the frequency with which they consume legumes in favour of animal proteins. The pilot may allow wealthier households to consume more fish and meat because of increased income from legume sales. Poorer households tend to consume legumes fairly often as well, eating beans at least a few times per week, but they may have gaps in legume consumption on a seasonal basis or consume lower quantities of legumes at meals than they would prefer. For these households, there may be some increase in legume consumption. Soybean, on the other hand, is not a significant part of the local diet in its unprocessed form, and as such we do not expect to see a nutritional impact from soy.

Figure 1-1. Uganda legume seed pilot theory of change[1]



[1]The AgResults Secretariat's officially stated goals for this pilot extend only to the point of increasing the availability (not adoption by smallholders) of improved legume seed in Uganda. However, in keeping with the overall objectives of the AgResults initiative and the specific evaluation questions we are charged with answering, this evaluation traces out the theory of change beyond seed availability. Our theory of change posits that smallholder adoption and subsequent benefits from improved legume seed are in fact integral components in the success and sustainability of the pilot.

Several key assumptions underlie the pilot's theory of change:

- **A sufficient portion of the improved legume seed will remain in Uganda (as opposed to being exported) for smallholders to represent the main group of end-users.** The Uganda pilot does not explicitly require participating solvers to market or sell their seed to smallholders, nor does it restrict seed exports. There is therefore the potential that some seed will be exported or sold to relief agencies that export it for distribution in neighbouring countries facing humanitarian crises, such as South Sudan or others.
- **AgVerify certification will ensure a high-quality product and consumers will trust that certification.** AgVerify has a robust plan to determine and certify the quality and legitimacy of improved legume seeds in Uganda. However, as a new programme it is untested. We cannot be certain ex-ante that it will not encounter some of the same pitfalls that the National Seed Certification Service has encountered – in other words, it may still be possible for quality issues to persist. This could happen if AgVerify's initiative becomes overwhelmed by volume, if its funding falls short of what is needed to provide comprehensive seed inspection and testing, or if corruption or tampering occur. Furthermore, AgVerify has the difficult job of earning consumers' trust in a country where counterfeiting is rampant. Farmers must trust that the higher price associated with AgVerify certification is worth paying. The pilot assumes that AgVerify and/or individual seed companies will find ways to convince farmers that AgVerify certification is worth paying for.
- **There truly exists significant latent demand for improved legume seed among smallholders.** Ugandan smallholders are largely unaccustomed to purchasing legume seed from the formal market. Purchasing seed from private sector companies would be a significant shift in behaviour for most farmers, who are used to saving their own seed, purchasing it locally from neighbours or the grain market, or receiving free or subsidized seed from non-governmental organisations (NGOs) and the government. Furthermore, particularly with regard to common beans, farmers cite strong taste preferences for the beans they grow using their own saved seed. This may be less of a challenge for soybeans, which are not commonly grown for consumption but rather are produced commercially in most cases, which means there is a profit incentive for farmers to invest and increase their productivity.
- **Smallholder buyers (as opposed to institutional buyers) will represent enough of the market that companies will sustain their investment in high-quality seed.** Large institutional buyers currently dominate the Ugandan legume seed market, purchasing from seed companies in bulk and then distributing the seed to farmers at no or subsidized cost. They do not tend to demand high quality of the seed they purchase and in fact purchase from the cheapest seller, which they may continue to do despite this pilot. In that case, seed companies may continue to market lower-quality seed to institutional buyers in addition to selling AgVerify-certified seed as long as it is profitable, muddling branding efforts and diverting seed company resources back to the status quo. This scenario would represent a challenge for the long-term sustainability of AgResults because it would keep open the channels through which seed companies profitably market lower-quality seed. If that occurs, the AgVerify-certified seed market may be too small to be sustained.

1.4 Context

This sub-section describes several key facets of the context in which the Uganda pilot will operate. These include both the legume and legume seed value chains as they currently function, the presence of other initiatives promoting the use of improved seed in Uganda, and gender issues.

1.4.1 Description of the legume value chain

Legumes are a traditional food staple in Uganda and are critical to the country's food security given that an average of 76% of Ugandans' protein comes from non-animal sources such as legumes (FAO 2015). There has also been growing domestic and regional demand for legumes for industrial food processing and animal feed, and Uganda has seen marked increases in net imports of soybeans over much of the past decade. While annual production of beans has grown over most of the past decade, this growth has been driven by expansion of acreage. Yields have stagnated and even declined. Meanwhile, soybeans have shown stagnant production, yields and acreage, despite the demand evidenced by consistently increasing imports.

Beans are produced throughout Uganda, with the Western region leading and the Northern region second in terms of national acreage, production, and productivity. Areas cultivated are quite small, averaging 0.3 hectares per bean-producing household per season in the Western region, 0.15 in the Northern region, and less than 0.1 hectares in the rest of the country (Uganda Bureau of Statistics 2010).

Soybean production is dominated by Northern region, which accounts for 72% of acreage and 67% of output. Farmers in the Northern region cultivate just over half a hectare each on average, which is more than 10 times the average area cultivated of the second-most important soybean region, Eastern. Despite the significant area under soybeans in Northern Uganda, yields are relatively low at 0.6 MT/hectare compared to 0.9 achieved in Western and 0.8 achieved in Eastern Uganda. Overall, soybean is a relatively minor crop in terms of acreage and output, accounting for only 5% of the area dedicated to bean production (Uganda Bureau of Statistics 2010).

Regional patterns in production and market access may affect seed companies' investments in developing distribution networks for improved seed. Such patterns may also affect development initiatives' involvement in seed and legume production and market development in different parts of the country, as well as farmer demand for improved seed and the responsiveness of production to improved seed across agro-ecological zones. Legume production is commonly undertaken as part of an intercropping system, with complementary crops varying regionally but including bananas, coffee, and maize. There is little use of improved inputs including seed and fertilizer, and yields tend to be low. Generally, few large-scale commercial farmers specialise in legume production due to sensitivity to production conditions that cause large yield fluctuations and make large-scale production risky from a commercial standpoint. Nonetheless, the profitability of legume production has been shown to approximately double under production systems using improved seed and fertilizer (aBi Trust 2014). This implies that significant gains are possible through improved legume seed.

Beans are traditionally produced as a food staple, with small-scale producers selling surplus production in local markets, where it is frequently aggregated along multiple stages of intermediation before being retailed in either domestic or regional export markets such as Rwanda, Kenya, and South Sudan. This implies that smallholder farmers who experience improved yields as a result of planting improved bean seed may benefit from improved food availability, improved incomes due to sales of surplus, or some mix of those two benefits.

In contrast to beans, soybean is produced primarily for sale rather than consumption (although soybeans are consumed in some areas as a snack food). It is much less widely grown in Uganda than beans – the most recent Agricultural Census (2008-09) reported that Uganda produced 929,278 metric tons of bean but only 23,609 metric tons of soybean that year. There is strong demand for soybean for domestic processing and for export to neighbouring countries such as the Democratic Republic of the Congo and South Sudan. The development of oilseed processing plants has increased demand, and more recently the depletion of *mukene* (a small fish used for protein in animal feed) in Lake Victoria has increased demand for soybeans for animal feed production. There is growing demand for

soybean for industrial processing into oil, flour, baby foods, and supplements. There is also demand for soybeans for use in supplemental feeding and nutrition programmes by organisations such as the World Health Organization (WHO). This implies that farmers who experience yield benefits of planting improved seed are likely to have improved incomes as a result of selling more soy.

Offtake markets for legumes are fragmented with little advance coordination of purchases. Some large-scale buyers have attempted to strengthen their supply base by contracting farmers to produce legumes on their behalf; however, these efforts have characteristically failed due to opportunistic side-selling on the part of farmers. Market contracts are relied on to a lesser extent. These entail less risk than production contracts because they do not involve advance provision of inputs or financing, the value of which is forfeited when there is side-selling. However, high rates of contract default on both the buyer and supplier sides are reported, as price movements in the market undermine incentives to adhere to the contracts and the legal framework for contract enforcement is weak. Commercial opportunities are broadly recognized to increase farmers' uptake of productivity-increasing inputs and management practices, and the presence of dynamic offtake markets in Uganda is associated with an increasing demand for improved legume seed, particularly among commercially oriented farmers. This implies that uptake of improved seed is likely to be less than it would be if functional contract farming systems were in place.

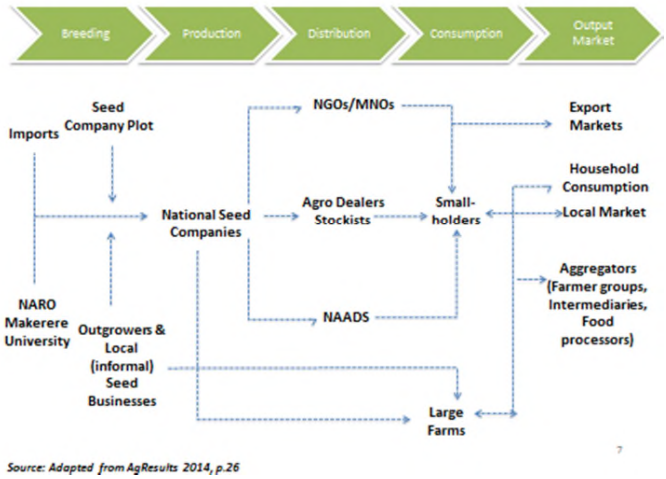
1.4.2 Description of the legume seed value chain

Figure 1-2 depicts Uganda's legume seed value chain. The first stage in the value chain is breeding, which includes the provision of breeder seed[3] through a public institution (the National Agricultural Research Organisation [NARO] or Makerere University) or from imports. During production, foundation seed[4] is then multiplied either on seed companies' own plots or by outgrower farmers and then certified by the National Seed Certification Service. Seed then reaches farmers through institutional channels (NGOs and government) or through commercial channels (either agro-dealer networks or sales direct to farmers). The graphic also depicts local seed businesses (LSBs), which are currently informal and produce seed either as outgrowers to national seed companies or for direct sale to farmers, and the role of commodity grain markets not only as a source of derived demand for legumes, but also as a source of seed (from grain).

[3] Breeder seed refers to seed that is directly controlled by the originating plant breeding institution.

[4] Foundation seed refers to the progeny of breeder seed.

Figure 1-2. Uganda's legume seed value chain



Source: Adapted from AgResults 2014, p.26

As shown in Table 1-1, farmers source legume seed almost exclusively from informal channels, including saved seed from their own production, purchases from friends and neighbours whose fields they can observe in production, and grain markets. Seed obtained through these sources may be of improved varieties but has been recycled many times (replanted season after season), implying that the purity and viability of the seed may have degraded. Farmers may also obtain legume seed from institutional sources such as government distributions or NGOs and relief or development projects. There is also limited distribution of legume seed (particularly bean seed) through commercial agro-input dealers, and some farmers source directly from emerging LSBs or national seed companies.

Table 1-1. Farmers' seed sources by crop (%)

	Maize	Beans	Soybean
Farmer-saved seed	54	46	30
Neighbour	4	2	9
Local market	30	43	30
LSB	3	2	5
Agro-dealer	5	2	15
Government	2	3	1
Project NGO	3	2	9
Total	**100**	**100**	**100**

Source: Integrated Seed System Development [ISSD] Uganda 2014.

There are 17 active national seed companies registered in Uganda. These seed companies concentrate their activities primarily in the highly profitable market for hybrid maize, and undertake secondary activities with products such as vegetable and legume seed. Seed companies tend to have weak management systems and financial footing, which limits their capacity to behave proactively in the market (Tetra Tech 2015).

National seed companies procure "certified seed" from three main sources—their own production, outgrowers (including local seed businesses), and the grain market ("standard seed"). Due to capacity limitations, there is limited oversight of seed production by the

national certification agency such that it is possible and common for these companies to market "standard" seed as "certified". Certified legume seed from national seed companies reaches farmers through two primary distribution channels—institutional and commercial. Institutional channels include government and NGO purchases. Commercial channels include distribution through decentralized agro-input dealers and direct sale to farmers. Historically, institutional channels account for approximately 70 to 80% of seed companies' sales. Purchasing institutions distribute seed to farmers for free or on a subsidized basis. NGOs' seed procurements support their technical programming, as well as seed distributions to countries that are recipients of their relief or development activities, such as South Sudan or Somalia.

There is also a small contingent of emerging local seed businesses, often based in farmers' organisations or cooperatives. These businesses are heavily involved in production of open-pollinated varieties of seed, including legumes. LSBs are the outcome of donor efforts (particularly the Integrated Seed System Development initiative) to establish a Quality Declared seed system (Mubangizi, Nandagire Ntamu et al. 2012). However, Uganda does not yet have a seed policy that will permit LSBs to legally sell seed. There are currently 30 LSBs in operation, with 70 more planned. LSBs currently operate with a significant degree of donor support, which includes facilitated access to foundation seed, business and seed production training, and oversight. They market their seed to local farmers (without formally labelling it as seed) or to seed companies for which they serve as outgrower seed producers. Key informant interviews during the initial qualitative assessment (IQA) revealed that LSBs and farmers planting LSB seed receive premium prices from neighbouring farmers who are confident of its quality because they can see the plants in the field (Mainville 2015). This implies that there may be a ripple effect of farmers who plant improved legume seed receiving premium prices for their output. Farmers may receive premium prices when they sell their output to neighbouring farmers (who buy it to use as planting material), rather than selling it on the grain market.

The Ugandan government purchases legume seed through the National Agricultural Advisory Services (NAADS) for large-scale distribution through initiatives such as its current programme, Operation Wealth Creation. Beans and soybean are priority crops under this initiative, and the procurements have a large impact on the market—currently, due to the government's Operation Wealth Creation, NAADS is estimated to account for 90% of seed purchases in the market. Accordingly, seed companies report their intention to increase their production in anticipation of forthcoming tenders.

Institutional (government and NGO) purchases typically take place as the result of tenders on which seed companies submit bids. Tenders are difficult for seed companies to predict and plan for, involve tight cost competition, and typically require rapid turnaround (e.g., a month) between award of the tender and actual delivery of the seed. Tenders are awarded on a cost basis and merely require that "certified seed" be delivered, opening the door for companies to economize on costs by packaging standard seed as certified[5].

Commercial distribution of seed accounts for a much smaller proportion of certified seed sales (20-30%). It includes sales from national seed companies directly to farmers as well as sales through agro-input dealers located in production areas.

1.4.3 Other nationwide initiatives promoting improved legume seed

In addition to the previously described AgVerify initiative (which grew out of the Agricultural Inputs activity of the U.S. Agency for International Development [USAID]) and the Operation

[5] There are few exceptions to this generalisation. The International Red Cross, for example, is known to undertake germination testing of the seed it procures in an effort to ensure acquisition of quality seed.

Wealth Creation initiative described in the previous section, there is substantial donor involvement in Uganda's legume and legume seed sector. These initiatives could influence farmers and other value chain players' involvement in seed and legume markets (Mainville 2015).

1.4.4 Gender issues in the pilot context

The identification and assessment of gender effects of the AgResults pilot requires a socially contextualized understanding of the many ways that gender may influence, or be influenced by, the pilot. Gendered lines of analysis relevant to the evaluation include gendered household structure[6], intra-household gender dynamics, and gender-differentiated activity in the legume value chain. In the following section we discuss these, and how they may influence or be influenced by the pilot and its outcomes.

Women are integral to agricultural production, particularly production of staple crops, in Uganda, and this participation has been increasing over time (Forum for Women and Democracy [FOWODE] 2012). Overall, approximately 70% of smallholder farmers are women, and women are responsible for 90% of the country's food production in contrast to only 50% of cash crop production (USAID 2014). This implies that in order for the pilot to have a marked impact on legume production—particularly designated for household consumption rather than market—then women will be an important beneficiary group. Nonetheless, as discussed below, the reality of women smallholders, the gendered structure of households, and intra-household dynamics, may mean that women and women-headed households are less likely to benefit from the increased availability of improved legume seed that the pilot may catalyse.

Women smallholders face particular constraints relative to men. They have lower levels of education and more limited exposure to media (Uganda Bureau of Statistics 2012), both of which limit their access to and ability to use new information. Their lack of land and other assets and their curtailed participation in markets limits their ability to get credit for inputs such as improved seed or for other uses (FOWODE 2012). They also have less access to government programmes facilitating access to finance or inputs, and are slightly less likely than males to access government extension programmes (FOWODE 2012). As a result of these and other constraints, women tend to have lower yields than men for the same crops (Peterman, Quisumbing et al. 2010). These realities are demonstrated in Table 1-2, which shows women-headed households' more limited access to land, specifically for legume cultivation. The widespread constraints faced by women smallholders imply that they may have lower access to improved legume seed that may be made available as a result of the pilot. At the same time, however, it is unclear how increased access to improved legume seed would affect production, unless women were also exposed to complementary information and inputs needed to realise the yield gains of improved seed.

[6] Gendered household structure refers to the structure of households with respect to gender—that is, the presence of men and women in different roles such as head of household. Gendered household structure has implications for the household's access to assets and opportunities, and consequently for outcomes (such as incomes or poverty) relating to these (Peterson and Lewis 1999).

Table 1-2. Land area (acres) under production of maize and legumes by male- and female-headed households

	Maize	Beans	Soybean
Overall	1.16	0.96	1.51
Female-headed households	0.76	0.59	0.91
Male-headed households	1.29	1.12	1.59

Source: ISSD Uganda (2014).

Approximately 30% of households in Uganda are female-headed (Uganda Bureau of Statistics 2012). Similar to female smallholders, female household heads are less educated than male household heads, have less access to capital, and are less likely to have access to government extension services. Female-headed households have on average less than half the cultivatable land that male-headed households have. Their limited access to productive assets means that female-headed households have higher concentrations of poverty, and their limited income-earning options often imply that they suffer asset depletion in their efforts to meet basic needs. Female-headed households' agricultural production is more oriented to meeting their household consumption needs. Female-headed households tend to dedicate relatively less land to cash crop production than male-headed households, and market less of their crop (FOWODE 2012). This implies that female-headed households may be less likely to use improved seed due to lower awareness and access. This is because they are less integrated into the market economy and have lower exposure to other sources of information or resources such as media and government programmes.

Most women in Uganda live in male-headed households. Women in male-headed households generally benefit from greater access to productive resources and the products of those resources, though they are commonly dependent on their male kin for access to those resources, lack decision-making authority over their use and disposition, and risk loss of those resources in cases of death or divorce (FOWODE 2012; Peterman, Quisumbing et al. 2010). This implies that women in male-headed households who are responsible for legume production may not have the decision-making influence to direct household expenditures to improved legume seed. This may be particularly true if the legume production is to be dedicated to household consumption (as is more common for beans).

The following section introduces our overall approach to answering the evaluation's seven questions.

2. Evaluation questions and overall research methods

This evaluation will assess whether the pilot has met its objectives and, as with our evaluation of each AgResults pilot, it will seek to answer the following questions. Evaluation Question 4 is not relevant for this pilot because the pilot is designed to shift the supply curve for improved legume seed, rather than the demand curve for legumes.

Table 2-1. Evaluation questions and approaches

#	Evaluation question	Evaluation method
1	What has been the impact of the pilot on private sector involvement in the development and uptake of improved legume seed?	Mixed methods using Structure, Conduct, Performance (SCP) conceptual framework; results from Evaluation Questions 2–3
2	What has been the impact of the pilot on smallholders' uptake of improved legume seed?	Performance evaluation using pre-post quantitative data complemented by qualitative inquiries using key informant interviews and focus group discussions
3	What has been the impact of the pilot on smallholders' incomes?	Performance evaluation using pre-post quantitative data complemented by qualitative inquiries using key informant interviews and focus group discussions
4	What has been the impact of the pilot on poor consumers' demand for legumes and derivative products?	n/a
5	What evidence exists that the effects of the pilot will be sustainable in the medium to long term?	Combination of SCP, qualitative smallholder interviews, and demand analysis
6	What is the evidence on the scale of any effect on private sector investment and uptake, and on the cost-effectiveness of the pilot as an approach?	SCP, with focus on market structure and per-unit cost-effectiveness of key outcomes
7	What lessons can be learnt about best practices in the design and implementation of pull mechanisms?	Synthesis of results from Evaluation Questions 1–6; Compilation of results from all AgResults pilot evaluations

The evaluation approach is designed to test the assumptions underlying the theory of change for the AgResults legume seed pilot using mixed methods, with tailored approaches for each evaluation question. Major themes of inquiry will include how impacts may vary by location, gender, socio-economic profile, and relative access to improved legume seed suppliers as well as the role of the AgVerify e-verification scheme in facilitating the pilot.

To assess the pilot's impact on the development of the improved legume seed market, including its sustainability, our approach is guided by the Structure, Conduct, Performance (SCP) framework. SCP is a theory-based approach to value chain or commodity systems analysis. The SCP framework links the underlying characteristics of a market to the strategic decisions that market players, including firms, farmers, and consumers, make about whether and how to engage in the market, given their perspectives on the underlying market conditions. The strategic decisions of numerous firms give rise to the market structure, which includes the numbers and characteristics of market participants, the predominant marketing channels, and modes of product transformation and value addition. Together, these factors affect the performance of the market, including such considerations as whether the market for improved legume seed expands and whether it benefits nutritionally vulnerable consumers.

To evaluate the impact of the pilot on smallholders (specifically, on their uptake of and income from using improved legume seed), we will complete a performance evaluation using both quantitative and qualitative data. We will also perform a seed quality assessment, detailed in the next section, to account for the unknown (and likely very low) level of quality of seed sold as "improved" in Uganda. In overview—with further technical detail to follow—our approaches to the various Evaluation Questions are as follows:

To answer Evaluation Question 1 on the impact on private sector involvement in the market for high-quality legume seed—we will apply the SCP framework to develop and test qualitative hypotheses regarding the effects of the pilot incentive on private firms' perceptions of, participation in, and outcomes in the market for high-quality legume seed. We will collect data for Evaluation Question 1 using key informant interviews and small-sample surveys of actors along the legume value chain and will also draw heavily on pilot monitoring data on improved legume seed production and sales and data collected from farmers to answer Evaluation Questions 2 and 3. Our SCP analysis will differentiate results on the basis of the type of legume seed being targeted, private sector actors, and region, and will include a nuanced analysis of the relative roles of the AgVerify certification scheme and the AgResults initiative in the development of the market for improved legume seed over the pilot period. We will collect data at all levels of the value chain and from diverse market actors including seed companies, agro-input dealers, farmers, and sector experts in government and the development community.

To answer Evaluation Questions 2 and 3 on smallholder uptake of improved legume seed and subsequent income impacts—we will employ a performance evaluation using a pre-post evaluation design. We will supplement this with qualitative research on farmers' knowledge, attitudes, and practice (KAP) to understand smallholders' conduct and decision making with regard to legumes and a national seed quality assessment. Results from these investigations will also inform Evaluation Question 1.

To answer Evaluation Question 5 on the pilot's sustainability—we will draw on results of Evaluation Questions 1 through 3 to examine the sustainability of the pilot's impact. In particular, we will examine whether conditions are right for the market developments the pilot stimulate, if any, to continue after cessation of the direct pilot incentives; that is, whether the preconditions for a sustainable market have been established or not. Qualitative contributions to the evaluation of sustainability will come from the SCP and farm-level analyses, and will focus on whether the basic conditions that provide incentives for continued private sector and farmer engagement in the market are present. We will also conduct a "sustainability" survey several seasons after the end of the pilot to assess whether smallholders are continuing to purchase and plant improved legume seed.

To answer Evaluation Question 6 on the pilot's scale and cost-effectiveness—we will determine the scale of improved legume seed activity in the market. In particular, we will investigate the extent to which the improved legume seed market is truly national and spans a range of buyers or whether companies have instead made "boutique" investments and limited their efforts to specific market segments. We will assess the cost-effectiveness of the pilot at endline when the total project costs are known, by estimating the cost of the pilot per unit of impact (smallholder uptake and income changes) measured in response to Questions 2 and 3.

To answer Evaluation Question 7 on best practices and lessons learnt—we will synthesise results from Evaluation Questions 1 through 6 to determine where the pilot intervention worked well and where it did not. Using a common framework across all AgResults pilots, we will also identify and draw lessons from the design and contextual conditions that influenced the pilot's outcomes and relate what has happened in Uganda to the learning form the other AgResults pilots.

3. Evaluation approach by question

3.1 Evaluation Question 1: What has been the impact of the pilot on private sector involvement in the development and uptake of improved legume seed?

In this section we present the methodological and conceptual frameworks that will inform our approach to Evaluation Question 1 on the pilot's role in improving the performance of Uganda's legume seed market. This is followed by a preliminary application of the framework to the market based on our IQA and a discussion of data sources for the analysis.

3.1.1 Methodological approach

Our evaluation of the pilot's impact on private sector involvement in the development and uptake of improved legume seed will analyse whether the pilot intervention strengthened the market for improved legume seed. We will also analyse how the development of the improved legume seed market differed by region and farmer characteristics and by the presence of complementary legume and seed interventions. We will document the structure of the market and the strategies of firms in the market. We will also evaluate whether the pilot's effects can be differentiated on the basis of major community characteristics such as socio-economic profile or proximity to key private sector actors engaged in the market. We will examine how private sector involvement in the pilot and improved legume seed market evolve over the life of the pilot. Finally, given the critical role of the AgVerify initiative in the pilot, we will conduct a nuanced analysis of the relative roles of AgVerify (which serves as a "push" in the AgResults parlance) and the AgResults "pull" incentives in motivating seed companies' behaviour and the effects of this behaviour on the development of the market for improved legume seed.

We will apply the SCP framework to develop and test qualitative hypotheses regarding the effects of the pilot incentive on private firms' perceptions of, participation, and outcomes in the market for high-quality legume seed.[7] SCP is a conceptual framework that links the underlying characteristics of a market (its "situation") to firm strategies about whether and how to engage in the market, to the overarching market structure, and to the market's performance. Below we detail the components of the SCP and their relationships. Then, we apply the SCP framework to the improved legume seed market in Uganda.

[7] The SCP paradigm is a product of the Industrial Organisation school of economics (Caves, 1987; Scherer & Ross, 1990). The use of SCP as an evaluation tool was pioneered by John Holtzman of Abt Associates (Holtzman, 2003). The seminal SCP framework delineates how the underlying conditions in a market influence the market's structure, which in turn influences individual firms' conduct in the market (such as decisions to invest in new market segments and technological and organisational decisions). Individual firms' decisions, at an aggregate level, lead to market performance outcomes of interest such as the adequacy of a product's supply in terms of volume and quality, prices, returns to investors, and responsiveness to consumer demand. Building on the basic SCP framework, Sutton (1992) introduced the practice of examining how endogenous and exogenous sunk cost investments influence industry structure. This approach, which we will apply in the current analysis, recognizes that firm strategic conduct is a direct response to market conditions and that aggregation of the outcomes of firm strategic behaviour gives rise to market structure. The paradigm thus follows the logical progression from situation (basic or underlying market conditions) to strategy to structure to performance.

Table 3-1. Sub-questions and outcome measures for Evaluation Question 1

Evaluation Question 1: What has been the impact of the pilot on private sector involvement in the development and uptake of improved legume seed?	
Sub-question	**Outcome measures**
Basic Conditions What are firms' perceptions of market conditions, and how do those perceptions influence their decisions around engaging in the market for improved legume seed? How do different value chain actors—particularly seed companies, institutional buyers, agro-input dealers, and farmers—perceive improved legume seed in terms of its relevance to them and their business models? • How does the AgVerify initiative affect market actors' perception of the improved legume seed market? What other factors—such as end-market demand, government/institutional activity, legislation, production conditions, and other socio-economic trends—indirectly influence demand for and supply of improved legume seed? • How does the institutional environment affect development of the improved legume seed market?	• Market situation for improved legume seed and market actors' perceptions of these conditions - Supply and demand conditions for improved legume seed, both AgVerify certified and non-certified seed - Transaction costs and risk in acting in the improved legume seed market , including with and without AgVerify certification - Perceptions of price and volume parameters underlying entry and sustained participation in market for improved legume seed, including with and without AgVerify certification - Existence of enabling legislation for development of improved legume seed market - Market actors' awareness of and perceptions regarding relevance of the AgVerify initiative - Activity of institutional buyers (including government) in the improved legume seed market - End-market demand and other exogenous factors as relevant
Strategy • What drives the decision of different market actors to buy or supply improved legume seed? • What are procurement, value-addition, merchandising, and distribution strategies for improved legume seed (as relevant to different market actors)?	• Market strategy for improved legume seed by market actors - Drivers for decisions to buy or supply improved legume seed including role of AgVerify - Decision to participate in the AgVerify initiative - Procurement, value-addition, merchandising, and distribution strategies for improved legume seed, both AgVerify certified and non-certified
Structure • How is the improved legume seed value chain structured in terms of product movement through the market? • How many private sector actors of different types participate in the market? • What volumes are transacted by different types of market actors? • What technologies, organisational arrangements, and logistical arrangements predominate? • What is the pattern of women's participation in the value chain for legumes?	• Improved legume seed market structure - Flow of AgVerify-certified improved legume seed through the value chain - Number and types of private actors who participate in the market - Participation of seed companies (both AgResults participants and others) in AgVerify initiative - Volume and share of volume transacted by different value chain actors - Technologies and logistical arrangements predominant in the market - Difference in how women participate in

	the improved legume seed value chain, particularly at the production level
Performance • Does a sustainable, private-sector-driven market for improved legume seed exist? • Is improved legume seed accessible (in terms of availability and price) to diverse producers? • Do smallholder farmers perceive benefits from use of improved legume seed? • What are smallholder farmers' experiences with different distribution modalities (e.g., ag-input suppliers versus organisational or institutional suppliers)? • Does the market disadvantage or otherwise affect specific stakeholders, such as women or other vulnerable groups?	• Market performance - Existence of market volume and quality to support sustained improved legume seed trade - Returns and perception of benefits of engagement in the market - Ability of vulnerable groups such as smallholder farmers and women to participate in the market and benefit from use of improved legume seed

The Structure–Conduct–Performance conceptual framework

The SCP conceptual framework analyses the underlying conditions of the market (its "situation") to the strategic decisions that market players, including firms, farmers, and consumers, make about whether and how to engage in the market, given their perspectives on the underlying market conditions. The strategic decisions of numerous firms give rise to the market structure, which includes the numbers and characteristics of market participants and the predominant marketing channels and modes of product transformation and value addition. Together, these factors affect the performance of the market, including such considerations as whether the supply of improved legume seed improves, and whether and which farmers have enhanced access to improved legume seed. While we refer to the overall paradigm as SCP, the specific analytical model that we will use in this evaluation reflects a causal flow from situation to strategy to structure to performance. We will apply the SCP analytical framework at baseline and endline to evaluate the pilot's impact on private sector involvement in the market, and the effect it has on the development of the improved legume seed market.

Situation: The underlying, or "basic", conditions of a market (also referred to as its "situation") are fixed in the short to medium term and include characteristics of supply and demand of a product and its market and the institutional environment. Supply and demand conditions include cost structures, seasonality of demand and supply, income distribution, and buyers' and suppliers' responses to changes in prices and income. Other salient characteristics of a market include the prevalence of information costs and asymmetry and asset specificity, which increase transaction costs and risk. The institutional environment includes both formal (legal) and informal (cultural) controls on behaviour, and is critical to establishing behavioural norms that reduce transaction costs and the risks to which potential buyers and suppliers in the market are exposed. Together, these conditions define the incentives and create interdependencies that shape individuals' and firms' decisions regarding whether and how to engage in the market (North 1990).

Strategy: Individuals' and firms' strategic behaviours reflect their attempts to pursue profit and utility objectives given the constraints imposed by external conditions. Strategic behaviour includes such decisions as whether to invest in production facilities or a new venture; pricing and service delivery choices; whether to register a company rather than continue as an informal entrepreneur; and the choice of institutional arrangements between market actors, such as the type of contract structure.

Structure: A market's structure is shaped by the aggregate decisions of many individual firms. Structural elements include the numbers of buyers and sellers in the market, the characteristics of production and value creation (such as the technological packages that

dominate), the degree and types of product differentiation, and barriers to entry and exit. Such structural features tend to evolve over the medium to long term and as such are represented among the basic conditions that influence firms' strategic behaviour.

Performance: The performance of a market can be understood in innumerable ways, but the main elements of interest for the legume seed pilot include whether a sustainable market for improved legume seed emerges and whether that market is accessible and appropriate (in terms of physical access, product form, pricing, and quality assurance) to the needs of its buyers.

Preliminary application of the SCP framework to the legume seed pilot

Preliminary research, including an extensive literature review, field visits during the IQA in April 2015, and discussions with the AgResults Secretariat, informed a preliminary application of the SCP framework to the legume seed pilot. The results of this preliminary application will serve as hypotheses to be tested through our baseline assessment; and are summarised here to guide discussion of our evaluation approach.

Situation: The basic conditions underlying Uganda's improved legume seed market include weak institutions (in particular the lack of effective quality certification) that increase transaction costs and risks of engaging in the market, and unpredictable and price-sensitive demand among buyers. Many farmers have limited awareness of the benefits of improved seed, but demand for improved seed is more robust among farmers who are cognizant of its potential benefits. Supply constraints include inadequate availability of breeder and foundation seed, lack of working capital, difficulty obtaining an adequate supply of legume seed, and limited potential profitability of legume seed production. Another consideration is the opportunity cost of applying working capital to legume rather than hybrid maize seed production, which is more profitable. Together, these market conditions contribute to a fairly risky market environment with limited profitability potential.

Strategy: In terms of their strategies regarding engagement in the improved legume seed market, national seed companies focus heavily on the profitable hybrid maize market but perceive the legume seed market to be a means of diversifying their portfolio. The current government procurement, which requires a large volume of seed without the imposition of stringent quality requirements, has also piqued seed companies' interest and investment in the legume seed market. Nonetheless, given the unpredictability of demand, seed companies base their production decisions on conservative estimates.

Structure: The structure of the improved legume seed market reflects the importance of public breeders as a source of breeder and foundation seed. National-level seed companies rely heavily on independent outgrower farmers or smallholder farmers working collectively in LSBs for multiplication of improved legume seed, complemented in some cases by production on their own land. When faced with inadequate improved legume seed to meet demand, many national-level seed companies fill their orders with "standard seed" (i.e., legume seed purchased from the grain market that is treated and sold as certified seed). Industry experts and seed company experts have estimated that 50 to 90% of the "certified" legume seed that is marketed in Uganda is actually standard seed (Mainville 2015).

Performance: Uganda's improved legume seed market suffers performance failures at virtually every level. The outcome of these issues is not only that inadequate volumes of high-quality legume seed are produced, but that Uganda's legume seed market is a classic "lemons market", where inability to effectively signal or enforce quality in the context of intensive price pressures leads to low-quality legume seed dominating the market. Few farmers have access to the benefits of improved legume seed, with consequent negative impacts on their yields, production, food availability, and incomes.

Seen through the SCP framework (and as described under the theory of change), the legume seed pilot incentives are intended to temporarily alter the underlying conditions of the market that national seed companies face. Specifically, annual prizes based on year-

over -year sales growth increase returns to investing in the market, while the at-cost provision of at cost refrigerated storage for unsold seed between seasons reduces the risk of loss if a seed company expands its production too aggressively in a given year. We hypothesize that, in the context of the AgVerify certification scheme, which is designed to resolve the “lemons market” disincentive to produce and sell quality improved legume seed, national seed companies will respond to the pilot incentives as changed market conditions. As a result, they will increase their investment in the improved legume seed market. This includes scaling up their production or procurement of improved legume seed and developing demand for their product (e.g. through advertising, demonstration plots, or any other promotional strategies they may conceive) and distribution channels to supply that demand. If this pilot is sustained and of adequate scale, it should catalyse engagement by other parties (such as institutional buyers and farmers) who will gradually become aware of the availability and potential benefits of improved legume seed, leading them to develop sustained demand for it. Eventually, these responses will serve on aggregate to alter the structure and performance of the market to the point where it becomes self-sustaining and responsive to the needs of its users.

We will apply the SCP framework to the improved legume seed market, guiding our assessment with the preliminary SCP analysis and hypotheses detailed above. We will adjust hypotheses for our endline analysis on the basis of baseline results and any unanticipated changes to pilot implementation. We will also account for important market developments that might affect pilot impact but are not themselves presumed to be due to or influenced by the pilot (such as the entry of new players, performance of the AgVerify certification scheme, or policy changes). We will compare outcomes between the baseline and endline periods, as well as track ongoing results (drawing on the monitoring data collected by the Pilot Manager) throughout the pilot.

3.1.2 Data sources

We will conduct interviews with representatives of seed companies, agro-input dealers, institutional buyers of seed, farmers, and other sector experts (e.g., in government or the NGO or donor communities) to refine and formalize our baseline assessment underlying Evaluation Question 1. We will speak to representatives of national seed companies, including all prospective pilot implementers as well as a selection of national companies that are or have been active in the improved legume seed market but are not prospective implementers. We will also interview representatives of development initiatives (such as those outlined in Section 1.4.3) and major institutional buyers of seed. These interviews will provide overarching insights into the market and will also enable triangulation of data and findings from the market actor interviews.

Within each of Uganda’s administrative regions, we will interview regional seed producers (such as LSBs) or distributors (such as seed retailers or wholesalers based in the major cities) in each of the administrative regions. We will also interview bean and soybean farmers and agro-input dealers that sell seed in two communities of each of the administrative regions. This geographic representation will ensure that we capture diversity in the importance of bean and soybean in farmers’ livelihoods and in the local economy. We will also vary our sampling of communities to obtain diversity in terms of the importance of bean or soybean in local agro-ecosystems (for example major producing areas versus marginal ones), in orientation of production for market versus home consumption, and in the presence of development initiatives targeting bean or soybean. Within each community we will speak to farmers of varying scales and commercial orientations, being sure to capture diversity in terms of gender, location, and participation in membership organisations and development initiatives. We will also collect data from agro-input dealers in these communities. We will identify candidates for data collection in consultation with local extension officers and through referrals solicited within the community.

To track the development of the market throughout the life of the pilot, we will monitor implementation through regular communication with the Pilot Manager and the Secretariat. We will review data on seed company production and sale of improved legume seed from the pilot's monitoring system as well as any data that can be made available from AgVerify. Then, at endline, we will conduct another large-scale qualitative data collection exercise in which we will interview the same sets of stakeholders discussed in the preceding two paragraphs.

The data we will draw upon to answer this evaluation question, while it will include quantitative data on seed production and sales, will be primarily qualitative. We will employ "best practices" in qualitative research to ensure the robustness of our qualitative methods. These include "naïve" questioning approaches (rather than "leading" questions which introduce bias), triangulation of data sources (for example seeking information from multiple levels of the marketing chain to obtain diverse explanations of phenomena), and the careful documentation of the evidence supporting results (Yin, 1994). Much like quantitative research, the validity of qualitative research is also bolstered by leading with theory-based models (such as the SCP framework), as well as actively seeking out disconfirming rather than confirming evidence. These best practices will allow for nuanced exploration of diverse factors, such as those identified above, which might also affect the pilot's outcomes of interest. We will collect baseline data beginning in February 2017 and endline data in February 2022.

3.1.3 Sampling plan

Anticipated sample sizes for each type of respondent are outlined in Table 3-2. It is important to note that our results for Evaluation Question 1 will also be informed by the results of the smallholder data collection described for Evaluation Questions 2 and 3.

Table 3-2. Sampling plan for evaluation question 1

Respondent group					Sample size, all communities
National-level sample					
Seed companies					6
Institutional buyers					4
Legume or seed-related development initiatives					4
Regional sample					
	Eastern	Western	Northern	Central	Subsample size, all communities
Regional seed companies producers or distributors	2	2	2	2	8
Communities	2	2	2	2	8
Agro-input dealers	2-3	2-3	2-3	2-3	16-24
Farmers	3	3	3	3	24

3.1.4 Data management and analysis

The evaluation team's Qualitative Lead will train the Uganda-based Agricultural Economist on the SCP model and appropriate data collection methods prior to implementation of baseline data collection activities. The Agricultural Economist will then collect data under the supervision of the Qualitative Lead. Data collection will use face-to-face interviews at the site of the respondent's legume and seed-related operations when feasible.

We will record the large majority of the data using sound recordings (with respondent permission and following best practices to ensure integrity of the data) or, where necessary, verbatim notes. The Agricultural Economist will enter data into a template provided by the Qualitative Lead, and transmit the populated template to her for analysis. The Qualitative Lead and Qualitative Methods Specialist will clean, code, and analyse data using Microsoft Excel, SPSS, and qualitative data analysis software such as NVivo as appropriate.

We will analyse data on market structure using descriptive statistical methods. For example, for sales and production data gathered from the pilot's monitoring system, we will present descriptive statistics and comment on the implications of any observed trends over time.

We will analyse data from key informant interviews using pattern analysis, in which we will evaluate preliminary hypotheses on the basis of field results to ascertain patterns and divergences among similar market actors. The analytic process and interactions with the Agricultural Economist who collected the data will facilitate an active search for disconfirming evidence and alternative explanations for observed outcomes, and we will further investigate results that do not align with the hypotheses.

The Qualitative Expert will be responsible for data analysis and reporting of results; however, the nature of qualitative research implies that there will be substantial communication with the Agricultural Economist based in Uganda to clarify questions, elicit further insights, ask follow-up questions (as necessary), and vet and review research results.

3.2 Evaluation Questions 2 and 3: What has been the impact of the pilot on smallholders' uptake of improved legume seed? What has been the impact of the pilot on smallholders' incomes?

In this section we present the methodological and conceptual frameworks that will inform our approach to answering Evaluation Questions 2 and 3 on the pilot's role in improving uptake of improved legume seed, and its consequent impact on smallholder income. This is followed by a discussion of data sources and the analysis plan.

3.2.1 Methodological approach

We will use a pre-post design performance evaluation to compare uptake of improved legume seeds, the quality of improved legume seeds, and smallholders' income (as well as intermediate outcomes such as productivity) after the five-year pilot.[8] We use the terminology "performance evaluation" as defined by DFID in the Terms of Reference of its Global Evaluation Framework Agreement: 'Performance Evaluation assesses an intervention on the basis of contribution to development outcomes and impacts within its context.' This is in contrast to the way it describes impact evaluation: "Impact Evaluation builds knowledge on what works in development by establishing causal attribution in relation to an intervention and its effects." These definitions are in agreement with those used by USAID and other donors.[9] Accordingly, our performance evaluation design will detect changes in our key outcomes of interest using a pre-post comparison, allowing us to investigate contribution of the AgResults pilot to observed outcomes but not to identify a causal effect of the AgResults pilot on those outcomes.

[8] If resources allow, we will also conduct limited quantitative and qualitative data collection and analysis around the middle of the pilot. This would be beneficial in terms of tracking progress and in capturing fluctuations in improved seed adoption that a strictly pre-post study might not capture. However, the decision to collect data at midline is mainly a budgetary one that remains under discussion as this design report is finalized.

[9] See, for example, USAID's official evaluation policy at https://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf.

We will collect household-level panel data both at baseline and at the end of the pilot, and describe and compare these data. Unfortunately, this approach lacks a rigorously-defined counterfactual group: we cannot know what smallholder outcomes would have occurred over the same period in the absence of the AgResults pilot. While we will not unequivocally determine causation, we will perform extensive qualitative research and secondary data review to understand how and why observed changes would plausibly have occurred (i.e., has the pilot played out as planned in the theory of change, and if not, what went differently than planned?). We will also search for and rule out alternative explanations. This process of assessing alternative explanations, called contribution analysis, uses a systematic application of logic to rule out external factors that could account for observed changes. This approach will allow us to reach an evidence-based conclusion about whether the pilot contributed to the observed changes.

The Uganda pilot is a nationwide private-sector effort operating in the context of a complex value chain where seed companies transact with different types of entities who distribute seeds in difficult-to-track ways, making it difficult to trace the distribution of seed geographically. Thus, the pilot design and the nature of the legume seed value chain does not allow for setting aside a control or comparison group of farmers once the pilot rolls out, even for a limited period of time. In addition, creative alternatives for designing an impact analysis are quite limited owing to an insufficient number of seed companies willing to compete for sales-based prizes in some areas and not others to provide a contemporaneous with-versus-without comparison.

In these ways, the Uganda pilot somewhat resembles the Kenya on-farm storage pilot, where most major maize-growing areas are covered by the pilot and there was no feasible way to set aside comparison areas. In Kenya, Abt is implementing a quasi-experimental interrupted time series (ITS) design (it is quasi-experimental because, although there is no comparison group of farmers, the design incorporates maize yield as an "unaffected outcome" as a standard of comparison – maize yield is affected by all of the same exogenous factors as the main outcomes of interest but is not expected to be affected by the pilot). The ITS design measures the intervention impact as a departure from the expected levels of the outcome measure (in that case smallholders' uptake of on-farm solutions, smallholder income or food security) were the treatment not introduced (e.g., Shadish, Cook, and Campbell 2002; Bloom 2003). Specifically, an ITS design entails (1) generating a counterfactual for the outcome measure, which represents the expected level of the outcome in the post-intervention period in the absence of the treatment and is constructed by projecting the trend in pre-intervention observations of the outcome measure, and (2) modelling the treatment effect as a deviation from this counterfactual. Multiple pre-intervention observations, over several prior plantings/harvests, are obtained by asking respondents retrospective "recall" questions in the baseline survey. When the question concerns adoption of on-farm storage technologies we expect strong recall accuracy, since the event in question represents an unambiguous event around a major investment in improved farming practice.

The parallel situation does not hold in the Uganda pilot. It is much more complicated for both smallholders and survey interviewers identify when adoption of improved legumes took place than for on-farm storage, especially using retrospective survey questions. Unlike purchasing an on-farm storage container (a binary variable) in the case of the Kenya pilot, there are many ways a Ugandan farmer might have acquired improved legume seed in the past, including:

- Purchasing it directly from a retailer,
- Receiving seed from various NGOs,
- Receiving seed from the Ugandan government,

- Re-using saved improved seed (the progeny of improved seed may be re-planted for up to three generations before it degrades to the point where it is no longer considered improved), or
- Acquiring second- or third-generation improved legume seed from someone else (such as a neighbour who had a good harvest using improved seed in the previous season).

Given the diversity of types of seed and means of procuring it that could be counted as having adopted improved seed, it would be necessary to ask a large number of questions about seed procurement at various points in time over past planting cycles to discern whether and when a respondent adopted improved seed. So while in Kenya the respondent can be asked in a fairly straightforward, easy-to-understand way when they first purchased on-farm storage, it is a much more complex undertaking to find out when a Ugandan farmer first used improved seed. Furthermore, it is impossible to know if the adopted seed is indeed improved despite the farmer's perception that it is (we discuss our strategy to address this issue, a seed quality assessment, later in this section). Given the difficulty that respondents often have with complex recall questions, we avoid recall bias by gathering data on improved legume seed use for the most recent year's planting seasons, rather than asking respondents complicated set of questions stretching back three seasons or more.

We also considered whether secondary data could serve as the basis upon which to build an ITS design. There is but one candidate in this regard, the World Bank's Living Standards Measurement-Integrated Surveys on Agriculture (LSMS-ISA). It includes questions on improved legume seed adoption and legume sales and exists for five time periods prior to the pilot, the latest of which is 2013. However, since 2013 was a full four years prior to the start of the pilot, that data set would not accurately represent the situation at baseline, so the ability to project into the AgResults pilot period would be severely compromised. Furthermore, that survey's current funding source has ended, and it is not clear whether additional funding will allow it to continue into the future, necessitating primary survey data collection for this evaluation even if we were to use prior year data from LSMS-ISA. We also considered using the LSMS-ISA's sample design taken forward in time using AgResults data collection resources to enable some ITS analysis. However, we cannot use the LSMS-ISA's sample frame exactly. This is primarily because the pilot's impacts will be concentrated in the heavier legume-producing areas of Uganda, and the LSMS-ISA survey does not focus on legume-producing areas. Therefore, while we will use the LSMS-ISA to descriptively assess any pre-pilot trends in improved legume seed adoption and legume-related revenue, we will not use it for ITS regression analysis of the trends.

Thus, for the sequence of reasons laid out above, we have ruled out an ITS analysis of adoption impacts because we do not see a feasible way to collect data on multiple pre-pilot time points in order to project a counterfactual trend into the post-pilot period.

As a result, the only feasible design is a pre-post evaluation comparing smallholders' income and uptake of improved legume seeds for the planting season immediately before the pilot (from the baseline survey) to the same indicators after the pilot concludes five years later (from an endline survey).

The risk in this approach is that the main outcomes of interest could change due to non-pilot factors, but that in the pre-post analysis may erroneously seem to attribute these changes to the pilot. To guard against this risk (but not eliminate it) we will conduct quantitative and qualitative data collection and analysis, as well as a contribution analysis. Specifically, to augment the quantitative pre-post comparison of outcome data, our contribution analysis entails searching for alternative explanations for the outcome. We will iteratively gather and examine evidence on all possible alternative causes of observed changes, reviewing our conclusions with stakeholders and local experts. For example, the quantitative pre-post comparison may reveal an increase in smallholder up take of improved legume seed. In our

report, we will present the size of the increase for the full sample of households we survey, as well as selected subgroups of households. Alongside this quantitative analysis, we will include a narrative from our qualitative analysis of possible alternative explanations for this increase (other than the pilot itself). In this narrative, we will discuss non-pilot factors such as a growing export market for soybean or by weather, pest, or crop disease events that affect local varieties. Our analysis will seek out information on all external factors like these and use logic to determine what combination of factors likely explains the changes we have observed in farm-level outcomes. The goal is to be able to say that a preponderance of evidence *suggests* the pilot contributed to the change in outcome, though we will not be able to say with certainty that the pilot caused results because the evaluation design does not involve a valid counterfactual and therefore cannot establish causal attribution.

The main outcomes of interest in our pre-post evaluation are uptake of improved legume seed, improved productivity and subsequent impacts on smallholder income through increased legume sales. We will define "income" for this pilot as profits from sales, i.e. revenue less the cost of production, since that is the primary avenue through which the pilot is expected to increase farmer income. As elements of income, we will also report on how much legume the farmer produces, how much legume the farmer sells, the price of legumes the farmer obtained when selling his product, and the farmer's costs in producing that crop. There are a number of intermediate outcomes we will also investigate because they comprise elements of the causal pathway toward the main outcomes of interest. Farmer knowledge of and positive attitude toward improved legume seed is a precondition for uptake, as is access to the seed, so we will investigate these intermediate outcomes. For example, we will ask how the farmer perceives quality, where s/he can buy improved seed, and whether s/he thinks the improved seed on the market yields a higher quality harvest. In addition, income increases are contingent on a superior product that produces a higher yield. Therefore we will include yield as an intermediate outcome (yield is calculated as production divided by land area planted, as self-reported by respondents). Finally, because many farmers grow legumes primarily for own consumption, we will also investigate household legume consumption, determining households' quantity consumed and whether respondents perceive the quantity available for consumption to be optimal or not.

As a supplement to the quantitative research, we will also undertake a national seed quality assessment to study the extent to which seed that is labeled as "improved" actually is improved (i.e. that it meets AgResults seed quality standards).This assessment is needed to address the issue of accurately estimating uptake of truly improved seed in a situation in which what is *purported* to be improved seed and what is actual improved seed differ drastically. Since a high proportion of seed that is labeled and sold as "improved" actually is not improved, we cannot measure uptake of improved seed based on respondents' understanding and self-reporting alone – this would overestimate uptake of improved seed, possibly drastically. Instead, we will use a seed quality assessment to adjust the self-reported rate of improved seed adoption by the proportion of "improved" seed in Uganda overall that is found to be truly improved through the seed quality assessment. We will also collect household survey data from the same households that supply seed for the quality assessment. This will allow us to examine the characteristics of households that are able to accurately, as opposed to inaccurately, judge the quality of the legume seeds they have acquired. For example, farmers with larger farms who are less credit-constrained, better educated, or belong to farmer groups may be more likely to have seed that is truly improved because they buy new seed every year, whereas farmers with smaller farms, lower access to credit, education, and farmer group measurement who are more credit-constrained may use recycled seed purchased from a neighbour or local market that they falsely perceive as improved. This analysis will help us to assess the reach of AgResults across various farmer types.

Further details for all of these elements of our plan are provided below.

Pre-post quantitative approach

For the pre-post evaluation we will interview a panel 1,500 households (plus the 300 households from whom we will collect seed samples at baseline and endline, the sampling plan for which is described below) in early 2017 before pilot implementation, and in early 2022 after 5 years of implementation of the pilot. This will provide quantitative data to be used in the performance evaluation described above and given more detail below. (Details on our sample size calculations and sample selection strategy for this data collection appear in sub-Section 3.2.3 below).

The full set of primary outcome measures we will study quantitatively includes[10]:

- Household planted any improved legume seed (self-reported) in the most recent season (binary variable)
- Household planted any AgVerify-certified legume seed in the most recent season (binary variable)
- Quantity of improved legume seed planted (self-reported) in the most recent season, in isolation and compared to total legume seed planted (continuous variable)
- Quantity of AgVerify-certified legume seed planted in the most recent season, in isolation and compared to total legume seed planted (continuous variable)
- Hectares the household planted with improved legume seed (self-reported) in the most recent season, in isolation and compared to total hectares planted with legume seed (continuous variable)
- Hectares the household planted with AgVerify-certified legume seed in the most recent season, in isolation and compared to total hectares planted with legume seed (continuous variable)
- Household income from legume sales (continuous variable)

For all outcomes that involve seed directly, such as the first two above, we will collect data on both AgVerify-certified seed and on improved seed of any kind, as long as it is considered "improved" by the respondent. (AgVerify-certified seed, identifiable by its AgVerify label, will serve as a proxy for legume seed sold under the AgResults pilot, since AgResults will not be directly promoted to farmers and will not be a brand name familiar to them.) Ideally, our set of outcomes would only include seed that was sold as part of the AgResults pilot. However, given the range of channels through which a farmer might access improved seed, respondents may or may not know what company their seed originated from and whether it was certified under AgVerify. This would lead to a downward bias in our estimate of uptake of AgVerify-certified seeds. By also asking questions about *any* improved seed, regardless of whether it was certified by AgVerify, we will be better able to triangulate our data on seed uptake during the AgResults pilot.

The first six of the above measures account for different definitions of "uptake" and will serve to answer evaluation question 2. The first two measures tell us whether the household used improved legume seed (self-reported or AgVerify-certified, respectively) during the most recent season. The third, fourth, fifth, and sixth measures tell us how much of that seed they planted and on how much of their land, both in isolation and as a proportion of total legume seed planted, since some households plant both improved and non-improved varieties

[10] For all outcomes that involve measures about the seed being promoted by AgResults, we will collect data on two separate measures: 1) knowledge/adoption of *any* improved seed and 2) knowledge/adoption of AgVerify seed.

during the same season. The final primary outcome is smallholder farm income from legume sales, the measurement of which we will use to answer evaluation question 3.

Recognising that improved seed adoption and other farming practices may vary within households, we will consider all of the indicators listed above using a gendered lens. To facilitate that, we will gather these data at the plot level and disaggregate results by the gender of the person who makes decisions about the plot.

To generate hypotheses supporting the likelihood that AgResults influenced observed changes over time in uptake and legume-related income, we will also study the following intermediate and secondary outcomes using the pre-post differencing approach:

- Farmer knowledge/perception of various characteristics of improved legume seed (binary variables for each characteristic)Farmer perception of availability of/access to improved and/or AgVerify legume seed (Likert scale)
- Legume yield (production divided by area planted; continuous variable)
- Quantity of legumes produced (continuous variable)
- Quantity of legumes sold (continuous variable)
- Quantity set aside for home consumption (continuous variable): legume consumption: number of meals per week involving legumes (continuous variable for each month of the past year) and perception among households that their legume consumption is sufficient (binary variable for each month of the past year to account for seasonal variation in legume consumption)

The models we will use in our analysis are described in Section 3.2.4 below.

Pre-post national seed quality assessment

Since seed quality is a central issue in this pilot, a national seed quality assessment is a critical component of the evaluation. Even if the evaluation were able to determine with certainty the proportion of farmers who have adopted seed that they think is improved, the fact remains that much of the seed labelled as “improved” actually is not. This means that self-reported measures of adoption, even from a nationally representative sample of legume farmers, would likely overstate the true adoption of improved legume seed. Therefore, we will conduct seed quality assessments at baseline and endline to enable us to compare the true penetration of improved legume seed before and after the pilot.

The outcomes of interest for the seed quality assessment, which will be used to determine whether a particular sample of seed meets the AgResults standards of being “improved”, are:

- Moisture rate (% wet);
- Purity (the percentage composition by weight of the sample being tested and the proportion of pure seeds, other seeds, and inert matter constituting the sample); and
- Germination rate (the proportion of the sample comprised of normal seedlings, abnormal seedlings, fresh seeds, hard seeds, and dead seeds).

Seed must meet COMESA standards on all of the above criteria to be considered “improved” under AgResults. (We do not propose genetic testing as there does not exist a genetic test of whether legume seeds are improved or not.)

To implement the seed quality assessment, we will work with Chemiphar, the seed lab which is responsible for seed quality testing for AgVerify. Chemiphar is the only International Seed Testing Association (ISTA) certified lab in Uganda. Chemiphar will be responsible for both sampling and testing.

Chemiphar staff members will travel to smallholder households during the August-September 2016 planting season to sample seed directly from households that self-report that they are using improved seed (see 3.2.3 for more information on how this seed will be sampled). Chemiphar will then transport the sampled seed to their lab in Kampala, where they will perform testing. Chemiphar will submit a final report to Abt detailing the proportion of seeds meeting the minimum quality standards to be considered improved among the seed nationally perceived to be "improved." All seed sampling and testing will be done in accordance with ISTA standards.

Once Abt has collected self-reported survey data from legume-growing households about use of what they believe to be improved seed, we will use the Chemiphar information to scale the results. For example, if we find that 10% of the seed reported planted by respondents in our survey is perceived as "improved" by respondents, and Chemiphar finds that only 50% of the "improved" seed in Uganda actually is improved, we will adjust our results to reflect that in fact only 5% of the total seed planted by respondents was improved seed. In so doing, we will be able to estimate the actual adoption, rather than the perceived adoption, of improved legume seed.

Additionally, we collect baseline and endline household survey data from the same households who provided seed samples. Using these data together with the results from the seed quality assessment, we will look for relationships between adoption of truly improved seed and characteristics like farmer group membership, gender, distance to roads and markets, and education of the household head.

Pre-post qualitative approach

We will use qualitative research to understand the "how" and "why" of improved legume seed uptake, production and consumption decisions among smallholders. We can evaluate the decision to purchase improved seed and to consume and/or sell legumes grown from that seed, from a behaviour change perspective in which farmers' decisions to grow, sell, or consume a crop are outcomes of their knowledge and attitudes about it. Knowledge of and positive attitude toward improved legume seed is a necessary pre-condition for growing it. Thus, we will evaluate the impact of the AgResults pilot on farmers' decision to plant improved legume seed and subsequently sell or consume it by conducting a more in-depth KAP assessment of farmers based on open-ended questions. The "knowledge" component (also assessed quantitatively) will evaluate whether consumers are aware of improved legume seed as an alternative to local varieties and, if they are, the extent of their knowledge about improved seeds and how they differ from local varieties. The "attitude" component will assess farmers' perceptions of the relative benefits of improved legume seed (such as higher yield) as compared to local varieties (such as lower cost), their consumption and taste preferences; their perceptions of the legume market, improved legume seed, seed companies, and AgVerify; and their ability to earn money through legume sales. The "practice" component will focus on farmers' production, sale, and consumption of improved legume seed and the factors driving and constraining those decisions.

Our qualitative inquiry will gather data through open-ended questions about smallholders' self-reported KAP and decision making surrounding improved legume seed uptake and income. We will use these data to develop our understanding of the "how" and "why" of changes detected in the quantitative household survey as outlined in 3-3. Adoption of improved legume seed is far from binary; some households may quickly switch to planting all or mostly improved seed, while many may try out a small amount during a season, while still others may split their production between local legumes they plant for consumption and improved seed they plant for sale. To enable the comparative analysis, we will aim to sample non-adopters, low- or medium- adopters, and high-adopters to understand the dynamics behind each group's adoption decisions. While the qualitative endline outcome levels will not be directly attributable to the pilot, we will nonetheless be able to draw useful insights about how household socio-economic characteristics, the legume market, smallholders' knowledge

and perception of improved seed, and household decision making, among other factors, may have influenced uptake and household income.

Table 3-3. Evaluation method and outcome measures for evaluation questions 2 and 3

Evaluation Question 2: What has been the impact of the pilot on smallholders' uptake of improved legume seed? **Evaluation Question 3: What has been the impact of the pilot on smallholders' incomes?**	
Evaluation method	**Outcome measures**
• Pre-post performance evaluation	**Primary outcomes of interest:** • Household planted any improved legume seed in the most recent season (self-reported) (binary variable) • Household planted any AgVerify-certified legume seed in the most recent season (binary variable) • Hectares planted with improved legume seed, self-reported and adjusted based on results of seed quality assessment (ha), both by itself and compared to total hectares planted with legume seed • Hectares planted with AgVerify-certified legume seed, both by itself and compared to total hectares planted with legume seed • Quantity of improved legume seed planted, self-reported and adjusted based on results of seed quality assessment (ha), both by itself and compared to total legume seed planted • Quantity of AgVerify-certified legume seed planted, both by itself and compared to total legume seed planted • Farmer legume-related income (UGX) **Intermediate and secondary outcomes:** • Proportion of "improved" seed that meets AgResults moisture, purity, and germination standards (%) • Farmer knowledge of improved legume seed (binary variables for knowledge of various improved legume seed characteristics) • Farmer knowledge, attitudes, and practices regarding improved legume seed (qualitative assessment) • Farmer perception of availability of/access to improved legume seed (Likert scale) • Quantity of legumes produced from improved legume seed (kg) • Legume yield (kg/ha) for both improved and non-improved seed • Quantity of legumes sold (kg) • Legume consumption among smallholder households: number of meals per week involving legumes (continuous variable) and perception among households that their legume consumption is sufficient (binary variable for each month of the year to account for seasonal variation in legume consumption) • Proportion of legumes consumed vs sold (%) • Farmer use of e-verification service for improved seed (if e-verification data is available)

Contribution analysis

As discussed above, as this evaluation is a performance evaluation, rather than an impact evaluation, contribution analysis will be our primary approach to ensuring rigour and to assessing the likely contribution of the pilot to observed changes.[11] Contribution analysis is an approach to reaching conclusions about whether a programme or intervention is likely to have contributed to observed results based on a preponderance of evidence that events did or did not play out as intended according to the programme's theory of change. It is a structured way of synthesising information to reach evaluation conclusions, and we will use it to structure our consideration of data that we collect throughout the evaluation (household survey data, interviews with a wide variety of stakeholders, secondary data that may include macroeconomic indicators, agricultural production data, pilot monitoring data, LSMS-ISA; any data that becomes available from AgVerify, and documentation of relevant policy changes or related donor programmes) to determine whether a preponderance of evidence suggests that the pilot has contributed to observed outcomes. Our use of contribution analysis recognises that, because identifying a causal effect is impossible given the context of this pilot, our best option is to identify its likely contributions.

Contribution analysis will help us to construct a plausible narrative about how the pilot contributed to changes in key outcomes by:

1. *First, working from a clearly defined theory of change.* We have set out the theory of change in Section 1 of this document, showing the theorised pathways to impact and describing some of the major assumptions that underlie it. Our contribution analysis will flow directly from the theory of change and the assumptions we have identified.

2. *Identifying the links within the theory of change for which there is less evidence (i.e., the assumptions).* For example, one major assumption that we identified in Section 1 was the assumption that there exists significant latent demand for improved legume seed in Uganda. This may be true, but we do not know for sure ex-ante. Alternatively, it may be the case that there is such a high taste preference for local varieties of beans that demand for improved varieties is actually quite limited. Or perhaps farmers who decide to invest in improved seed also prefer to invest in a higher-value cash crop than beans, or perhaps the majority of farmers simply cannot afford the higher cost of improved seed at all. We simply do not yet know what the true demand picture looks like.

3. *Investigating further the links and assumptions for which we have less evidence.* In the previous step, we identified the links and assumptions in the theory of change for which evidence is lacking. In this step, we search for evidence that supports or opposes those assumptions. To continue with the example of whether the latent demand for improved legume seed is as large as assumed in the pilot design, we will gather as much data as possible throughout the evaluation (as discussed above) on both sales within Uganda and on farmer perceptions and behaviours. We will triangulate all data by incorporating as many data sources as possible, including at a minimum pilot monitoring data on seed company sales and primary data on farmer knowledge, attitudes, and practices. We will similarly collect and triangulate data on all links and assumptions underlying the theory of change.

4. *Identifying other possible causes of outcomes and determine whether they also likely contributed or whether they can be eliminated as contributing factors.* For example, if improved soybean seed sales to smallholders increase dramatically during the course of the pilot, this may be due to the pilot or it may be due to other factors such as increases in demand for soybean in the export market. We would therefore assess whether demand for soybean has increased, looking at macroeconomic data on trade and prices as well as

[11] For more background on the use of contribution analysis as an evaluation tool, see http://betterevaluation.org/en/plan/approach/contribution_analysis.

interviewing aggregators and exporters. If we did find that export demand had increased dramatically, we would have to build that into our narrative as having been a likely contributor to the increase in soybean seed sales to smallholders. We will use both primary and secondary data to understand all non-pilot factors that could have impacted outcomes in addition to the pilot. We will use all of this information to construct a plausible narrative about whether the theory of change played out as expected, and why or why not.

5. *Assessing the narrative we have developed by running it by key informants*, including the AgResults Secretariat and Pilot Manager, international and local seed sector experts, and beneficiaries including seed companies and farmers. We will review with these key informants the patterns of results that we have identified to make sure they agree a) that our narrative captures all of the most important factors that could have been at play and that b) the preponderance of evidence supports our conclusions.

Having followed these steps, if the pilot is implemented according to the theory of change and we observe the expected changes (i.e., increased adoption of improved legume seed and increased smallholder income), we would be able to say that a reasonable claim of contribution can be made. We would be able to make that claim because we would have determined that other potential influencing factors either were not the sole causes of the changes or have not made a significant contribution. Accordingly, our final report would present results describing how the theory of change is supported by the available evidence and has not been disproved.

3.2.2 Gender considerations in the evaluation

Given the considerations outlined in Section 1.4.4, we hypothesize that women and women-headed households are likely to see lower uptake of improved legume seed, and consequently potentially lower output and income effects, due to their lower participation in crop and input markets in general. We hypothesize that participation in organisations and development projects may mediate this tendency towards lower participation. We will test these hypotheses using both quantitative and qualitative methods, collecting data on gendered household structure and on intra-household roles and responsibilities in production, marketing, and decision-making regarding legumes. We will also analyse how uptake and income effects differ for women (and men) participating in agricultural organisations or development initiatives. We will use our qualitative inquiries to enrich our understanding of these issues and to identify any other unanticipated, gender-differentiated outcomes at the farmer level.

3.2.3 Sampling plan

The pre-post design described above requires baseline and endline measures of the key outcome variables for the evaluation. We will collect baseline and endline data through a household survey and qualitative interviews administered to smallholders and by sampling and testing seed for the seed quality assessment. Timing of each wave of data collection is discussed in Section 4. In the event that resources allow, we will also collect limited data at midline to confirm trends in monitoring data and information from stakeholders.

In addition to the primary data collection we discuss below, we will also request data on e-verification messages sent to AgVerify (if AgVerify and/or the participating seed companies do roll out an e-verification scheme, as is currently being discussed). We will request information such as the quantity of e-verification messages sent about improved legume seed and, to the extent these data are available, location of the message sender. This will help us to understand the volume and spatial distribution of e-verification, which would be an indicator that AgVerify-certified seed is being traded at a particular time or in a particular area. This will allow us to develop a rough mapping of the spread of improved legume seed over the life of the pilot, providing context to our results at the smallholder level.

Sampling for quantitative data

Because the pilot is nation-wide, we have designed the household survey sample to be representative of all legume farming households in the nation, drawing from districts in all regions of Uganda. Ideally, we would survey a random sample of smallholders within many villages, selecting villages with probability of selection proportional to the size of the average legume harvest in those villages. However, there does not exist a sample frame for households and we do not have village-level data on legume harvests. Instead, we will use data from Uganda's National Agriculture Census on legume production by district as a proxy for the quantity of legume seed planted.[12] Therefore, each village within a district will have the same sampling probability because our legume production data is at the district level. The village-level sampling probability will be directly proportional to a district's percent share of total national legume production (the number of selected villages in a district may be zero). We cannot know ex-ante which of the villages may be targeted by AgResults-related efforts. By drawing a random sample of villages across all districts in the country, we hope to obtain a representative mix of AgResults implementation efforts.

After selecting the villages, we will select households for interviews. At this point, we do not anticipate being able to obtain listings of households from which we could randomly select a sample. One approach, though not ideal as it may be biased against the more vulnerable, is to use linear systematic sampling, wherein we choose households using a random walk pattern through the village. The random walk method involves randomly selecting a starting point and walking through a village asking every household at a pre-determined interval to participate. Another option is to obtain lists of households from village leaders upon arrival, though this method can also lead to bias or recall error on the part of the village leader supplying the list, particularly if it is supplied from memory as is likely to be the case. We will work with the survey firm and local leaders to explore the feasible means of achieving an unbiased, random sample during survey pilot testing.

We conducted a power analysis to determine how many households to interview, and arrived at a total of 10 farming households each from 150 villages, to which we will add 300 households from whom we collected seed samples for the seed quality assessment. We assume that households in a village will have correlated responses, and look to recent sub-Saharan agricultural estimates of intra-cluster correlation coefficients to arrive at a conservative estimate of an ICC of 0.1 for all outcomes (Geyer, Davis, and Narayan 2016). We assume that the pre- and post-pilot outcomes are correlated, with R-squared of 0.2. We also assume a sample attrition rate of 20 percent, i.e. that we will not be successful reaching 20 percent of the baseline survey respondents at the time of the follow-up survey. With these assumptions, we will be able to detect a difference of 0.14 in standardized differences between pre-and post- measures. For example, if the standard deviation of legume-related income is 8,000 Uganda shillings (UGX), there is an 80 percent chance that we will detect a difference between pre and post farmer income of 1,120 UGX with a 10 percent chance of a Type 1 error. For another example, if the current adoption rate of improved seed is 40 percent, then there is an 80 percent chance that we will detect a difference between pre and post adoption rates of 7.0 percentage points (with a 10 percent chance of a Type 1 error). Table 3-4 displays the minimum effect size we could obtain with a variety of possible sample allocations.

[12] National data on legume seed planted is not available, but data on legume production serves as a good proxy.

Table 3-4. Minimum effect size detectible with different sample allocations

Number of villages sampled	Number of households sampled in each village	Minimum effect size
50	10	0.249
80	10	0.196
100	10	0.174
150	10	0.142
200	10	0.123

Sampling for seed quality assessment

The objective of the seed quality assessment is to compare the actual quality (i.e. whether it meets AgResults standards) of "improved" seed planted by farmers before and after AgResults. Working with Chemiphar, Abt will determine the number of seeds to sample and the process for collecting seed samples for both the baseline and endline smallholder surveys. The sample size required depends on the percent of seeds labelled "improved" that is actually improved, both at endline and at baseline. The goal of the power analysis is to select a sample size for which a 95% confidence interval for this percentage is smaller than the expected difference between the baseline and endline percentages of truly improve seed. We present several scenarios assessing this factor in Table 3-5.

Table 3-5. Seed quality assessment power analysis scenarios

Number of seeds sampled:	Percent of seed that is considered "improved" by smallholders that is actually improved		Difference between baseline and endline	Confidence Interval
	At baseline	At endline		
50	60%	75%	15%	+/-15%
100	60%	75%	15%	+/-10%
200	60%	75%	15%	+/-7%
300	60%	75%	15%	+/-6%
50	60%	90%	30%	+/-13%
100	60%	90%	30%	+/-9%
200	60%	90%	30%	+/-6%
300	60%	90%	30%	+/-5%

Given the uncertainty surrounding these estimates—no comprehensive study has ever been made of the quality of legume seed in Uganda—it is prudent to consider a baseline rate of 40 to 60 percent (both 40 and 60 percent will yield the same confidence intervals). As we cannot know how effective the pilot will be at boosting sales of improved seed, we conservatively consider that it will increase the percent of seed that is labelled "improved" that is actually improved by 15 percent. In this conservative scenario, 100 seed samples are required to detect a statistically significant difference of a 15 percentage point improvement. However, owing to the minor cost implications but great improvement in precision, we will use 300 seed samples.

In addition to statistical precision, the other design goal for the seed quality study is representativeness. The seed selection plan includes the number of districts, villages, and farmers to sample and the number of seeds to be collected from each farmer. We would ideally select a sample of legume and soybean seeds that is representative of all legume and soybean seeds planted by smallholder farmers in Uganda. However, this is infeasible in practice. Instead, we will use the same process used to select villages and households for

the quantitative study, based on data from Uganda's National Agriculture Census, which provides legume production totals at the district level. We will randomly select the number of villages to visit in each district. The probability that a village is selected will be proportional to the percent of the national bean and soybean production from that village's district.[13] With our current plan to gather 300 samples, we will select 60 villages and gather seed in each village from 5 households who self-report that they are growing improved legume seed, using the random walk method to find households. Since we cannot know the legume and soybean production in each village or by each farmer before data collection, we will select the villages and households from whom to sample seeds randomly. For this survey, we do not consider attrition, because it is not necessary to sample the same households at the follow-up data collection. The goal of the seed sample study is cross-sectional.

Sampling for qualitative data

We will use non-probabilistic, multi-stage sampling to select research participants. We aim to select a minimum of 48 smallholder households from the following sub-populations of interest:

- Region (Northern, Western, Eastern, Central)
- Level of improved legume seed adoption (none, low, and high)
- Gender of household head and of respondent (i.e. we will sample women from male-headed households in addition to male and female household heads)
- Soybean production (since it is not yet common)
- Homestead distance from tarmac road (and thus likely market access).

We will perform qualitative interviews slightly later than our quantitative interviews and will use responses given to the quantitative survey to purposively select households for interviews in accordance with the sampling plan outlined in Table 3-6 below. The purposive sampling plan is intended to ensure that respondents represent a range of important sub-populations along lines of gender, region, and level of improved legume seed adoption.

Approximately half of respondents will be adult males and approximately half will be adult females, who are either the household head or spouse of the household head, to capture perspectives of men and women. As feasible, in each region, one or two of the women will be household heads, and one or two of the men or women will be part of remotely located households that fulfil other sampling criteria. Due to resource and logistical constraints and uncertain pilot participation, we may not be able to follow the same smallholder households from the baseline through the endline. Therefore, we will make comparisons across household characteristics and case studies rather than longitudinally for a fixed set of households.

[13] The households selected for the seed quality assessment cannot be selected as a subset of households from the quantitative survey because the quantitative survey will occur after the seed quality assessment. The quantitative survey must occur after the season's harvest because it involves variables like production, yield and prices received by farmers for estimation of income. But the seed quality assessment sampling must be done at the beginning of the growing season before farmers have finished planting. This timing prevents us from selecting households for the seed quality assessment from among the respondents to the quantitative survey. Furthermore, the low rate of improved legume seed adoption would likely mean that fewer than the required 300 households from the quantitative survey plant improved legume seed, so we might not have enough households to sample seed from. Because of this timing, we will return to collect household survey data from the households who provided seed samples, in addition to the 1,500 households we will select randomly. This will result in 1,800 total households represented in the survey.

As outlined in Table 3-6, we will replicate the sampling plan across each region of Uganda. Within each region, we will sample from the list of districts and villages we cover in our quantitative survey. We will consider aggregate soybean production in selecting districts, aiming for one district (from among those represented in our quantitative survey) with high soybean production to ensure capturing some soybean producers in our sample. We will aim for least one of the villages in each region being two or more hours (by car) from a tarmac road and the other(s) being closer to the nearest tarmac road, as a rough proxy for that village's access to markets and farming information.

Table 3-6. Household sampling plan by region, adoption level and gender

Adoption Level	Gender (respondent/ HH head)	Northern	Western	Eastern	Central	Total
No adoption	Male	3	3	3	3	*24*
	Female (male-headed HH)	2	2	2	2	
	Female (female-headed HH)	1	1	1	1	
Low to medium adoption (<50% of legume seed improved)	Male	3	3	3	3	*24*
	Female (male-headed HH)	2	2	2	2	
	Female (female-headed HH)	1	1	1	1	
High adoption (>50% of legume seed improved)	Male	3	3	3	3	*24*
	Female (male-headed HH)	2	2	2	2	
	Female (female-headed HH)	1	1	1	1	
Total		*21*	*21*	*21*	*21*	*72*

For qualitative studies, there is little available guidance on how to operationalize the concept of "data saturation," or the point at which no new information or themes are observed in the data. Two studies that do address this question in several sub-Saharan African countries determined that new concepts plateaued after 6 qualitative interviews and data saturation occurred after 12 interviews (Guest, 2006; Francis et al., 2010). We are using this as a guide to determine our minimum sample size. Based on this work, the smallholder sample size will likely be adequate to capture differences between sub-populations and achieve overall data saturation for common experiences and cultural concepts. The results of our data collection will not be generalisable to or representative of the population as a whole, but will rather be applicable to the farmers from whom we collected data.

3.2.4 Data management and analysis

For quantitative data, Abt will contract with a local data collection firm to collect data electronically using tablets or smartphones and a to-be-determined survey software package (likely SurveyCTO, SurveyBe, or SurveyToGo). The firm will script not only the survey questions and response options but also data quality control mechanisms such as range checks and skip patterns to reduce data entry error. Additionally, the field supervisors will perform back checks for a minimum of ten percent of households interviewed as well as sit-ins to observe at least ten percent of interviews. Data will be reviewed daily during the survey period for accuracy, consistency, and adherence to the sampling plan by field supervisors and, members of the Abt team. Data will then be uploaded to secure servers

and cleaned in several stages by the data collection firm and the Abt team using SPSS or Stata.

Members of the Abt team will perform quantitative data analysis using Stata. We will perform both descriptive and regression analysis. Using regression analysis, we will model the dependent variable (Y: outcomes as listed above in Table 3-3 including knowledge of improved/AgVerify seed, adoption of improved/AgVerify legume seed, quantity of seed/hectares planted with improved/AgVerify seed yield, legume income, legume consumption) as a function of baseline household characteristics (X: such as household size, gender of head of household, poverty index) and an indicator for the post-implementation survey date (P) as written in Equation 1.

$$\boldsymbol{Y_{i,t} = \alpha + \beta X_{i,t=1} + \delta P_t + \varphi_i + \mu_{i,t}} \quad (1)$$

Equation 1 includes two unobservable household characteristics. One of the unobserved characteristics is constant over time (φ_i: for example, skill/ability or willingness to try new technologies). The other unobservable characteristic varies over time ($\mu_{i,t}$), for example health status. The pre-post analysis is *more efficient* if we remove the former (we cannot remove the latter) from the data. We remove time invariant unobservable household characteristics by subtracting the pre-pilot measure of the outcome variable, $Y_{i,t=1}$, from the post-pilot measure, $Y_{i,t=2}$for each household, as shown in Equation 2. This subtraction also removes all the fixed-throughout-time covariate terms$X$. Therefore, the statistical analysis is paired down to a simple test of whether the mean of a normally-distributed random variable is equal to zero: a t-test of the term $Y_{i,t=1} - Y_{i,t=2}$across households.[14]

$$\boldsymbol{Y_{i,2} - Y_{i,1} = (\alpha - \alpha) + (\beta X_{i,t=1} - \beta X_{i,t=1}) + \delta(1 - 0) + (\varphi_i - \varphi_i) + (\mu_{i,2} - \mu_{i,1})} \quad (2)$$

$$\boldsymbol{Y_{i,2} - Y_{i,1} = \delta + (\mu_{i,2} - \mu_{i,1})}$$

Although household characteristics are not required for the full-sample pre-post analysis, at baseline we will collect data about household characteristics in order to study the pre-post difference for various subgroups defined by gender, poverty level, proportion of legume crop that is soybean, farm size, and baseline legume income. In equation 3, the term (δ_x) will not be equal to zero if the pre-post difference in outcomes varies by baseline characteristic (X).

We will also ask about exogenous factors (Z) that might play a role in legume farming, such as weather, as these factors could help explain the difference between baseline and endline outcomes (dependent variable $\boldsymbol{Y_{i,2} - Y_{i,1}}$). Equation (3) illustrates how these exogenous factors could enter into the analysis: they help explain changes in the outcome variable over time to the extent that they change over time as well, accounting for their appearance in the equation as$\boldsymbol{Z_{i,2} - Z_{i,1}}$. We will use the linear regression model with baseline covariates and exogenous factors (Equation 3) for both binary and continuous outcomes.

$$\boldsymbol{Y_{i,2} - Y_{i,1} = \delta + \delta_x X_{i,1} + \gamma(Z_{i,2} - Z_{i,1}) + (\mu_{i,2} - \mu_{i,1})} \quad (3)$$

For our seed quality study data, we will compare the proportion of seed that meets AgResults quality standards before and after the pilot. We will also incorporate the results of

[14] One alternative to Equation (2) is to "residualize" the outcome data by first regressing the dependent variable at both time periods on baseline characteristics ($\boldsymbol{Y_{i,t} = \alpha + \beta X_{i,t=1} + \delta P_t + \epsilon_{i,t}}$), and then estimating whether the coefficient on $\widehat{\epsilon_{i,1}}$ in the following regression is equal to one: $\widehat{\epsilon_{i,2}} = \beta\widehat{\epsilon_{i,1}} + \omega$. However, we do not anticipate that this method would be more efficient.

the seed quality study into our household-level quantitative analysis. The household survey will take place several months after the seed sampling survey, because the seed sampling needs to occur near planting time and the household survey needs to occur after harvest and marketing. Still, we will survey the households that took part in the seed sampling survey in order to (a) explore household characteristics that are associated with use of truly improved seed, and (b) contrast harvest outcomes of households using truly improved seed versus seed falsely perceived to be improved. These analyses will be exploratory, offering suggestive narratives about adoption of improved seed and the consequences of adoption before and after the pilot. We will use simple linear regressions, and possibly logit models, to explore household characteristics associated with use of truly improved seed. We will use only simple linear regressions to explore contrasting harvest outcomes of households using truly improved seed versus seed falsely perceived to be improved, because harvest outcomes are continuous numbers.

For qualitative data, members of the Abt team will perform all data collection and analysis directly. To ensure quality, we will record interviews (with respondent consent) or, if necessary, take verbatim notes. Then, we will perform content analysis on the interview transcripts and/or notes. Broadly, we will analyse in-depth qualitative data to understand the potential degree and diversity of pilot effects on smallholder improved legume seed uptake. We will draw comparisons across different subgroups differentiated by gender, farm scale, farm distance from major transport networks, and regional variations to account for the influence of differing socio-economic and contextual factors.

We will use NVivo to organise, code, and analyse the qualitative data.[15] Analysis of qualitative data begins with coding, i.e., flagging pieces of data that relate to a theme or concept of interest (thematic codes) or to a specific research question, sub-question, or objective (structural codes). The codes will be informed by a priori concepts that the pilot theory of change, SCP framework, desk research, and the IQAs suggest will factor into the success of the pilot. We will apply this deductively developed codebook to enable content analysis, a form of text analysis that enables hypothesis testing (Bernard, 2006).

During initial, exploratory data analysis, we will broadly analyse thematic coded data for common patterns, cultural categories, themes, and outliers related to the topics of interest. After initially exploring the data, we will perform content analysis in NVivo to test hypotheses developed based on the pilot theory of change to determine if farmers' improved legume seed uptake and subsequent outcomes fit the expected pattern and which factors impact improved legume seed uptake and outcomes (e.g. gender, farm scale, membership in farmer groups, etc.).

3.3 Evaluation Question 4: What has been the impact of the pilot on poor consumers' demand for legumes and derivative products?

The legume seed pilot is intended to spur the market for legume seed only; the pilot does not aim to directly impact consumer demand for legumes or derivative products. In other words, it aims to shift the supply curve, not the demand curve. Therefore, we do not expect to see changes in demand for legumes due to the pilot in the time period of our analysis. Still, we will monitor any exogenous developments in consumer demand to understand the shifting context in which the pilot operates. For example, demand for soybeans is increasing in both the domestic and export markets for industrial processing into oil, flour, animal feeds, and other derivative products independently of the pilot. As part of our response to the other evaluation questions, we will examine the interplay between (exogenous) rising demand for

[15] Due to resource constraints, we will primarily code notes taken during data collection, audio recordings of interviews, photographs, and short videos rather than coding transcripts.

legumes, especially soybean, and any attendant increase in legume production in response to that demand because this external factor could affect pilot outcomes.

We will also, as part of our evaluation for questions 2 and 3, assess whether there are any significant price changes in the legume market that could be due to the pilot. If the pilot were to cause legume production to rise significantly, we would expect legume prices to decrease. While this is not the same thing as an impact on poor consumers' demand, it would be part of the overall welfare effects that the pilot has in Uganda and is therefore important to consider.

3.4 Evaluation Question 5: What evidence exists that the effects of the pilot will be sustainable in the medium to long term?

In the AgResults context, sustainability means that the pilot has initiated significant and long-lasting contributions to the development goals that motivate it *after the pilot has concluded*. Assuming a positive initial impact, the sustainability of the legume seed pilot will depend on whether market developments that the pilot has stimulated will continue following cessation of the direct pilot incentives; that is, whether the preconditions for a sustainable market have been established or not.

Qualitative contributions to the evaluation of sustainability will come from the SCP and household-level analyses, and will focus on whether the basic conditions that provide incentives for continued private sector and farmer engagement in the market are present by 2022. These include:

- Whether actors engaged in the improved legume seed market (including farmers) feel that their engagement is adequately rewarding for them to want to sustain it following the conclusion of the pilot;
- Whether AgVerify continues to certify legume seed in the absence of AgResults funding and whether seed companies continue to pay for that certification, signalling a commitment to sustaining higher seed quality;
- How farmers perceive the trade-off between the yield benefits of improved seed and other characteristics, such as taste and cost;
- Whether seed distribution mechanisms have sufficiently evolved that there is reliable, affordable access to improved legume seed for those who wish to plant it (for example, through NGO or governmental distribution schemes or through local agricultural input suppliers so that farmers do not have to travel long distances to access seed); and
- Whether there are other exogenous factors that will affect the sustainability of the pilot's effects, such as evolving demand or changes to the enabling environment.

We will also evaluate market actors' actual behaviour and their perspectives on the viability of the market and their intentions for continued engagement in the market following cessation of the pilot's activities. Specifically, we will:

- Request seed company production and sales data after the conclusion of the pilot;
- Ask seed companies and other market actors about their interest and intentions around continuing higher production and marketing of improved legume seed;
- Inquire into the details of any plans these sources report to gain a sense of their nature and the market actors' commitment to them; and
- Monitor seed companies' continued involvement in the AgVerify scheme, including their behaviour after the subsidy for AgVerify is phased out.

- Inquire of market actors what conditions they see as necessary to successfully carry out their plans and their assessments of the likelihood that these conditions will be fulfilled.

Table 3-7. Evaluation method and outcome measures for evaluation question 5 summarises the key evaluation method and the key outcome measures to answer Evaluation Question 5. We will collect data from private sector actors and farmers, and other seed industry experts in the development community as well as among other public and private stakeholders. The in-country Agricultural Economist, who is responsible for conducting the questionnaires, will compile the results. The Qualitative Lead will analyse and report the data in conjunction with the in-country Agricultural Economist.

Quantitative information will also be used to address this evaluation question. For one, we will make use of secondary data on pilot-related outcomes (particularly on smallholder uptake of improved legume seed and on their legume-related income) to the extent possible. One potential source of data is the LSMS-ISA, if data continue to be collected at regular intervals. Examining this data after the conclusion of the pilot to determine the prevalence of improved legume seed use by smallholder farmers will be a cost-effective way to gauge the sustainability of the pilot, assuming the data collection continues at regular intervals and is made available to us. In early 2024, two years after the pilot concludes, we will also conduct a "sustainability survey" of brief interviews with smallholder households focused on current perceptions of the improved legume seed market (including the availability of improved legume seed), smallholders' most recent behaviour with regard to improved legume seed production, and their current legume consumption and sales. Information from this survey will be provided in a Sustainability Report following data collection.

Table 3-7. Evaluation method and outcome measures for evaluation question 5

Evaluation Question 5: What evidence exists that the effects of the AgResults pilots will be sustainable in the medium to long term?	
Evaluation method	**Outcome measures**
• SCP analysis • Analysis of farmer uptake and income effects (evaluation questions 2 and 3)	• Extent to which farmers feel they have reliable, affordable access to improved legume seed • Farmers' perceptions of the attractiveness and likelihood of continued use of improved legume seed following the pilot • Suppliers' perceptions of the attractiveness and likelihood of continued engagement in the market following cessation of the pilot • Continued participation in AgVerify following the pilot • Suppliers' legume seed production and sales after the pilot • Existence of other exogenous factors, such as market and enabling environment evolution, that affect attractiveness of legume seed market participation.
• Sustainability survey (quantitative assessment of sustainability of key outcomes)	• Same as primary outcome measures for evaluation questions 2 and 3

3.5 Evaluation Question 6: What is the evidence on the scale of any effect on private sector investment and uptake, and the cost-effectiveness of the pilot as an approach?

This question involves two separate but related elements: scale and cost-effectiveness. We will determine the scale of pilot impact on private sector investment and uptake of improved legume seed by drawing on the results of the SCP analysis, used in Evaluation Question 1.

Market structure estimates from the SCP will provide information on the numbers and characteristics of private sector investors and participants in the value chain, as well as volumes of improved legume seed transacted as a result of the pilot.

We will also examine the pilot's cost effectiveness, an important question in its own right to assess the pull mechanism as a use of donor funding but also as an important element of the ability of the approach to be scaled up. Central to the motivation behind the use of incentive-based pull mechanisms is the expectation that they will be more cost-effective than traditional development interventions, and hence more scalable. The private sector, it is argued, can be closely attuned and responsive to the needs of agricultural markets if the sector's incentives can be aligned to support the development of those markets. At the same time, however, incentive-based mechanisms have not yet been applied to any significant extent in agricultural development programming, so evidence about their cost effectiveness is as yet unavailable.

We will compute the cost effectiveness of AgResults at endline once the total project costs are known. Cost effectiveness is measured as a ratio, cost per unit of impact. Its determination will require estimates of both cost and pilot impact. We will use as our numerator the gross cost of the pilot. As our denominator the results of our analysis for Evaluation Questions 2 and 3; namely, measured change on improved legume seed adoption by smallholders and any attendant increase in smallholders' legume-related income.[16] The gross costs of the pilot will be based on actual project expenditures from the start of the project through its conclusion using project monitoring data. These expenditures will cover incentive payments, verification procedures, payments to AgVerify, and other types of expenses incurred in the course of pilot implementation. This accounting will also include pertinent AgResults project administration and management costs, which we will distribute over all of the pilots and also discount to account for the time value of money.

We will also compare the cost effectiveness ratio of a given pilot to that of other AgResults pilots. This will not be a cost-benefit analysis—that is, we will not assign a monetary value to changes in the affected agricultural markets and will not compare the pilots' overall value to their costs. Comparisons of AgResults cost effectiveness results to the findings for other interventions will include discounting adjustments so that costs are expressed in comparable dollars when measured in different years. In addition, the cost effectiveness analysis will include sensitivity tests for alternative discount rates. Table 3-8. Evaluation methods and outcome measures for evaluation question 6 summarises the evaluation methods and key outcome measures to answer Evaluation Question 6.

[16] Note: because this is not an impact evaluation, we are able to estimate only the change over time during the pilot period, not the impact attributable to the pilot. Our cost effectiveness results will thus be indicative only. Nevertheless, it is valuable to compute cost effectiveness ratios for the key outcomes of interest to determine the scale of costs versus change in smallholder outcomes.

Table 3-8. Evaluation methods and outcome measures for evaluation question 6

Evaluation Question 6: What is the evidence on the scale of any effect on private sector investment and uptake, and the cost-effectiveness of AgResults as an approach?	
Evaluation methods	**Outcome measures**
Scale of private sector investment and uptake	
• SCP market analysis	• Market linkages, interest, and investment in continued involvement in market for improved legume seed by value chain actors (including those not directly engaged with AgResults)
Cost-effectiveness	
• Cost-effectiveness analysis comparing cost against outcomes of the pilot • Synthesis of qualitative results from all AgResults pilots	• Cost per unit of change (defined as changes in smallholder adoption of improved legume seed and legume income)

3.6 Evaluation Question 7: What lessons can be learnt about best practices in the design and implementation of AgResults?

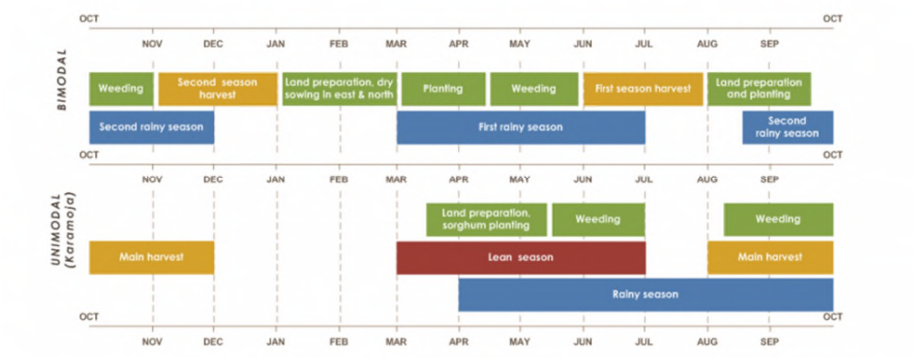
An assessment of design effectiveness and identification of best practices for pull mechanism design are central to the evaluation and learning framework of the AgResults initiative. The Abt team is currently in the process of developing an overarching design paper for lessons learnt across all AgResults pilots. The framework and methodology for addressing lessons learnt for the Uganda pilot will be forthcoming as part of that paper.

4. Implementation timeline and other considerations

4.1 Implementation timeline

This section presents the timeline for all major activities in the Uganda pilot evaluation plan. Our approach to implementing the evaluation involves an IQA followed by baseline and endline household quantitative surveys and qualitative data collection. We will collect data with Uganda's planting and harvesting schedules in mind (see Exhibit 4-1). For example, because we wish to ask farmers about their legume harvest, we will conduct farmer surveys after the harvest period. Baseline data collection will reference the season immediately preceding the start of the pilot, which is currently scheduled for early 2017. Because our key outcomes of interest include both adoption (which occurs before or during planting) and income (which would be affected by sales after harvest), we will collect each round of household data slightly after harvest for the second rainy season (i.e. beginning in February 2017), according to the Uganda seasonal calendar shown below in Exhibit 4-1.

Figure 4-1. Uganda agricultural season calendar



Source: FEWS Net

4.1.1 Initial qualitative assessment and evaluation design missions

The evaluation team undertook the first phase of our IQA in April 2015. The purpose of this trip was general scoping and gathering information on the pilot context. Then in March 2016, to coincide with the AgResults Steering Committee meeting in Kampala, members of the team returned to Uganda for the second phase of the IQA (in which we gathered inputs for our design for the evaluation for questions 2 and 3) in March 2016.

4.1.2 Baseline data collection

The AgResults pilot was initially scheduled to begin in August 2016 to coincide with planting for the second rainy season. However, a number of delays have occurred and the pilot is now scheduled to begin *after* the second rainy season. Given the delay, the first rainy season 2017 will be the first season in which pilot-related sales could occur. Therefore, the second season of 2016 will serve as the baseline reference season. Accordingly, we have collected seed samples for the seed quality assessment during the planting period of this baseline reference season (August-September 2016). Seed testing will occur from September to November 2016. Data collection to inform evaluation question 1, such as key informant interviews with solvers and other value chain actors, will be timed to follow the pilot's official launch event (scheduled for February 2017). We will conduct household-level qualitative and quantitative data collection through interviews and surveys, respectively, starting in February 2017. It is not possible for seed sampling and household data collection to happen concurrently; our household data collection will involve questions about what

happened during and shortly after the season (production, harvest, consumption, sales, etc.) and therefore we can only conduct this data collection *after* the reference season has ended.

4.1.3 Ongoing stakeholder engagement and secondary data review

Following baseline data collection, the evaluation team will continually monitor the pilot implementation as part of our ongoing qualitative assessment. This will consist of regular communications with the Pilot Manager, the Secretariat, DFID, and the Steering Committee to keep track of any issues that arise, their importance to the pilot's implementation, and how they are eventually resolved. This will continue up to the point of endline data collection.

We will also request access to data collected by AgVerify on e-verification activity, if applicable. This information will help us understand the uptake of AgVerify-certified products in general and of improved legume seed in particular, providing important context to the evaluation.

4.1.4 Midline data collection

Because the five-year Uganda pilot is longer than most AgResults pilots, it would be ideal to collect and analyse some data in the middle of the pilot in 2019 to gauge progress. However, the question of whether or not resources will allow for midline data collection was still under discussion at the time this report was finalized. If Abt and DFID later determine that midline data collection can be funded given available resources, we will conduct quantitative and/or qualitative data collection and analysis around the middle of the pilot. This would be beneficial in terms of tracking progress and in capturing fluctuations in improved seed adoption that a strictly pre-post study might not capture. If it is agreed that we will do a formal midline, a memo describing the data collection and plan for analysis will be attached to this design report. In the event that resources do not allow for midline data collection, we will use our notes from regular Secretariat, Pilot Manager, and Steering Committee meetings, secondary data like LSMS-ISA, and pilot monitoring data at endline to retrospectively determine as much as possible about the timing and magnitude of changes during the pilot without collecting primary data at midline.

4.1.5 Endline data collection

In 2022 (at the conclusion of the five-year pilot) we will collect endline data. We will re-survey the same panel of smallholder households surveyed at baseline to collect endline quantitative data. For qualitative data we will aim for the same balance across our primary sampling criteria (region, level of improved legume seed adoption and gender) that we used in the baseline. We will also conduct an endline seed quality assessment, using the same methodology used at baseline. We will submit our analysis of endline data, including findings on results realised in the intervening years of implementation, to DFID in the quarter following endline data collection in early 2022.

4.1.6 Sustainability survey

Assuming the pilot proceeds according to the current schedule, we will present the sustainability assessment by the end of 2024. We will base this assessment both on the farmer-level sustainability survey (conducted in late 2023, two years after the pilot concludes), on our interviews with various market actors, and on secondary data that includes information on improved legume seed use and production, such as LSMS-ISA survey data. To the extent that seed company data are made available to us, we will also analyse legume seed production and sales trends after the pilot.

4.2 Deliverables and communication plan

We will post all evaluation updates and reports on the AgResults website following DFID approval, and where relevant on the Abt Associates website and social media (e.g., Facebook, Twitter). We will also provide an update on the evaluation at each semi-annual Steering Committee meeting. We will submit the baseline and endline reports to DFID for formal review, after which they will be posted on the DFID external website. The evaluation

team will also present the salient lessons learnt to the Steering Committee as a part of a dissemination workshop to take place at a place and time to be determined in consultation with DFID and the Steering Committee. We will summarise the content of this dissemination workshop in a one-page technical project summary, which will also be available on the Abt website, and we will assist the Secretariat's efforts to further disseminate the evaluation results. We will submit evaluation reports to DFID on the approximate dates shown in Table 4-1.

Table 4-1. Projected submission dates of AgResults evaluation reports

Deliverable	Projected submission date
Baseline report	June 2017
Endline report	June 2022
Dissemination workshop report	TBD

We will further disseminate project findings through presentations, academic papers, or other means when opportunities arise, as deemed appropriate by both the evaluation team and DFID.

4.3 Evaluation risks and mitigation approach

The risk of collecting biased or poor-quality data exists in every evaluation. To avoid these issues, we will employ a number of safeguard measures and data collection best practices. For quantitative data, we will use an electronically-scripted survey instrument with built-in skip patterns and range checks to reduce data entry error. We will supervise observation and back-check at least ten percent of interviews. Both Abt and our survey sub-contractor will review data at least weekly during data collection to ensure accuracy, consistency, and adherence to the sampling plan. For qualitative data, we will use "naïve" questioning approaches (rather than "leading" questions, which introduce bias), triangulation of data sources (for example, seeking information from multiple levels of the marketing chain to obtain diverse explanations of phenomena), and the careful documentation of the evidence supporting results (Yin 2003). Much like quantitative research, the validity of qualitative research is also bolstered by leading with theory-based models (such as the SCP framework we are using), as well as actively seeking out disconfirming evidence in addition to confirming evidence.

Other risks are more specific to this evaluation design. One significant risk is that our results could be incorrectly interpreted as causal when in fact they are not. To mitigate this risk, we will use clear language in our final report and any dissemination materials that this evaluation wasn't designed to determine attribution or causality. We will emphasise to readers that, while it may seem plausible that the pilot contributed to changes, we as evaluators cannot claim to know that with confidence. Even after performing contribution analysis and attempting to eliminate external explanations for the observed changes, we will not claim that our results tell us the *impact* of the pilot. Instead, the strongest statement we can make, if changes do occur and if we have qualitatively ruled out alternative explanations, is that it is *likely* that the pilot *contributed* to changes. During our internal review, we will give reviewers specific instructions to check that the final report and all other materials make it clear to the reader that this is the nature of our findings.

Another risk is that the pre-post data collection approach will miss important changes that occur closer to the middle of pilot. The Uganda pilot is longer than most AgResults pilots, at five years. However, we may not have the resources to perform a complete midline survey. To mitigate the risk of missing important variations in outcomes that occur within the pilot, we will take the following steps:

- Frequent communication with in-country stakeholders including the Pilot Manager and AgVerify staff to stay abreast of developments;

- Regular review of secondary data (including monitoring data, AgVerify data, and any other available secondary data) during the pilot;
- In-depth qualitative investigation, as described in Section 3.2, at the smallholder level at baseline and endline to understand farmer perceptions of changes over the course of the pilot ;
- Limited quantitative and qualitative midline data collection in 2019, if resources allow, to confirm patterns observed from stakeholder consultation and secondary data.

Additionally, though more of a limitation than a risk, the evaluation will not be able to separate the effects of the AgResults pull mechanism from the effects of AgVerify, since all AgResults solvers are required to participate in AgVerify as well. Because these two initiatives are operating hand-in-hand, it will not be possible to quantitatively determine the incremental effects of each one. That said, our extensive qualitative research at all levels of the legume seed value chain will include questioning of actors at all levels as to the relative influence that each element of the pilot exerted on their decision-making.

Finally, we must remain aware of the fact that, because Chemiphar is involved in the implementation of the pilot, its role in the evaluation could be perceived as a conflict of interest. We do not see a financial interest to Chemiphar in altering its behaviour or the results of its evaluation seed tests and therefore judge that there is no true risk to the evaluation or the pilot. In any case, as the only ISTA certified lab in the country, Chemiphar is the only choice of subcontractor to perform this work.

4.4 Ethical considerations

To ensure that we collect data in an ethical and responsible way, the team will submit the data collection instrument and draft design report to Abt's Internal Review Board (IRB) for review prior to conducting any data collection. Further, as we begin work with the contracted survey firm, we will work with the IRB to execute the project's data security plan—a continuously updated document that will track how data will be handled and by whom, and what security measures are undertaken to maintain respondent confidentiality.

4.5 Quality assurance

Quality assurance is an integral part of our evaluation. Abt employs both internal and external quality assurance to review the data collection instruments, the study design, and all results. Dr Kyle Emerick from Tufts University and Dr Andrew Kizito from Makerere University provide external quality review, covering both the evaluation's methodology and its results. Dr Stephen Bell—an Abt Vice President and Senior Fellow—provides internal quality control as the team member responsible for the quality assurance of all evaluation documents and methodologies. In addition, Abt's Research and Evaluation Expertise Center and our Data Capture and Analytics Center will provide expert advice on our evaluation design, data collection, and analysis to help us ensure quality and rigor. Finally, our Journal Author Support Group will allow us to present our evaluation results and aid in preparing the results for publication.

References

aBi Trust (2014). *Synthetic budgets for common beans, groundnut, and soy.*

AgResults (2014). *Seed production in Uganda: Increasing smallholder farmer access to improved legume seed varieties: Pilot business plan.* AgResults: Innovation in Research and Delivery.

Akerlof, George (1970). The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 89.

Bardach, Eugene (2011). *A Practical Guide for Policy analysis: The eightfold path to more effective problem solving.* Thousand Oaks, CA: Sage.

Bloom, H. (2003). Using 'short' interrupted time-series analysis to measure the impacts of whole-school reforms: With applications to a study of accelerated schools. *Evaluation Review, 27*(1), 3–49.

Caves, R. (1987). *American Industry: Structure, conduct, performance.* Upper Saddle River, NJ: Prentice Hall.

Coryn, C. L. S., & Hobson, K. A. (2011). Using non-equivalent dependent variables to reduce internal validity threats in quasi-experiments: Rationale, history, and examples from practice. In S. Mathison (Ed.), *Really new directions in evaluation: Young evaluators' perspectives. New Directions for Evaluation, 131,* 31–39.

FAOSTAT (2015). Uganda legume statistics. Rome: Food and Agriculture Organization of the United Nations.

FOWODE (2012). *Gender policy brief for Uganda's agriculture sector,* Forum for Women and Democracy (FOWODE).

Geyer J, M Davis, T Narayan (2016). Intracluster Correlation Coefficients of Household Economic and Agricultural Outcomes in Mozambique. Evaluation Review I-20.

Guest, Greg, Emily E. Namey, and Marilyn L. Mitchell (2013). "Sampling in Qualitative Research." In Collecting Qualitative Data: A Field Manual for Applied Research, pp. 41-74. Thousand Oaks, CA: SAGE Publications.

Holtzman, J. S. (2003). Using subsector analysis to assess the impact of policy reform. In G. Ender and J. S. Holtzman (Eds.), Does agricultural policy reform work? The impact on Egypt's agriculture 1996-2002. Cambridge, MA: Abt Associates Inc., 111–134.

ISSD Uganda (2014). *Baseline report on farmers' access to seed and other planting materials,* Integrated seed sector development programme in Uganda. Kampala: Wageningen University and Research Center.

Kilimo Trust (2012). *Development of Inclusive Markets in Agriculture and Trade (DIMAT): The nature and markets of bean value chains in Uganda.*

Mainville, Denise (2015). Initial Qualitative Assessment Report for Uganda Improved Legume Seed Pilot, AgResults Impact Evaluation.

Mubangizi, E., D. Nandagire Ntamu, W. Mwesigwa Thembo and M. Thijssen (2012). *Uganda seed sector assessment.* Briefing Note, ISSD Africa.

North, D. C. (1990). *Institutions, institutional change and economic performance.* New York, Cambridge University Press.

Peterman, A., A. Quisumbing, J. Behrman and E. Nkonya (2010). *Understanding gender differences in agricultural productivity in Uganda and Nigeria.* Washington D.C., International Food Policy Research Institute.

Peterson, J., and Margaret Lewis (Eds.) (1999). Income distribution. *The Elgar Companion to Feminist Economics*, Cheltenham: Edward Elgar, 457-63.

Scherer, F. M. and D. Ross (1990). *Industrial market structure and economic performance.* Chicago, Rand McNally & Co.

Shadish, William R., Cook, Thomas D., and Campbell, Donald T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton, Mifflin and Company.

Sutton, J. (1992). *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration.* Cambridge, MA: The MIT Press.

Tetra Tech (2015). *USAID Feed the Future Agriculture Input: Seed sector workplan.*

Trochim, William, and James P. Donnelly. 2007. *The research methods knowledge base*, 3rd edition. Cincinnati, OH: Atomic Dog Publishing.

Uganda Bureau of Statistics (2010). *Uganda Census of Agriculture: Crop Area and Production Report.*

Uganda Bureau of Statistics (2012). *Uganda demographic and health survey.*

USAID (2014). *Uganda gender assessment.*

Yin, R. K. (2003). Case Study Research: Design and Methods. Thousand Oaks, Calif., Sage Publications.

Appendix A: SCP interview guide

Qualitative interview guide for seed companies

1. Summary identifying information
    a. Name of interviewee
    b. Position
    c. Organization/firm
    d. Contact information
    e. City
2. Brief characterization of firm and its activities with seed
    a. Ownership and management structure of firm (family owned and operated, multinational corporation, etc.)
    b. Scale of operations/industry position in seed market
    c. Prior to 2016, did the firm work with legume seed?
        i. If yes, please provide background in terms of what seed, what market outlets, and feedback on the experience of working in this market.
    d. Does the firm currently work with legume seed? If yes, please complete table below.
    e. What activities does the firm undertake in the legume market?
        i. Breeding/varietal development
        ii. Multiplication of foundation/basic seed from breeder seed
        iii. Multiplication and processing of seed for sale
        iv. Other

For 2015/16 (baseline period)	Bean	Soy
Share (%) of company seed sales		
Volume sold (MT)		
Volume unmet demand		
Buyers		
Direct to farmers		
NGOs/Donors		
Government (NAADS)		
Industry players (e.g., aggregators)		
Other (specify)		
Procurement		
% from own land		
% from direct contract farmers		
% from associations, LSBs		
% other		

    a. Procurement strategy
        a. What challenges has the firm faced with respect to procurement of legume seed? (As appropriate, discuss bean and soy separately.)
    b. Merchandising strategy (how distributes product, organizational and institutional arrangements, characteristics of buyers)
        a. What size packages does firm use for legume seed distribution?

b. What challenges has the firm faced in merchandising legume seed? (As appropriate, discuss bean and soy separately.)

3. Perspective on legume seed market (for beans and soy)
    a. How attractive is legume seed market relative to other markets the firm works in (maize, vegetables)?
    b. How is the market trending in general?
    c. What are important constraints to activities in this market? What are root causes of these constraints?
    d. Where does the firm see its activities heading in the coming years with regards to legume seed? Increase activity? Decrease? Why?
4. Perspective on AgVerify
    a. Does firm know of the AgVerify initiative? If yes,
    b. Describe what knows of initiative and how learned of it.
    c. What is firm's perspective on AgVerify and its potential to help the seed market develop?
    d. What is firm's perspective on the relevance of AgVerify to its own activities? Is it interested in participating? Why or why not?
    e. Does firm intend to pursue AgVerify certification? If yes,
        i. For maize?
        ii. For legumes?
        iii. What is firm's strategy for engaging with AgVerify?
        iv. How will firm use AgVerify to strengthen its own position in the market?
        v. What are constraints and/or reservations about working with AgVerify?
5. Perspectives on AgResults
    a. Is firm familiar with AgResults? (If not, describe)
        i. Is firm interested in/planning to participate? Why or why not?
        ii. What aspects of the incentive—annual prize, storage, end-of-pilot prize—does firm find most attractive? Why?
        iii. What challenges are anticipated if it does participate?
    b. Does firm intend to apply/has firm applied?
        i. What has been experience to date?
    c. Does firm see the pilot attracting other firms? What firms are most likely to want to participate and why?
    d. Does firm see the pilot succeeding in stimulating the emergence of a dynamic, private-sector-driven market? Why or why not?
    e. How do the different incentives relating to AgResults, and the AgVerify certification scheme, affect the firm's perspective on and participation in the market for improved legume seed?
    f. If other firms were to engage in the pilot, or independently increase their activity in the legume seed market, would this at some point affect the firm's own interest in getting involved in the market? In what way?
    g. Is there any potential downside to participating in the pilot?

Qualitative interview guide for agro-input dealers

1. Summary identifying information
    a. Name of interviewee
    b. Position
    c. Organization/firm
    d. Contact information
    e. City
2. Brief characterization of firm and its activities with seed
    a. Ownership and management structure of firm (family owned and operated, MNC, etc.)
    b. Prior to 2016, did the firm work with legume seed?
        i. If yes, please provide background in terms of what seed, what market outlets, and feedback on the experience of working in this market.
    c. Does the firm currently work with legume seed? If yes, please complete table below.

For 2015/16 (baseline period)	Bean	Soy
Share (%) of retailer's seed sales		
Volume sold (MT)		
Volume unmet demand (if any)		
Buyers		
Direct to farmers		
NGOs/Donors		
Government (NAADS)		
Industry players (e.g., aggregators)		
Other (specify)		
Procurement		
What brands legume seed?		
(list brands)		

    c. Procurement
        a. Does firm have trouble getting adequate legume seed to meet demand? Please describe, including when/how often this happens, and what strategies it takes when there are shortages.
        b. What other procurement challenges, if any, does firm experience with respect to legume seed?
    d. Merchandising
        a. What size packages does firm use for legume seed distribution?
        b. What are the characteristics of legume seed buyers (e.g., farmers versus organizations, scale of production)?
        c. Does firm do any re-packaging of legume seed to meet buyer demand for smaller units? Describe.
        d. What challenges has the firm faced in merchandising legume seed? (As appropriate, discuss bean and soy separately.)
3. Perspective on legume seed market (for beans and soy)

a. How attractive is legume seed market relative to other markets the firm works in (maize, vegetables)?
b. How is the market trending in general?
c. What are important constraints to activities in this market? What are root causes of these constraints?
d. Where does the firm see its activities heading in the coming years with regards to legume seed? Increase activity? Decrease? Why?

4. Perspective on AgVerify
    a. Does firm know of the AgVerify initiative? If yes,
    b. Describe what knows of initiative and how learned of it.
    c. What is firm's perspective on AgVerify and its potential to help the seed market develop?
    d. What is firm's perspective on the relevance of AgVerify to its own activities?
    e. Has firm had opportunity to stock AgVerify-certified seed? If yes, please describe.
    f. Is it interested in stocking AgVerify certified seed? Why or why not?
    g. What AgVerify-certified seed is firm most interested in stocking?
        i. Maize?
        ii. Legumes?
        iii. Would firm promote AgVerify-certified seed with its buyers? How?
        iv. What are constraints and/or reservations about working with AgVerify?
    h. Does firm know of any other local agro-input dealers stocking AgVerify-certified seed? If yes
        i. Please identify
        ii. What has been their experience (to firm's knowledge)
        iii. How does this affect the firm's own interest in participating?

Qualitative interview guide for farmers

Note 1: This guide assumes that basic contextual data on farmer production systems, scale, markets, etc., are collected in the quantitative survey.

Note 2: 'Legumes' is used to refer to bean or soy. Interviews will be targeted to major bean/soy areas, and interview questions will focus on the specific commodity based on this targeting.

1. Household legume production and consumption
    a. Are legumes an important food item for your household? If yes,
        i. What legumes (bean, soy, groundnut, other)?
        ii. How often do you eat them?
        iii. Do any household members eat more legumes than others? Who? Why?
        iv. Would you like your family members to consume more legumes? Why or why not?
    b. What do you do with legumes that you produce—how much do you sell versus keep for home consumption or other uses?
    c. What would you do if you produced more legumes? Sell more? Consume more? Why?
    d. How is the market for legumes?
        i. Is it attractive?
        ii. How is it trending?
        iii. Who do you sell to if/when you sell?
        iv. If you were to produce more legumes to sell, would you change how you market them? In what ways?
        v. If you were to produce more legumes to sell, would you change how you produce them? In what ways?
2. Intra-household dynamics
    a. Who is responsible for different activities involved in legume production? Who primarily performs the following activities? What are means of production?

Activity	Male/female	Hired/family/community	Mech/manual/animal/chemical
Land preparation			
Planting, weeding			
Harvest			
Processing			
Storage			
Sale/market			

    b. Who in the household makes decisions around legume production (what to produce, how much, how)?
    c. Who in the household makes decisions about sales of legumes (what, when, and how much to sell, and to whom)?
    d. Who in the household retains money from the sale of legumes? What is it used for?

3. Farmer perspectives on legume seed sources
    a. What seed (local versus improved) do you use for bean and soybean production?
    b. How does the seed perform?
    c. If you haven't used improved seed for legume, why not?
    d. How often do you renew your seed for these crops?
    e. Seed sources
        i. Where do you obtain legume seed from?
        ii. What other sources of legume seed are there?
        iii. Why do you choose the source that you do?
        iv. How much does seed cost (relative to food equivalent, relative to commercial sources)?
    f. Do you think there is a shortage of quality legume seed available? Have you ever wanted to buy legume seed and couldn't obtain it? Or had trouble getting the quality or type/variety of seed that you wanted? Describe.
    g. Where would you most like to buy legume seed from (e.g., agro-input dealer, LSB, project, association)? Why?
    h. Would you be willing to pay more for legume seed if you were confident of its quality? If yes, how much more?
    i. Have you ever purchased legume seed from a commercial agro-input dealer?
        i. Why or why not?
        ii. What was the experience? I.e., did the seed perform?
        iii. What are the benefits of purchasing from commercial seed sources?
        iv. What are the drawbacks of purchasing from commercial seed sources?
4. Any other thoughts on legume seed and sources?

Qualitative interview guide for other key informants

- Interviewee identification, contact information, and relevant to the pilot
  - Name of interviewee
  - Position
  - Organization/firm
  - Contact information
  - City
- What does respondent/organization do that relates to legume or legume seed production/markets?
- Does respondent see the lack of quality legume seed as a problem? If so, please characterize the problem.
- Is respondent familiar with the AgVerify initiative? (If no, briefly describe)
  - What does respondent know about AgVerify?
  - Does AgVerify have potential to help resolve the problem?
  - What would be challenges to successful implementation of this intervention (e.g., verification, obtaining necessary approvals)
  - What factors would affect the success of this initiative?
  - What seed companies are most likely to participate?
  - What issues and challenges does respondent see arising that might affect the initiative's success?
- Is the respondent familiar with AgResults? (If no, briefly describe.)
  - What is respondent's perspective on specific components of the pilot incentive and structure and how they will affect pilot outcomes?
  - How will the pilot incentives and mechanisms be perceived by the pilot's intended participants? How are they likely to respond?
  - Is the pilot tackling the right issues to achieve the objective of increasing legume seed production and consumption?
  - Is the pilot incentive likely to stimulate the investment that is sought? Why or why not? How will different market actors respond and why?
  - What will be required for the pilot to succeed?
  - What other issues are likely to influence the success of the pilot? How will they affect it?
  - What alternative approaches, if any, would the respondent expect to be more effective?
- Does respondent have any other observations or comments on the challenge of increasing farmers' access to improved legume seed?