Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *Paths of Convergence for Agriculture, Health, and Wealth*

Agriculture, nutrition, and health in global development: typology and metrics for integrated interventions and research

William A. Masters,[1] Patrick Webb,[1] Jeffrey K. Griffiths,[1,2] and Richard J. Deckelbaum[3]

[1]Friedman School of Nutrition Science and Policy, Tufts University, Boston, Massachusetts. [2]Department of Public Health, School of Medicine, Tufts University, Boston, Massachusetts. [3]Mailman School of Public Health, Columbia University, New York, New York

Address for correspondence: William A. Masters, Friedman School of Nutrition Science and Policy, Tufts University, 150 Harrison Ave., Boston, MA 02111. william.masters@tufts.edu

Despite rhetoric arguing that enhanced agriculture leads to improved nutrition and health, there is scant empirical evidence about potential synergies across sectors or about the mix of actions that best supports all three sectors. The geographic scale and socioeconomic nature of these interventions require integration of previously separate research methods. This paper proposes a typology of interventions and a metric of integration among them to help researchers build on each other's results, facilitating integration in methods to inform the design of multisector interventions. The typology recognizes the importance of regional effect modifiers that are not themselves subject to randomized assignment, and trade-offs in how policies and programs are implemented, evaluated, and scaled. Using this typology could facilitate methodological pluralism, helping researchers in one field use knowledge generated elsewhere, each using the most appropriate method for their situation.

Keywords: integration; agriculture; health; nutrition; research; development

Introduction

Agriculture, nutrition, and health are interrelated through biological and behavioral pathways that make these sectors distinctively tied to ecological and socioeconomic conditions. Place-based factors constrain and influence peoples' nutrition and health outcomes at every scale and in each context. For example, rainfall, temperature, and soil nutrients interact with crop genetics to influence plant growth and food quality as well as farm incomes, which in turn influence food purchase and health-seeking behaviors, all of which combine to influence individuals' nutrition and health outcomes. Ecological conditions also influence the growth and reproduction of pathogens, parasites, and disease vectors of all kinds, weighing heavily on the effectiveness of interventions aimed at meeting development goals.

Since the 2007/2008 food-price crisis, increasing attention to agriculture and nutrition as a factor in health and development has led donors and governments to pursue more integrated national plans, with a view to capitalizing on potential synergies between sectors in each location.[a] For example, countries as diverse as Nepal, Haiti, and Kenya all promote intersectoral coordination and interministerial collaboration to achieve common

[a]This wave of attention to integration is, in some ways, a return to efforts of the 1970s articulated by Winikoff: "Is there any way in which programs can be effective if they focus on any single problem without being combined in a so-called 'integrated attack'?"[1] Some of the most prominent interventions of that era were integrated rural development projects,[2] inspired in part by the systems-modeling approach popularized by *Limits to Growth*.[3] What has changed since then is a sharp reduction in the extent of global poverty,[4] and three decades of accumulated capacity to achieve and evaluate success in eradicating the remaining pockets of geographically concentrated extreme deprivation, combined with breakthroughs in research capacity ranging from spatial geocoding to genome sequencing.

doi: 10.1111/nyas.12352

goals around enhanced nutrition, health, and food security.[5–7] Indeed, the call for integrated action represents a new global agenda, as highlighted by the L'Aquila Joint Statement on Global Food Security, which argued that "food security, nutrition and sustainable agriculture must remain a priority issue on the political agenda, to be addressed through a cross-cutting . . . approach."[b,8]

The potential gains from integrated interventions call for enhanced research methods that explicitly account for biological and ecological relationships at their natural scale, to reveal regional-level effects that may be quite different from the sum of individual-level changes potentially observed in a limited-scale randomized control trial. However, the increasing focus of policies and programs on intersectoral integration to solve location-specific problems poses deep challenges for researchers, regarding how the value added from integration is most appropriately measured. The challenge is illustrated by controversy and confusion regarding what constitutes appropriate evidence in relation to multisectoral programming at the village level[9] or in economic development more generally.[10,11] Programs that seek to link food, water, health, and nutrition at a regional scale are already being implemented around the globe, but few have been rigorously analyzed, and most remain focused on measuring siloed outcomes within sectors and then aggregating these as if their impact equaled the sum of their parts. To measure synergies, we need a research agenda that explicitly addresses intersectoral linkages and the cost-effectiveness, replicability, and scalability of integrated processes.

This paper describes plausible causal pathways from agriculture, nutrition, and health interventions to measurable outcomes at their natural scale and provides a new typology of interventions and outcomes to help researchers estimate the effectiveness of combined program elements. By natural scale, we mean the smallest demographic or geographic unit over which the intervention can occur and potentially be replicated or randomized. The natural scale for medical intervention is typically the individual, whereas for agriculture it may be an ecosystem or marketplace. Various aspects of an intervention may have different scales and units of observation. The simplest such problems can readily be handled with clustered research designs, but combining interventions at different scales greatly increases the problem's complexity.

Although our typology of interventions and outcomes could be applied to any number of integrated sectors, we focus here on just the agriculture–nutrition–health nexus that has received heightened attention recently in the context of successive world food-price crises, new global commitments to scaling up nutrition actions, and concerns over the ecological stability of natural resources that underpin global food production and health outcomes. The adoption of this kind of integration typology would allow researchers to describe policy or program interventions and choose research methods to evaluate their success in ways that reveal system-level interactions among sectors and across regions that are typically ignored by traditional evaluations focused on individual changes in individual subjects.

Interactions across scales and sectors: a typology of interventions and outcomes

Table 1 provides a typology of program interventions and outcomes designed to reveal interactions between agriculture, nutrition, and health at various scales. The rows show specific interventions and desired outcomes in the agricultural, nutrition, and health sectors, as conventionally defined. The outer columns show place-specific attributes of regions, which serve as effect modifiers for person-specific interventions and outcomes shown in the inner columns. The potential causal mechanisms involved connect everything with everything else, both horizontally between interventions and results at various scales, and vertically among these sectors as well as others such as education, finance, or infrastructure. As shown in the bottom row of Table 1, the research techniques needed to identify these causal relationships depend crucially on the geographic scale of analysis.

[b]The L'Aquila Joint Statement was endorsed by the G8 countries plus Algeria, Angola, Australia, Brazil, Denmark, Egypt, Ethiopia, India, Indonesia, Libya (Presidency of the African Union), Mexico, the Netherlands, Nigeria, the People's Republic of China, the Republic of Korea, Senegal, Spain, South Africa, Turkey, the Commission of the African Union, the FAO, the IEA, the IFAD, the ILO, the IMF, the OECD, the Secretary General's UN High-Level Task Force on the Global Food Security Crisis, the WFP, the World Bank, and the WTO.

Table 1. **A typology of interventions and outcomes in agriculture, nutrition, and health**

	Interventions		Outcomes	
	Regional (places)	Individual (people)		Regional (places)
Agricultural interventions and outcomes	• Locally adapted genetics and other crop or livestock improvement • Local infrastructure, institutional, and policy improvement • Local soil and water management, pest control, and other agro-ecological improvement	• Access to or provision of seeds and other inputs • Access to or provision of market services and information	• Higher productivity • Higher income • Improved diets • Reduced toxin exposure	• Local wages and employment opportunities • Local credit and insurance options • Local land, water, and other resources
Nutritional interventions and outcomes	• Locally appropriate fortification, supplementation, and food quality assurance • Locally appropriate services and information for child care and breastfeeding	• Access to or provision of food nutrients services and information • Access to or provision of food safety services and information	• Improved behaviors, physical growth, and cognitive development • Improved micronutrient status • Enhanced reproductive potential	• Local supply of diverse and nutritious foods • Local norms regarding diet, infant feeding, hygiene, and sanitation
Health interventions and outcomes	• Local water, sanitation, and hygiene • Local deworming, vaccination, and vector control • Local health system services for all-cause prevention and treatment	• Access to or provision of healthcare products, services, and information	• Lower morbidity and mortality • Enhanced human productivity • Improved maternal and child health	• Local exposure to disease • Local supply of healthcare services
Integrated research methods		Impacts and causal mechanisms can be identified using both natural and controlled experiments varying exposure across individuals		
	Impacts of region-wide interventions can be identified only by inference from observational data, using knowledge of individual-level causal mechanisms			

The middle columns of Table 1 list interventions and outcomes that can be identified at the level of individuals and households. The causal relationships involved have long been subject to both observational and experimental study, primarily across rows within the sectoral boxes. Most often, a specific intervention is associated with a specific outcome. Ideal experimental designs using fully double-blind assignment, preregistration of research protocols, and other criteria may not be feasible for many interventions, but researchers are increasingly able to use some degree of randomization from both controlled and natural experiments to identify mechanisms and magnitudes of causal linkages. In each case, the unit of observation for these impacts is an individual person, exposed to a particular treatment or intervention at a particular time and place.

The outer columns of Table 1 show regional effect modifiers. These attributes of a particular place are themselves subject to change, influenced by regional-scale policies and programs, but by definition they are conditioning factors that do not vary among the population of intended beneficiaries or observed subjects affected by the specific interventions listed in the middle columns of this typology. Such places can be defined at any scale up to and including the world as a whole. The causal effect of interventions in the middle columns can be identified through randomization over a variety of scales, for example, in studies that randomize over clusters or exploit natural experiments at the level of a district or state, but the impacts of many important interventions in the outer columns of Table 1 must be inferred from observational data and knowledge about causal mechanisms in the inner columns.

The typology of interventions and their effects reveals how increasingly integrated agriculture–nutrition–health policies and programs can and should be subject to increasingly diverse research methods. Causal mechanisms within and across rows in the inner columns can be identified by complex research designs designed to capture interaction effects, and variation in the conditioning factors around any given study can then be addressed by conducting multiple studies in diverse locations. A meta-analysis of those results can then be used to identify which causal relationships are robust to variation in these potential effect modifiers. But by definition there is no random assignment of the regional-scale phenomena, and their role in any given causal pathways can be understood only through inference from observational data, interpreted using knowledge of underlying causal mechanisms.

For example, the experiment by Miguel and Kremer, which studied the introduction of deworming medication in Kenyan schools, can be credited with sharp increases in funding for these interventions in recent years.[12,13] Their results were regional in nature, in the sense that much of deworming's impact occurred not only to the individual child who received medication, but through ecological effects to other children. Relatively few studies are capable of detecting such effects: for deworming, only five of the 42 trials used in the most recent "Cochrane Review" were clustered to detect such externalities,[14] and the updated series in the *Lancet* on maternal and child undernutrition dropped deworming as an evidence-based proven intervention for nutrition.[15] Results also vary by the degree to which children are exposed to helminthic infections in the first place. The Miguel and Kremer intervention was conducted in a region of high and rising infection rates, and none of the other studies were conducted across sufficiently diverse regions to identify interaction effects between treatment and exposure. Observational studies are needed to capture interaction with nonrandom prior conditions. For the effect of parasitic diseases on regional development, Bleakley[16] showed how differences in initial disease prevalence affected the impact of hookworm eradication in the United States, and McMillan *et al.*[17] documented how onchocerciasis control in Burkina Faso changed agricultural and other institutions in river valleys where the disease had been most widespread.

Cross-sector interactions can originate in any sector. For example, the top left corner of Table 1 shows regional interventions in agriculture, including the creation of better, more locally adapted crop and livestock techniques, market arrangements, and environmental innovations. Such interventions can make a region more prosperous and hence more able to afford and implement all of the other interventions. Similar region-specific programs exist in the nutrition and health sectors, as shown in the second and third rows of Table 1. Region-wide improvements affect the availability of any given product or service in a given locality, and also change individuals' access to that product or service. The middle two columns of Table 1 show how interventions that can be provided to individuals translate into outcomes that can be measured at the individual level, first in agriculture such as crop or livestock productivity or farm income, diets, and consumption of food and specific nutrients, then for nutritional outcomes including height and weight anthropometrics, and biomarkers in blood, urine, and tissue samples that are known to correlate with immunity to disease, cognition, and other aspects of maternal and child health.

The final column of Table 1 shows results defined at a regional scale, such as the wage rate and employment opportunities that are available to workers, food prices, and other market outcomes.

This kind of outcome arises from the interaction among people, and is shared among the people in a given region. For health, the comparable regional phenomena include disease transmission, which is a function of place, population density, and population (herd) immunity. The analysis of anthropometry, biomarkers, and disease prevalence rates can of course be analyzed at the regional level, for example, through documented cases in a given population, but the underlying data are individual in nature.

The typology presented in Table 1 offers a partial catalog of policy, program, and intervention potentials at various scales, linked to expected results that researchers might measure. Causal pathways between interventions and results could run in all directions. For example, agricultural interventions, which decrease food and water contamination (on-field aflatoxin control, penning livestock away from human water sources, controlling schistosomiasis in irrigation programs) can be posited to causally influence both nutrition and health. The typology reveals the inherent need for methodological pluralism when addressing increasingly integrated research questions: first to take account of linkages between agriculture, nutrition, and health within individual households, but especially to consider feedback effects at the ecological or regional scale that might drive population-wide impacts.

Impact pathways: metrics of integration between agriculture, nutrition, and health

Researchers can draw on the typology of Table 1 using a variety of logical frameworks regarding the possible pathways between interventions and outcomes. Studies addressing agriculture–nutrition–health linkages typically present their causal framework in a flow chart such as Figure 2, which is based on a widely used logical framework developed by the International Food Policy Research Institute (IFPRI).[18] In this diagram, causality flows from left to right, as assets and employment generate income that is spent on food and care practices that enter maternal and child nutrition and health. The possibility of reverse causality is shown in arrows from right to left around the outside of the flow chart. Such flow charts are composed of elements such as those in our Table 1, with each box representing something that is potentially measurable at one or another geographic scale, and the arrows representing directions of causality between the measurable elements.

The causal framework illustrated here augments the IFPRI framework with potential confounders for any observed relationship: for example, in recent years developing countries have experienced diverse and rapid changes in their labor–land ratios, age structure, and gender balance, posing various demographic challenges, even as climate change and disease exposure also vary widely. These trends and differences are not just random noise, but systematic confounders that threaten the external validity of any localized study, and may mask the effect of a multilocational program or trial. For any given set of interventions in the typology of Table 1, a causal framework, such as Figure 1, permits program design and evaluation to take account of potential synergies and interaction effects in ways that are potentially measurable.

Graphical frameworks such as Figure 1 can be expanded into infinitely complicated spaghetti diagrams, or simplified into highly reductionist linear models. Whatever their degree of granularity, however, they allow us to define and measure program integration in a novel way. Using Figure 2 or any other logical framework for program impact, we propose to define metrics of integration for operational research in two distinct dimensions: the breadth of integration across various arrows and boxes in the causal framework, and the depth to which those program elements are delivered together in any given intervention on the ground. The product of these two dimensions could then be used to measure the overall extent of integration.

Conceptually, a metric for the breadth of integration across causal pathways could start at zero (for a highly specialized, single intervention such as vitamin A supplementation that is not integrated with anything else) and go to infinity (for a universal intervention that does everything in the most detailed imaginable causal framework). In practice, a useful metric might use just three categories, with an intervention that spans one or two elements in a given framework being classified as having some integration (a score of 1), an intervention that spans three to five of the framework's elements being classified as having more integration (a score of 2), and one that spans six or more elements having an even wider breadth of integration (a score of 3). Such a measure could take values of 1, 2, or 3. What is

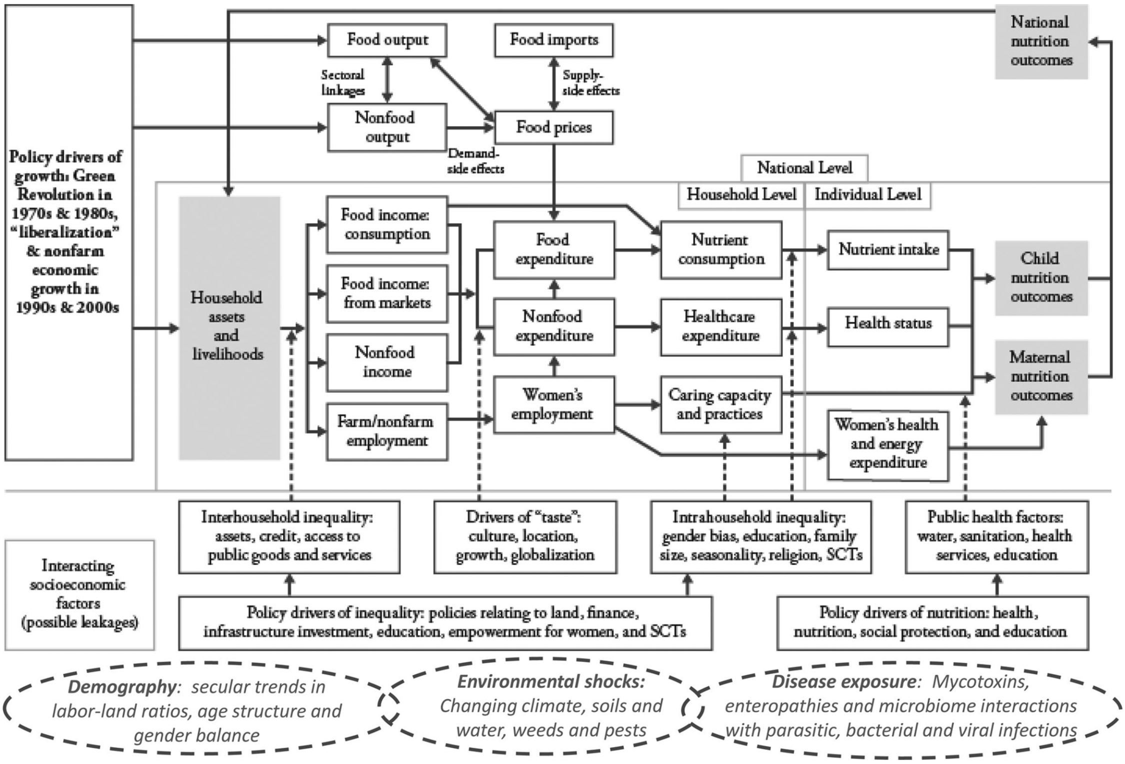

Figure 1. Example of a causal framework with confounders used for metrics of integration. Causal framework in square boxes and shadowed text is that of Ref. 18, augmented by the authors with social and ecological confounders in dashed ovals and italic text.

actually being more or less integrated would depend on the causal framework being used to identify pathways, but a larger number would always correspond to a greater breadth of integration.

Similarly, a metric for the depth of integration in program implementation could be scored using three categories. The simplest form of integrated programming is colocation, in which separate programs are placed adjacent to each other in a given village or region. A deeper form involves organizational partnerships and mergers, in which staff members from different organizations are asked to work together in each location. The most extreme form of integration is cross-training and resource sharing, in which individuals in a multifunctional organization are asked to provide multiple synergistic services. These could be scored as 1 (colocation), 2 (partnerships or mergers between organizations), or 3 (cross-training of multifunctional individuals). Again, what is being integrated would vary by program, but the metric is designed so that a larger number would always correspond to a greater depth of integration.

Clearly, a greater extent of integration in breadth or depth could improve a project's cost-effectiveness, if integration allows the organization to take advantage of synergies between program elements. But greater integration could also reduce cost-effectiveness if it prevents the organization and its staff from specializing in what they do best. In any given setting, there might be an inverted U-shaped relationship between a project's cost-effectiveness and its breadth or depth of integration, with the most effective level of integration being somewhere between pursuing each element in separate silo projects and pursuing every element simultaneously in undifferentiated all-purpose programs. The measures of integration that we propose offer the simplest possible scale with which to identify such a relationship, if it exists, with three levels of integration from some (1) to more (2) and most (3).

Table 2. **Breadth and depth of integration in a sample of 17 Non-governmental Organization (NGO) programs**

	Breadth of program (number of elements)		
	1–2	3–5	6+
Depth of integration between elements			
1. Colocation of program elements			2
2. Partnership among organizations	1	1	8
3. Cross-training and multifunctionality	1	1	3

SOURCE: Data shown are the number of programs in each breadth and depth category, from authors' classification of 17 integrated agriculture-nutrition projects implemented since 2008 by Catholic Relief Services (CRS), from internal CRS file data.

An opportunity to apply this metric of integration depth and breadth arose in 2013 for an internal review of agriculture–nutrition integration in projects implemented by Catholic Relief Services (CRS). That review covered all CRS English-language proposals involving agriculture and nutrition funded between 2008 and 2013 in which there was explicit use of the terms "integrated" or "integration" or other unambiguous linkage, such as interventions to control aflatoxin in food crops. The result was a sample of 17 projects implemented across Africa, Asia, and Latin America.

To measure the integration in this sample of projects, we began by constructing a causal framework similar to Figure 1, in which the cells were defined as all of the distinct program elements actually used by CRS in its work on agriculture and nutrition, and the arrows were causal pathways showing how those program elements affect development outcomes. Two consultants then scored the proposals in terms of breadth (number of different program elements in each proposal) and depth (whether two or more elements were proposed to be colocated, delivered by different organizations in partnership for the project, or delivered through cross-training of personnel). Each proposal was scored separately by the consultants, who generally reported identical scores but in a few cases identified ambiguities that required e-mail correspondence with CRS field staff to clarify the degree of integration linked to our framework.

Results of this pilot test are summarized in Table 2, showing that 13 of the 17 projects included six or more program elements, and that eight of them did so through partnerships among organizations (subgrantees). Five projects used cross-training and multifunctionality, mostly for projects with more than six program elements. From our initial observation, it appears that CRS's integrated agriculture–nutrition projects typically aim for relatively large numbers of program elements (more than six), with an intermediate depth of integration between them (using partnerships among specialized organizations, rather than colocation or cross-training). These classifications were based on textual analysis of project proposals, rather than field observations, and serve primarily to demonstrate the feasibility of measuring integration in this way. Next steps might include describing trends and patterns of integration across multiple agencies and then testing hypotheses such as whether the more integrated programs with more diverse elements are actually more cost-effective than more specialized programs and whether partnerships are actually more cost-effective than colocation or cross-training of multifunctional service delivery personnel.

The pilot data in Table 2 show that classifying projects using this metric is feasible and could usefully allow researchers to estimate how the degree of integration among interventions is associated with expected impacts. In so doing, a variety of research designs are likely to be useful, as detailed later.

Trade-offs in research: integration in delivery science, impact evaluation, and study design

To identify causal pathways and estimate effect sizes for more or less integrated programs and policies, researchers can choose among a complex set of methodologies. Choices among methods always involve trade-offs, as illustrated in Figure 2. This diagram illustrates a sequence of three-dimensional

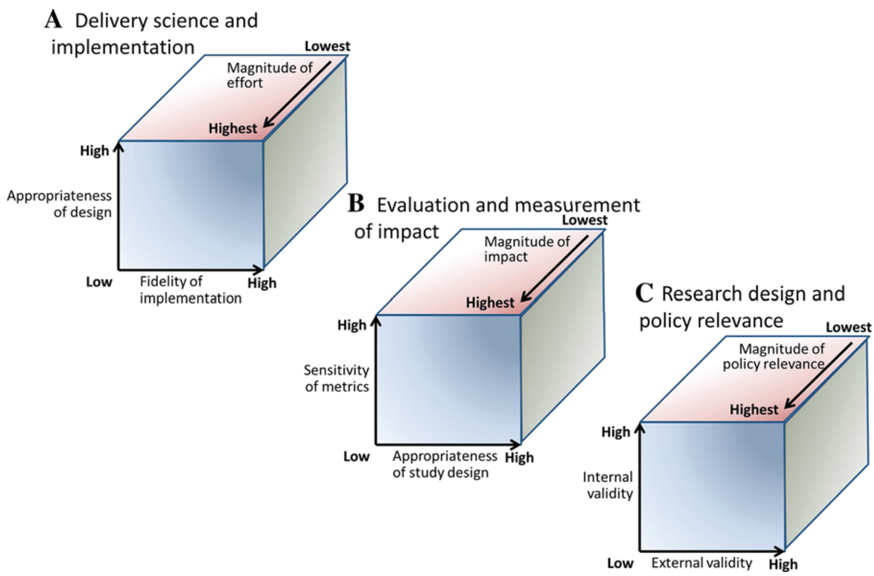

Figure 2. The many dimensions of integrated program delivery, impact evaluation, and research relevance.

spaces in which any intervention and research project might fall. In each three-dimensional space, the more desirable direction is upward along the vertical axis, to the right along the horizontal axis, and diagonally along the third axis showing depth, so that the most desirable situations are at the peak point shown at the center of each cube. The three cubes illustrate dimensions that characterize different aspects of integration in research: first, delivery science (how programs are implemented); second, evaluation and measurement of impact (whether the program met its objectives); and finally, the relevance for future policies of a given research design (whether the research itself is influential). Success in each of these can lead to a virtuous circle of improvement. Conversely, weakness in any dimension will limit what can be learned.

Delivery science and program implementation

The first set of trade-offs in Figure 2 illustrates the delivery science aspect of integrated programs. Delivery science focuses on overcoming implementation constraints and promoting cost-effectiveness at scale. This emerging field addresses the "how" as well as the "what" and "where" of a given intervention. Figure 2 shows the degree to which program delivery moves from a total lack of effectiveness along each axis toward the highest possible degree of success.

In Figure 2A, the vertical axis illustrates the appropriateness of an intervention's design. This might be measured as the degree to which iron folate supplements are effectively targeted to individuals at risk of anemia, for example, by taking account of whether malaria prevention and control measures are in place where malaria risk is endemic. In this case, appropriateness relies on researchers' *a priori* knowledge that iron folate supplementation might be desirable only under certain conditions.[19] Similarly, appropriate design may focus on promoting exclusive breastfeeding and birth spacing in populations where age of first pregnancy is low and fertility rates are high. Resolving anemia and reducing maternal pregnancy risks will have effects on nutrition of the mother, but also on the next generation of children to be born, which has long-term relevance to productivity (including in agriculture).

The horizontal axis of Figure 2A illustrates fidelity of implementation, which might be measured as the degree to which an intervention actually reaches the intended beneficiaries as designed, which might depend, for example, on how program staff are compensated and the incentives they have to screen beneficiaries according to defined protocols. Those two dimensions are the typical focus of research on intervention effectiveness, typically in controlled trials where design and delivery are closely monitored for a specific sample of individuals.

The third dimension, along the depth axis, shows the program's magnitude of funding and effort relative to the geographic region where results are expected. This dimension is especially important because of the nonlinearities and threshold effects that characterize bioecological systems, for example, in disease-control vaccination programs that require a high level of coverage to achieve herd immunity.

Taken together, these three dimensions of delivery science are critical to any understanding of how to maximize the benefits of complex programming. Design and implementation strategies are of course important, but analysis of fidelity, coverage, and intensity of application of programming is essential to the identification of pathways toward impact.

Program evaluation and measurement of impact

Given the quality of program implementation along the three dimensions of Figure 2A, the evaluation methods that a researcher might use can be described along the three dimensions of Figure 2B. Here, the vertical axis shows the sensitivity of metrics relative to expected results. For example, measurement of wasting prevalence, defined as the fraction of children below a given level of weight-for-height, would miss relevant changes in weight-for-height among children at other levels of body weight that might matter for health.[20] Measures that capture both under- and overnutrition may be particularly valuable given the increasing prevalence of both forms of malnutrition. Similarly, growing more food can result in less impact on nutrition where that food is contaminated with mycotoxins that potentially impair immunity of the consumer to diseases. The horizontal axis shows the appropriateness of study design, with respect to whether observed changes were actually caused by the intervention or might be due to other factors. The value of data depends not only on measurement accuracy, but also on how treatments were randomized and other aspects of data quality.

In Figure 2B, the third dimension on the depth axis shows the intervention's magnitude of impact. That dimension is closely related to the magnitude of effort in Figure 2A, but it also depends on the program's design and fidelity of implementation. The net result is crucial for program evaluation because the results of low-magnitude programs are difficult to measure with statistical significance. An intervention's magnitude is defined relative to the administrative or ecological region of interest, which is subject to changes other than the program intervention. Distinguishing signal from noise depends in part on sample size. In clinical settings, the expected magnitude of impact relative to variance associated with other factors informs study design and sample size through power calculations, but in field research the sample size is often dictated by the intervention itself, for example, through the number of distinct operations that can be randomized.

Research design and policy relevance

Figure 2C illustrates the quality of study design, in terms of the validity and relevance of its research findings. Here, the vertical axis shows the internal validity of causal attribution within the observed population, while the horizontal axis shows its external validity for other people to whom the intervention might apply elsewhere. Internal validity is typically highest in double-blind, placebo-controlled randomized trials, but the degree of internal validity depends on how the specific protocol is implemented. For example, if behavioral responses are involved, to avoid the Hawthorne effect it may be preferable to conduct trials in ways such that subjects do not even know they are in a trial.

The horizontal axis of Figure 2C represents external validity, namely whether the effect observed in a trial can be generalized to other populations. Research with high internal validity might not be generalizable, perhaps because the intervention would be implemented differently, or because subjects are not representative of the general population. Returning to the previous example of iron folate supplementation to combat anemia, a clinical study of its effectiveness among children in a malaria-free area may produce a highly valid attribution of causality for that population, even though that study may have very little external validity for anemic children in other locations. Evaluations of integrated agriculture–nutrition–health studies are particularly vulnerable to low external validity due to the location-specific, agroecological constraints they address. This puts a premium on drawing appropriately sized representative samples for data collection, which in turn makes it difficult to implement the randomization strategies needed for internal validity.

In Figure 2C, the third dimension of depth shows the study's policy relevance. This attribute of study design is different from either internal or external validity. The relevance of a study depends on whether the metrics and methods are salient to a decision maker's needs. Relevance may therefore involve the timing of a study, for example, to inform decisions during the period when decision makers are responsive to research results. Relevance may also involve metrics, for example, to inform decisions that are bound by law or other constraints to target a specific kind of change such as the Millenium Development Goals (MDGs). However, relevance can also color other aspects of study design. Studies may have high validity but low relevance to policy decisions. Researchers often face tradeoffs between evaluations tailored to specific decision makers or to address policy questions and those designed to produce results of the highest internal or external validity.

Taken together, the three panels of Figure 2 describe how integrated research can drive better project implementation and the best possible outcomes in Figure 2A, the highest measured results in Figure 2B, and the greatest policy relevance in Figure 2. Not all research methods will succeed in all dimensions, but research projects that take all of these dimensions into account are more likely to be successful.

Currently, such research activities are all too rare. A gap analysis undertaken by Hawkes *et al.* reviewed ongoing and planned research projects around the world that focused explicitly on understanding the linkages between agriculture and nutrition.[21] The authors identified 151 research undertakings, more than half of which targeted Sub-Saharan Africa. Of the 151 studies characterized, 92 reported a particular focus on nutritional impacts for women and children. Despite this explicit focus on (1) interactions among sectors and (2) nutrition outcomes for key target groups, the vast majority of the research so identified approached their subject matter through one of the three cubes presented in Figure 2, while none at all attempted to link data gathering and analysis across all three research domains. As Hawkes *et al.* stated, "no single project completes the research chain fully." While some studies come close, the main links missing in the research chains identified by that gap analysis included paying attention to value chains (beyond production of commodities and their consumption), and to local and regional food environments (determined by market access, retail price dynamics, and local integration with national and international systems), as well as appropriate measures of actual nutrition impacts using valid methodologies. Such gaps are seen as preventing a more complete understanding of the full pathways of change, and as such, a better understanding that "the most appropriate methods and metrics is [sic] needed to be able to translate research into practice, and plan and design future research."

Conclusions: toward more effective integrated research methods

Policies and programs designed to overcome location-specific biological and ecological constraints in agriculture, nutrition, and health have distinctive features that call for a new language of intervention design, delivery, and evaluation. It has been argued that such language is essential to progress in attaining global targets for health and nutrition.[c] This paper offers a typology of interventions and outcomes, with corresponding metrics of integration among them, to help guide evaluation approaches that can help researchers capture key features of program integration in ways that better inform policymaking and program design. These features cannot all be tested empirically, especially not in all combinations, but this typology allows us to think about integration in a way that could be tested in practice.

A first dimension of special concern for integrated agriculture–nutrition–health programs is their geographic scale. For ecological as well as administrative reasons, programs and evaluations are often scale-dependent, with interventions and expected results that are defined over a geographic region rather than an individual person or household. The typology proposed here therefore begins with a catalog of the interventions and results that are defined

[c] Shekar *et al.* state that the progress toward improved integrated programming aimed at enhancing nutrition remains constrained by the "absence of explicit training in these competencies in leading universities, absence of research on the delivery sciences including operational research, lack of or poor-quality assessments, scarce funding for delivery research, and the reluctance of journals to publish it."[22]

at the geographic as opposed to the individual level, and then provides three-dimensional criteria along which a given program intervention and evaluation method could be situated. These dimensions identify the relevance of delivery science for characterizing how an intervention is implemented, beyond the characteristics of what, where, and for whom the program operates. The typology then identifies specific needs of integrated programs in terms of evaluation methods, and the specific issues in study validity and policy relevance that arise for these programs. Metrics of integration, in turn, capture both the breadth of integration across diverse program elements, and the depth with which those elements are integrated in a given program on the ground. The value of such metrics can be tested retrospectively and prospectively in relation to selected groups of integrated interventions at various scales of operations.

The typology, metrics, and trade-offs described here can help researchers design more effective studies by matching the most appropriate methods to specific interventions and effects. No single research method can answer all questions. Each study adds to previous work by extending it in particular directions, looking under the lamppost illuminated by a particular approach and set of circumstances. The challenge posed by the need for greater integration in research methods will not be met by a new orthodoxy in research methods, but by methodological pluralism. In so doing, the challenge of integration is not only to match problems with appropriate methods, but more importantly to bring a wider range of previous research using other methods to bear on each new study. For example, credible observational studies of regional differences must take account of what is known about individual-level responses from experimental studies, and vice versa: a controlled experiment must attend to both external and internal validity, by appropriate choice of location and other circumstances. Researchers will also need increasing sophistication about the trade-offs illustrated in Figure 2, seeking the highest possible impact in terms of program delivery, evaluation design, and policy relevance.

As the pursuit of global development targets under the MDGs reaches the 2015 deadline, new quantified, time-bound goals for international development are needed. These will be defined in terms of metrics that have major implications for program design and evaluation. Researchers will need to design studies that capture the impact of cross-sectoral programs on the new targets. These are likely to go beyond the siloed, sector-specific goals of the past and offer greater understanding of how people interact with their resources in specific locations. The concepts presented here are designed to be of broad applicability, permitting program designers and researchers to identify specific characteristics likely to produce the highest possible measurable impacts from the interventions with which they work.

Acknowledgments

The authors thank the United States Agency for International Development for making this paper possible, through the USAID Feed the Future Innovation Labs for Collaborative Research on Nutrition in Africa and Asia. We are also grateful to Valerie Rhoe at Catholic Relief Services for permission to use internal program review results conducted under her leadership by Mackenzie Sehlke, Nadira Saleh, and William Masters, using the methodology described in this paper.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Winikoff, B. 1978. Nutrition, population, and health: some implications for policy. *Science* **200:** 895–902.
2. Ruttan, V. 1975. Integrated rural development programs: a skeptical perspective. *Int. Dev. Rev.* **17:** 9–16.
3. Meadows, D.H., D.L. Meadows, J. Randers & W.W. Behrens. 1972. *The Limits to Growth*. New York: Universe Books.
4. Lam, D. 2011. How the world survived the population bomb: lessons from 50 years of extraordinary demographic history. *Demography* **48:** 1231–1262.
5. Government of Haiti. 2012. *Politique nationale de nutrition*. Port au Prince, Haiti: Ministère de la santé publique et de la population.
6. Government of Nepal. 2012. *Multi-Sectoral Nutrition Plan for Accelerating the Reduction of Maternal and Child Under-Nutrition in Nepal*. Kathmandu, Nepal: National Planning Commission.
7. Government of Timor-Leste. 2012. *National Nutrition Strategy 2012–2017*. Dili, Timor-Leste: Ministry of Health. Draft as of September 11, 2012.
8. G8 (Group of Eight Forum Countries). 2009. "L'Aquila Joint Statement on Global Food Security. Accessed January 5, 2014. http://feedthefuture.gov/sites/default/files/resource/files/afsi_jointstatement_2009.pdf.
9. Pronyk, P. 2012. Errors in a paper on the Millennium Villages project—Editors' reply. *Lancet* **379:** 1946–1947.

10. Banerjee, A. & E. Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York, NY: Public Affairs.
11. Deaton, A. 2010. Instruments, randomization, and learning about development. *J. Econ. Lit.* **48:** 424–455.
12. Miguel, E. & M. Kremer. 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* **72:** 159–217.
13. Givewell. 2013. Givewell Metrics report—2012 Annual Review. Cited March 12, 2013. http://www.givewell.org/files/images/GiveWell%20Annual%20Review%202012%20-%20Metrics%20Report.pdf.
14. Taylor-Robinson, D.C., N. Maayan, K. Soares-Weiser, *et al.* 2012. Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin and school performance. *Cochrane Database Syst. Rev.* **11**, Art. No. CD000371.
15. Bhutta, Z., J. Das, A. Rizvi, *et al.*; for Lancet Maternal and Child Nutrition & Interventions Review Groups. 2013. Evidence based interventions for improving maternal and child nutrition: what can be done and at what cost? *Lancet.* **382:** 452–477.
16. Bleakley, H. 2007. Disease and development: evidence from hookworm eradication in the American South. *Q. J. Econ.* **122:** 73–117.
17. McMillan, M.S., W.A. Masters & H. Kazianga. 2013. *Disease Control, Demographic Change and Institutional Development in Africa.* National Bureau of Economic Research Working Paper No. 19245. Cambridge, MA: NBER (http://www.nber.org/papers/w19245).
18. Gillespie, S., J. Harris & S. Kadiyala. 2012. *The Agriculture-Nutrition Disconnect in India: What Do We Know?* FPRI Discussion Paper 01187. Washington DC: IFPRI.
19. WHO. 2012. *Guideline: Daily Iron and Folic Acid Supplementation in Pregnant Women.* Geneva: World Health Organization.
20. Bhagowalia, P., S.E. Chen & W.A. Masters. 2011. Effects and determinants of mild underweight among preschool children across countries and over time. *Econ. Hum. Biol.* **9:** 66–77.
21. Hawkes, C., R. Turner, J. Waage, *et al.* 2012. Current and planned research on agriculture for improved nutrition: a mapping and a gap analysis. A report for DFID. August 21, 2012.
22. Shekar, M. et al. 2008. Delivery sciences in nutrition. *Lancet* **371:** 1751.