Designing Quasi-Experimental Impact Studies of Agricultural Research at Scale

J.V. Meenakshi, Nancy Johnson, and Mina Karasalo

Technical Note N. 10

Designing Quasi-Experimental Impact Studies of Agricultural Research at Scale

J.V. Meenakshi, Nancy Johnson, and Mina Karasalo

June 2021

Contents

1 Introduction

To address the call for greater rigor in impact assessment, the credibility of causal claims is under careful examination (Stevenson et al., 2018). Experimental approaches may not be feasible for documenting the long-term and large-scale impact of agricultural research on human welfare, the environment, and other development outcomes (SPIA, 2020). In many such cases, quasi-experimental approaches—which aim to demonstrate causal evidence of the impact of an intervention on a set of outcomes—are the best option. While many estimation approaches exist, their causal claims rest on convincing and well-supported identification strategies.

In this technical note, we review several recent and ongoing quasi-experimental impact studies that offer promising approaches for measuring impact at scale. The studies cover different types of CGIAR-related innovations (ranging from crop varieties and fish strains to conservation agriculture and approaches to strengthening collective action) and different types of development outcomes (such as changes in national income, infant mortality, forest cover, and atmospheric pollution). These studies also employ innovative ways of combining a multiplicity of nationally-representative and global datasets to measure impact, thereby reducing the need for special-purpose surveys. Moreover, the studies offer detailed explanations of empirical strategies that rely on information and data on innovation rollout and diffusion over time and space to address the sources of endogeneity that traditionally plague ex post impact assessments.

The purpose of this technical note is to help CGIAR impact assessment specialists and practitioners elsewhere plan for and design high-quality impact studies at scale. With a focus on methodologies and data sources, we seek to collect and summarize attempts by the wider SPIA community to establish a credible basis for claiming causal impact of research on development outcomes. We do not emphasize impact results, and in most cases, these are not available yet as the studies are still ongoing. The goal is to familiarize readers with different approaches that are being used, and to facilitate learning that can help strengthen current and future impact assessments.

2 The Challenges of Measuring Impacts at Scale

While scale is a complex term in agricultural research for development (AR4D), in this note scale refers to the number of people or households reached by, and potentially benefiting from, an innovation. In the case of environmental outcomes, scale may be measured in units of land or other resources. Because the definition of scale is context specific and varies by innovation, it is difficult to establish standard thresholds for determining when an innovation has scaled. A threshold of, say, one million beneficiaries, as used for example by Kremer et al. (2019) in a different context, may be of little use to define large scale in the context of many CGIAR innovations. For example, one of the studies we review below considers environmental and health impacts of exposure to sustainable agricultural practices across 250,000 plots spread across Mexico. Another focuses on the impacts of stress-tolerant rice varieties in regions of Bangladesh that are prone to flooding in the *aman* rice-growing season (which is one of three). Other examples are national or global in scope.

The phrase *at scale* therefore can refer not only to the number of beneficiaries but also to the share of total expected beneficiaries or to aspects of the duration of dissemination efforts or use. What is common across the definitions of scale is that there is some degree of spontaneous, rather than incentivized, spread. Duration can be important for certain outcomes that can only be achieved as a result of use over

time. Environmental outcomes such as improved soil fertility may begin to manifest only after years of improved management practices. The opposite is also true; unintended negative consequences take time to appear. Impacts on social or demographic variables may result not only from long-term use but also from the general equilibrium effects of, for example, improved agricultural productivity on the broader economy. Thus, scale refers to the distance down the impact pathway from immediate outcomes such as agricultural productivity, to more distant outcomes related to human, social, and/or environmental welfare.

However they define scale, in all cases the studies reviewed here attempt to document the impacts of *big wins*. A big win is thought to have happened when anecdotal evidence suggests not only that research has been successful in generating new knowledge and technologies but also that these innovations have achieved widespread use and are generating benefits. The value of benefits from big win cases is expected to far outweigh the research and development costs of the entire portfolio (SPIA, 2020). Documenting the benefits of and returns to these big wins is important from an accountability perspective. That does not mean, however, that the results are not also useful to broader audiences of policymakers, practitioners, and academics nor that the standards of rigor are lower than for other types of impact evaluations.

One of the main reasons that measuring impact at scale is difficult relates to the endogeneity of adoption that results from self-selection (de Janvry et al., 2011). Outcomes for adopters cannot be directly compared with those of non-adopters unless access to the innovation is randomly assigned. While this is often the case in pilots, it is not a common way of rolling out agricultural innovations. This means that having good-quality data on adoption at scale, while needed, is not enough to measure impact. It is also necessary to account for both the observable and unobservable differences between adopters and non-adopters. Quasi-experimental identification strategies can help address endogeneity either by exploiting variation in variables that are correlated with adoption but are arguably exogenous so they can be used as instruments, or by using methods such as difference-in-difference or matching that are able to account for systematic differences between adopters and non-adopters. Balance tests and parallel trends can be used to assess whether evidence is consistent with the identifying assumptions, based on what is known about the sources or factors explaining variation in the rollout and/or access to the innovation. What constitutes sufficient evidence of support for identifying assumptions will necessarily be context specific and vary from case to case.

Identifying which innovations will ultimately become big wins, when, and where takes time and is hard to predict. As a consequence, planning rigorous impact studies in advance can be a challenge. In many cases it will be necessary to use quasi-experimental approaches to measure impacts at scale, and these approaches will depend on being able to identify appropriate datasets after the fact, often repurposing data that were collected to meet a different objective. Doing this well will always require some creativity, and determining whether a causal claim is convincing enough is ultimately a judgment call. A growing number of good examples are becoming available. This note attempts to document some of them in one place.

3 Global and National Assessments of the Long-Term Impacts of the Green Revolution

Gollin et al. (2021), von der Goltz et al. (2020), and Bharadwaj et al. (2020) are good examples of the overall approach of exploiting past diffusion patterns and using appropriate strategies to quantify the

impacts of CGIAR interventions on various outcomes. They consider the *set* and *stream* of adoption of several high-yielding crop varieties (HYVs) across several countries (von der Goltz et al., 2020, Gollin et al., 2021) or across districts in India (Bharadwaj et al., 2020) over several decades. However, they each use a somewhat different strategy to quantify the causal impact of the Green Revolution on outcomes that are realized only following a relatively long impact pathway. These outcomes include national incomes, population growth (Gollin et al., 2021), and child health (von der Goltz et al., 2020).

Gollin et al. (2021) first make the case that the pattern of spread of high-yielding varieties across 80 countries and over time (as reported in Evenson & Gollin, 2003) was determined largely by factors that could plausibly be exogenous to each country, relating to the establishment of institutions that were to become CGIAR centers and the crop-specificity of available technologies. They demonstrate, using an event study[1], that the yield growth induced by the introduction of HYVs was not driven by preexisting differential trends in yields. This finding enables them to establish the counterfactual increase in yield growth that would have occurred in the absence of the Green Revolution, which forms the basis of the instrumental variable they employ. In other words, the instrument represents the increase in yields that would have occurred across crops, keeping land and labor unchanged, aggregated using pre–Green Revolution area shares of various crops as weights. They conclude that the adoption and diffusion[2] of HYVs led to substantial increases in national incomes (relative to scenarios where there had been no Green Revolution or it had been delayed by a decade), and to a consequent reduction in population growth.

Von der Goltz et al. (2020) consider the extent to which the Green Revolution enabled increases in food supply that in turn prevented infant mortality, using three distinct types of data. They calculate infant mortality rates using mother's recall of child deaths from the Demographic and Health Surveys, and exploit the fact that these data are georeferenced. For diffusion, they combine national-level adoption data (from Evenson & Gollin, 2003) with gridded crop maps (with 10-km by 10-km cells) to derive spatially and temporally disaggregated information on the spread of modern varieties. The overlap between the availability of adoption and georeferenced infant mortality information yielded an assessment that encompassed 37 countries. The key to their identification strategy is their use of the agroecological specificity of crop cultivation. In particular, their argument is that HYV share in a particular area and time is driven by crop suitability, which in turn is determined by exogenous agroclimatic characteristics. The modern variety diffusion indicator they use is a weighted average of the national-level diffusion of various crops, with weights determined by area shares determined through crop maps. It is this computed, disaggregated, cell-specific diffusion that serves as a source of exogenous variation. They repeat the exercise using different crop maps as a check on robustness. Their results show that an increase in the share of HYVs to 50 percent is associated with a decrease in infant mortality of about 33 deaths per 1,000 births. This estimation strategy also enables them to provide supportive evidence showing that increased food supplies are the causal mechanism at work. They are unable, however, to discern the effects on childhood stunting (if anything, the signs are perverse).

---

[1] An event study can be thought of as a generalization of a difference-in-difference framework, with different units (say households) being treated at different periods of time. Thus there are a series of (say m) 'post' dummy variables (rather than a single post dummy variable in the basic DiD), representing households that began to be treated j (j = 1,…,m) periods ago. The identification stems from a comparison of outcomes of treated households with (a) pure control households (that is, those that will never receive the treatment) and (b) units that are not yet treated.
[2] In keeping with the literature, we distinguish between diffusion, implying the natural (and often unmonitored) spread of adoption of an innovation, and dissemination, connoting organized and structured attempts to introduce new agricultural innovations. Diffusion is typically preceded by dissemination efforts, especially for agricultural innovations.

Bharadwaj et al. (2020) address the same question (that is, how the Green Revolution led to reductions in infant mortality) but focus only on India. They use districts (subunits of states) as the unit of analysis and rely on official statistical records on the adoption of HYVs. They then compare two children from the same district and with the same mother who were exposed to different levels of HYV adoption depending on the year of their birth, over and above any unobserved shocks to infant mortality that differ by year of birth and any long-term changes in the region of birth. They use an event-study specification to provide support for their key identifying assumption, by interacting eventual HYV adoption with year fixed effects and showing that there were no differential time trends in mortality before the Green Revolution. They show substantial decreases in infant mortality that can be attributed to the adoption of Green Revolution varieties and are able to rule out factors such as investments in child health as the mechanism by which these decreases were achieved. Other work, notably by Brainerd & Menon (2014), points to the adverse consequences of the intensive use of fertilizers associated with the Green Revolution on seasonal child health and mortality outcomes.

4 Impacts of Recent CGIAR Innovations

Similar strategies can be used to assess the impacts of particular CGIAR innovations (in contrast to the aggregate of the Green Revolution), including innovations that are more recent or are from research areas outside of crop breeding. We provide summaries below of selected ongoing studies that are attempting to use past dissemination data, in combination with data from other sources, to assess ex post impact.[3] For more detailed information about these studies and any updates on estimation strategy, please refer to the 2021 SPIA webinar series page, where recordings from presentations by all three study teams are available.

- **Conservation agriculture, Mexico.** The impact study of the Sustainable Modernization of Traditional Agriculture program (MasAgro) is notable for at least two reasons: it is one of the few assessments of conservation agriculture at scale, and it successfully integrates monitoring and evaluation data into its empirical strategy. This administrative database, designed and compiled with technical support from CIMMYT, was integral to the scale-up strategy in Mexico. The authors note that the rapid scale-up of the extension services meant that it was unanticipated at a grid cell level, yielding, in turn, a substantial variation in the timing of exposure to treatment (extension services about conservation agriculture) across plots. They employ an event-study strategy to identify causal impacts. By combining satellite imagery on fires on agricultural land (as distinct from unrelated urban or forest fires), they are able to establish that conservation agriculture led to reduced burning of crop residue. In particular, there was a reduction in the incidence of crop fires in cells that received greater extension services. They use another administrative dataset on registered infant deaths and map them back to the region of residence to produce an outcome measure. Using natural exogenous variation in wind direction affecting exposure to air pollution resulting from fires, they are able to show impacts on infant deaths as a consequence of lower exposure to air pollution from fires. To the extent that that this study uses information on dissemination rather than diffusion, these results can be vested with an intent-to-treat interpretation.
- **Stress-tolerant rice varieties (STRVs), Bangladesh.** Because large parts of Bangladesh are prone to flooding in the post-monsoon season, popular rice cultivars were bred to tolerate submergence for longer periods of time. The target area for the STRVs is therefore defined as all *aman* (post-monsoon) rice susceptible to floods. Using administrative data from national

---

[3]We highlight these studies because of the way they use dissemination data to look at impacts on particular outcomes. In some cases, the studies include other components looking at additional outcomes that we do not review here. SPIA is also supporting other studies using quasi-experimental approaches. For more information please visit: cas.cgiar.org/spia/impact-evidence

partners, researchers have undertaken a retrospective exercise to track the location and quantity of STRV seed produced and disseminated across all districts in Bangladesh, including in areas that are relatively more flood prone, over time. They can then map this information on the spatial and temporal variation in dissemination of STRVs onto the incidence of floods as estimated by satellite imagery. The location and duration of floods act as a natural experiment that allows researchers to quantify the extent to which regions that were exposed to sustained sales and dissemination of STRVs were able to insulate against losses in rice yields relative to those that saw relatively little or no dissemination. They use a proxy measure for yield loss based on the normalized difference vegetation index (NDVI). Subsequent efforts will focus on further ground-truthing yield losses as derived from the NDVI.

- **Approaches to restoring degraded common lands, India.** To address land degradation in many of India's ecologically fragile ecosystems, the Foundation for Ecological Security (FES) drew on CGIAR research to design a location-specific intervention model founded on the tenets of collective action and property rights to restore degraded common land. Since the early 2000s, FES has scaled out its intervention model to approximately 20,000 villages, comprising over 5.5 million acres of common land and 6.25 million people. To assess the extent to which the interventions contributed to the restoration of common lands, the researchers obtained secondary data on the variables FES used as selection criteria, both for villages in which it has intervened, and for villages in which it will intervene in the future. The researchers plan to use information on the phased rollout to determine whether areas that had been treated for longer were more likely to see larger impacts through a difference-in-difference strategy. Their outcome measures rely both on remote-sensing data on vegetative cover in georeferenced locations over time and on information canvassed through a survey.

5 Using Dissemination/Diffusion Data in Designing Impact Studies

The examples reviewed in the preceding sections relied on information and data about the spread of innovations to construct identification strategies. The studies in Section 3 use, to different degrees, estimates of the timing and diffusion of HYVs to claim exogenous variation at the national and subnational levels. Gollin et al. (2021) use the differences in the timing of development and introduction of HYVs, to propose an instrument that can be considered exogenous to the individual country. Similarly, von der Goltz et al. (2020) argue that the timing of the onset of the Green Revolution was exogenous to individual countries, as it was the result of international crop research programs. They exploit this variation in conjunction with differences in the agroecological suitability of crop cultivation (again exogenously given) as measured by crop maps to generate variation. Another approach is pursued by Bharadwaj et al. (2020), who use an event study to show there were no differing trends in infant mortality before the introduction of new technologies across districts. To assess the impacts of HYVs, they exploit the variation in the year of birth of children who were born in the same area (or even had the same mother) but were, as a consequence of year of birth, exposed to different levels of HYV diffusion.

The studies in Section 4 use information and data on dissemination efforts over time and space (regions/villages/localities) to build identification strategies. For the FES program, dissemination happened according to a specific set of targeting criteria—11 specific and observable criteria were selected to identify areas that were ecologically fragile, degraded, and marginalized and had high proportions of poor, tribal, landless, and smallholder populations. They then used these data to compute propensity scores and undertook a matching exercise based on these scores to obtain a set of treated and untreated villages. They were able to exploit village-level information from the national census of 2001 and 2011 to identify sites similar to the selected sites, at the time of selection. Using these same historical data, they also confirmed that the targeted sites did indeed match the original targeting

criteria. It is possible that selection bias can still operate at the community level. This is less likely to be a problem at the more aggregated level of a district. In the FES case the assumption is that any remaining unobservable features that governed site selection would be differenced out. The researchers plan to undertake a second round of matching following primary data collection to generate single difference estimates of impact on household level variables (such as income).

The STRV study relies on administrative data on temporal and spatial variation in seed production and sales/distribution. The authors were able to track STRV seeds from regions where they were produced (which are typically not flood-prone) to the flood-prone districts in which they were distributed. This involved collating data across both public and private seed sectors, and over time from 2010 to 2018, although the public sector is the predominant seed supplier. Some flood-prone districts had different diffusion patterns over time, but may have been exposed to similar exogenous flooding events. Both sets of information are therefore necessary for identification.

The MasAgro study uses dissemination data from CIMMYT's monitoring and evaluation system. In particular, it employs an administrative dataset, Bitacora Electronica MasAgro, which contains georeferenced information on plot, farmer, and commodity characteristics for nearly 250,000 plots over a period of eight years, but data from before the implementation of MasAgro go back 19 years. Since these data were plot-specific and georeferenced, the authors could map this monitoring and evaluation database with remotely sensed information on fires, account for wind direction, and relate extension efforts to infant deaths via lower crop burning. Key to the ability to identify impact was CIMMYT's investment in the monitoring and evaluation system, which made it possible to collate micro-level data and provide real-time feedback that incentivized farmers and extension agents to actively engage with it.

In designing a rigorous impact assessment study using dissemination data, one key piece of information is why dissemination occurred the way it did. The FES study provides a clear example of how explicit information about the *why*, namely the targeting criteria, can be used to build a strong identification strategy. Even in cases when the *why* is not explicitly addressed, it is important to consider. A strong identification strategy ensures that, if there is incomplete information about the *why*, there is at least enough information to need to know it. Empirical tests such as pre-trend analysis or placebo tests can be used in support of the identification strategy, so that any unknown factors influencing rollout are assumed to be consistent with it.

Unlike adoption/diffusion data, dissemination data has not been a priority for collection by impact evaluation specialists. Some of the approaches reviewed here suggest that dissemination data, while clearly not a substitute for adoption/diffusion data, can potentially be valuable for designing rigorous impact evaluations. Collecting these data may cost relatively little. Even where data are not regularly compiled by public agencies or by large-scale development programs, it may be possible to reconstruct the data for cases of potential big wins (Box 1). Going forward, collection of dissemination data can be built into the monitoring and evaluation systems for large-scale dissemination programs (Box 2). This is also an opportunity to think about monitoring, evaluation, learning, and impact assessment (MELIA) in an integrated manner.

Box 1. Supporting the collation of dissemination data for an ex post assessment

SPIA is working with several study teams that have identified potential big wins but do not have the necessary data to justify and design a rigorous impact study.

Improved short-season lentil varieties developed by ICARDA and partners in Bangladesh are thought to have enhanced the diversity of its predominantly rice-based cropping system, improving diets and the environment. Following the example of the STRV study, ICARDA set out to gather historical data on dissemination of improved lentil varieties from the Bangladesh Agricultural Development Corporation (BADC). As there was no systematic documentation of historical data, the ICARDA team asked regional offices to compile data from paper files and in some cases visited the regions and districts, to help compile data at subdistrict level. A similar process was followed with the two other major lentil seed distributors in Bangladesh.

Genetically improved farmed tilapia (GIFT), first introduced in Bangladesh by WorldFish and the Bangladesh Fisheries Research Institute in 1994. There is widespread adoption of GIFT (Hamilton et al., 2020), and WorldFish is keen to estimate the impacts on a range of outcomes. No data are publicly available on GIFT dissemination through private hatcheries. Further, private hatcheries do not limit sales to certain geographic areas. Thus, the challenge is not only to obtain dissemination data but also to determine whether it is possible to link sales of GIFT from specific hatcheries to specific geographies. To identify hatchery catchment areas, the WorldFish team compiled a list of tilapia hatcheries and surveyed them to elicit information on their customers and on location and price of their GIFT and non-GIFT sales. They also asked where hatcheries sourced brood stock. The information was collated to define a set of potential catchment areas at the subdistrict level which were then validated through farmer surveys. This triangulation exercise confirmed the catchment areas and documented widespread GIFT diffusion.

Orange-fleshed sweet potato and high-iron beans, developed by CIP and CIAT, have been widely disseminated in Uganda. Several large-scale adoption studies have been conducted and one is planned for 2021 (SPIA, 2021). HarvestPlus tracks seed production and dissemination at the national level, but sub-national data are kept by a range of national and local actors. To get a more detailed picture of where, when, and why planting material and trainings were disseminated across Uganda, SPIA led efforts in 2019 to reconstruct data at the subcounty level over the period 2009–2019. Existing data from monitoring and evaluation (M&E) systems were compiled, standardized, and assessed for quality and completeness. To fill identified gaps, a series of regional workshops were undertaken with stakeholders during 2019-2020. The result was both a complete mapping of the dissemination and sensitization efforts in the region and an approach that could be adapted for other innovations and/or locations.

The AfricaRice-SAED-ISRA (ASI) rice thresher was developed and released in 1997 with the aim of reducing postharvest losses, alleviating postharvest labor bottlenecks, and ensuring that the threshed rice was free of the dirt and debris associated with manual processes. The thresher has been disseminated in many countries in West Africa. Evidence exists on its effectiveness, and there is anecdotal information on spread. AfricaRice would like to estimate impact on postharvest losses, income, and women's labor at scale. Researchers are compiling data on the production and dissemination of threshers in two key countries, Nigeria and Senegal. They are collecting data on the number of ASI threshers manufactured per year since introduction (1997 and 2015) and the total number of ASI threshers available in each district and year. Data will be collected from the Ministry of Agriculture, the extension service, local manufacturers, and service providers.

Box 2. Planning for future impacts

Aflasafe, a set of biocontrol products for mitigating aflatoxin contamination, is in the process of going to scale in Africa for use on maize (10 countries), groundnuts (9 countries), and sorghum (in Ghana) (Konlambigue et al., 2020). Each Aflasafe product contains four atoxigenic isolates (of the fungus Aspergillus flavus, which causes aflatoxin contamination) that are native to the target country. Aflasafe products have been registered with pesticide regulatory authorities for use in each country, beginning with Nigeria in 2014 and mostly recently in Malawi in 2020 (Bandyopadhyay et al., 2019). Products are also being developed for another 10 countries. To date, Aflasafe has been widely promoted at scale only in Nigeria, as part of the AgResults Nigeria Aflasafe project (2014–2019). In other countries, IITA is in the process of implementing a scaling strategy that involves identifying and contracting licensing and distribution partners and developing a commercialization strategy (Konlambigue et al., 2020). With support from SPIA, IITA's socioeconomic and M&E teams are working with the Aflasafe technical team to develop and put in place a system for tracking the spatial and temporal rollout of Aflasafe across countries. The multicountry nature of the rollout, the stage it is currently at, and the key role that IITA plays in providing technical support to public and private sector partners in each country, are some of the factors that make this case both feasible (in terms of access to data) and promising in terms of measuring potential future impact.

6 Discussion and Conclusions

The studies reviewed here use rigorous quasi-experimental approaches to support causal claims about the impacts of agricultural research on development outcomes. The examples confirm the value of collecting diffusion data at scale. They also highlight the potential of dissemination data to strengthen study design. There has been little systematic effort to compile data on dissemination of CGIAR innovations. Given the potential value of these data and, at least in some cases, their relatively low cost, it is worth exploring how this collection can be done better, especially moving forward. As more CGIAR research programs partner with large-scale development efforts to increase the likelihood of achieving outcomes in line with their theories of change, opportunities for building dissemination data collection into program monitoring should increase.

The studies and approaches reviewed here not only provide examples of identification strategies but also show how a multiplicity of data sources—ranging from remote sensing and administrative and municipal records to the Demogaphic and Health Surveys—can be effectively intertwined to build a convincing case. In each case, the choice of data sources was driven by the theory of change, especially when the outcomes are further along the causal chain, as described in Section 2. The fact that the variables within these diverse datasets—for both the interventions and the outcomes—are georeferenced is key to being able to link them. This feature also removes some of the data collection burden from research teams, who can focus on assembling existing datasets rather than collecting new data.

Most of the completed and ongoing studies mentioned here were designed after the impacts were hypothesized to have occurred. However there is also a case for considering impacts *prospectively*, as for example is the case with several large-scale dissemination initiatives being planned by CGIAR and its partners. Approaching impact in these situations in a prospective manner has the advantage that the theories of change can be used to inform the selection of potential outcome variables, for which appropriate secondary data sources could be identified from the growing number of available survey and

geospatial datasets. The Aflasafe intervention described above is a good example that suggests that an integrated and prospective-looking MELIA approach is promising, and likely achievable at a relatively low cost.

It is worth reiterating that a focus on documenting the *why* of dissemination, and its contribution to identification, does not substitute for collecting high-quality diffusion data. Supporting the collection of adoption data at scale is a core part of SPIA's mandate (Evenson & Gollin, 2003; Walker & Alwang, 2015; Stevenson & Vlek, 2018; Kosmowski et al., 2020). SPIA has made important contributions to methods for collecting adoption data, focusing on measurement error (Stevenson, Macours, & Gollin, 2018), and making data available in the public domain for users[4]. While diffusion data are valuable on their own as evidence of uptake and of likely benefits for adopters, the possibility of using these datasets in rigorous impact assessments enhances their value, especially over time.

The methods outlined above are useful ways of establishing *average* treatment effects at scale on welfare metrics such as incomes and health, as well as environmental outcomes in the form of reduced air pollution or arrested land degradation. These methods may be extended to look at impact heterogeneity where suitable levels of disaggregation and statistical power are available.

---

[4] Datasets from several SPIA-supported studies are available on the SPIA dataverse. SPIA has also awarded 14 small grants to early-career researchers for studies of agricultural innovations in Ethiopia making use of the datasets of the Ethiopia Socioeconomic Survey (ESS), where SPIA has worked in partnership with the Central Statistics Agency of Ethiopia (CSA) and the LSMS-ISA team at the World Bank to integrate data collection protocols for scaled CGIAR-related innovations.

References

Bandyopadhyay, R., Atehnkeng, J., Ortega-Beltran, A., Akande, A., Falade, T.D.O., & Cotty, P.J. (2019). "Ground-Truthing" Efficacy of Biological Control for Aflatoxin Mitigation in Farmers' Fields in Nigeria: From Field Trials to Commercial Usage, a 10-Year Study. *Frontiers in Microbiology* 10, 2528.

Bharadwaj, P., Fenske, J., Kala, N., & Mirza, R. A. (2020). The Green Revolution and infant mortality in India. *Journal of Health Economics,* 71, 102314.

Brainerd, E., & Menon, N. (2014). Seasonal effects of water quality: The hidden costs of the Green Revolution to infant and child health in India. *Journal of Development Economics,* 107, 49–64.

de Janvry, A., Dunstan, A., & Sadoulet, E. (2011). *Recent advances in impact analysis methods for ex-post impact assessments of agricultural technology: Options for the CGIAR*. Report prepared for the workshop "Increasing the rigor of ex-post impact assessment of agricultural research," organized by the CGIAR Standing Panel on Impact Assessment (SPIA), 2 October 2010, Berkeley, CA. Rome: Independent Science and Partnership Council Secretariat.

Evenson, R. E., & Gollin, D. (2003). Assessing the impact of the Green Revolution, 1960 to 2000. *Science,* 300(5620), 758–762.

Gollin, D., Hansen, C. W., & Wingender, A. (2021). *Two blades of grass: The impact of the Green Revolution*. *Journal of Political Economy*.

Hamilton, M.G. Lind, C.E., Barman, B.K., Velasco, R.R., Danting, M.J.C. & Benzie, J.A.H. (2020) Distinguishing between Nile tilapia strains using a low-density Single Nucleotide Polymorphism panel. *Frontiers in Genetics*

ISPC (Independent Science and Partnership Council). (2018). *Adoption and impact of improved lentil varieties in Bangladesh, 1996–2015*. Brief no. 63. Rome: ISPC.

Konlambigue, M., Ortega-Beltran, A., Bandyopadhyay, R., Shanks, T., Landreth, E., & Jacob, O. (2020). *Lessons learned on scaling Aflasafe® through commercialization in Sub-Saharan Africa*. A4NH Strategic Brief. Washington, DC: International Food Policy Research Institute.

Kosmowski, F., Alemu, S., Mallia, P., Stevenson, J., & Macours, K. (2020). *Shining a brighter light: Comprehensive evidence on adoption and diffusion of CGIAR-related innovations in Ethiopia*. Rome: Standing Panel on Impact Assessment (SPIA).

Kremer, M., Gallant, S., Rostapshova, O. & Thomas, M. (2019). Is development innovation a good investment? Which innovations scale? Evidence on social investing from USAID's Development Innovation Ventures. Research working paper. https://scholar.harvard.edu/files/kremer/files/sror_div_19.12.13.pdf.

Rashid, S., & Zhang, X., eds. (2019). *The making of a Blue Revolution in Bangladesh: Enablers, impacts, and the path ahead for aquaculture*. Washington, DC: International Food Policy Research Institute. https://doi.org/10.2499/9780896293618.

SPIA (Standing Panel on Impact Assessment). (2018). *Adoption and impact of improved lentil varieties in Bangladesh, 1996–2015*. Brief no. 63. Rome: SPIA.

SPIA. (2020). *SPIA's approach to impact assessment for CGIAR*. Technical Note no. 8. Rome: SPIA.

SPIA. (2021). Uganda AGDPG virtual meeting: Measurement of innovations in national surveys in Uganda. https://cas.cgiar.org/spia/events/uganda-agdpg-virtual-meeting-measurement-agricultural-innovations-national-surveys.

Stevenson, J., Macours, K., & Gollin, D. (2018). *The rigor revolution in impact assessment: Implications for CGIAR*. Rome: Independent Science and Partnership Council (ISPC).

Stevenson, J. R., & Vlek, P. (2018). *Assessing the adoption and diffusion of natural resource management practices: Synthesis of a new set of empirical studies*. Rome: Independent Science and Partnership Council (ISPC).

von der Goltz, J., Dar, A., Fishman, R., Mueller, N. D., Barnwal, P., & McCord, G. C. (2020). Health impacts of the Green Revolution: Evidence from 600,000 births across the developing world. *Journal of Health Economics, 74*, 102373.

Walker, T. S., & Alwang, J., eds. (2015). *Crop improvement, adoption and impact of improved varieties in food crops in sub-Saharan Africa*. Wallingford, UK: CABI.