From Biological to Program Efficacy: Promoting Dialogue among the Research, Policy, and Program Communities[1,2]

Jean-Pierre Habicht* and Gretel H. Pelto
Division of Nutritional Sciences, Cornell University, Ithaca, NY

ABSTRACT

The biological efficacy of nutritional supplements to complement usual diets in poor populations is well established. This knowledge rests on decades of methodologic research development and, more recently, on codification of methods to compile and interpret results across studies. The challenge now is to develop implementation (delivery) science knowledge and achieve a similar consensus on efficacy criteria for the delivery of these nutrients by public health and other organizations. This requires analysis of the major policy instruments for delivery and well-designed program delivery studies that examine the flow of a nutrient through a program impact pathway. This article discusses the differences between biological and program efficacy, and why elucidating the fidelity of delivery along the program impact pathways is essential for implementing a program efficacy trial and for assessing its internal and external validity. Research on program efficacy is expanding, but there is a lack of adequate frameworks to facilitate the process of harmonizing concepts and vocabulary, which is essential for communication among scientists, policy planners, and program implementers. There is an urgent need to elaborate these frameworks at national and program levels not only for program efficacy studies but also for the broader research agenda to support and improve the science of delivering adequate nutrition to those who need it most. *Adv. Nutr. 5: 27–34, 2014.*

Introduction

A successful dialogue to translate nutritional science knowledge into improving nutrition in populations is based on an initial agreement that interventions to improve nutrition in populations show evidence of biological efficacy. Without a strong evidence-based foundation, we cannot expect interventions, no matter how well organized and conducted, to achieve the goals of improved nutrition. Decades of research and dialogue about codifying the evidence for the biological efficacy of a number of nutrients have led to improving child growth and preventing illness and death. This has led to definitive evidence of efficacy for vitamin A, iron/folate, and mixtures containing these nutrients (1). Moreover, the evidence on efficacy in specific segments of the population (e.g., pregnant women) has been established. These very significant accomplishments were achieved under the leadership of WHO, in collaboration with many other organizations and individuals. This demonstrates how effectively the international nutrition community can pull together when institutions provide leadership to fulfill internationally agreed-upon mandates.

It has taken decades of investments and dedicated research to produce this biological foundation. Similar investments must now be made to acquire evidence on how to deliver these nutrients, as well as to explicate the roles of other factors that determine effectiveness. That investment will need to encompass many different issues and will require inputs from different types of scientists, policy makers, and program implementers and from the communities in which interventions are seated. Furthermore, it will also require a dialogue to establish a consensus about basic constructs and terms.

A primary purpose of this article is to describe some basic constructs related to the delivery of efficacious interventions, to the evidence that giving a nutrient can improve

[1] Presented at the symposium "The WHO Evidence-Informed Guideline Development Process: Implications for Vitamin and Mineral Research Priorities" held 20 April 2013 at the ASN Scientific Sessions and Annual Meeting at Experimental Biology 2013 in Boston, MA. The symposium was sponsored by the American Society for Nutrition (ASN) and supported in part by the World Health Organization. A summary of the symposium was published in the September 2013 issue of *Advances in Nutrition*.

[2] Author disclosures: J.-P. Habicht and G. H. Pelto, no conflicts of interest.

* To whom correspondence should be addressed. E-mail: jh48@cornell.edu.

nutrition, and to how one might deliver that nutrient to a population.

A second purpose is to propose some steps to enhance the mutual understanding among interlocutors concerned with improving public nutrition (2). As implementation science develops in nutrition, it will be important to establish venues and opportunities for cross-disciplinary and cross-community communication.

Biological Efficacy: Dialogue Is Essential for Demonstrating Efficacious Biological Interventions in Free-Living Populations Outside of the Laboratory and the Clinic

This section is a brief review of the first step of the process to demonstrate biological efficacy in a community setting. Efficacy is "the extent to which a specific intervention … produces a beneficial result under ideal conditions." "Effectiveness is a measure of the extent to which a specific intervention …when deployed in the field, does what it is intended to do for a defined population" (3).

The difference between efficacy and effectiveness is clear for therapeutic drugs used in medical practice. The "ideal conditions" for drugs are controlled trials in clinical settings; effectiveness is the impact in usual medical practice. In contrast to testing drugs, the "ideal condition" to ascertain public health efficacy is a field trial where the deployment is as perfect as can be devised, especially for the delivery of the intervention. For example, in a vitamin A efficacy study using capsules, a research staff member directly provides the capsule to the child or observes its ingestion. "Effectiveness" relates to the outcomes in a usual public health program. For example, in a vitamin A effectiveness study in which capsules are delivered by community health workers (CHWs)[3], the ingestion is not observed.

Laboratory and clinical research should precede field efficacy trials to establish a high likelihood that the intervention is likely to lead to the desired outcome (3). Progress in the demonstration of biological efficacy in the community has depended on improvements in scientific techniques of sampling and data collection. It also has depended on the development of agreements about the interpretation of data, with a common terminology, that permitted the codification of evidence to "prove" efficacy. This codification required an enormous investment of time and resources over more than half a century and continues today. The process of this codification has been presented by Peña-Rosas et al. (4) and was updated by Tovey (5) in this issue. The codified methodology establishes proof that nutrient molecules prevent impaired growth, health, and survival due to malnutrition.

An essential step in the codification for biological efficacy is the demonstration through a randomized controlled trial (RCT) of an effect on a nutritional outcome when a nutrient or combination of nutrients is directly administered to individuals who are either already malnourished or who are very likely to become malnourished. The interpretation of positive effects of RCTs is generally straightforward when the results show statistically significant effects. When they are replicated in further studies with the same results, confidence grows that biological efficacy is established. However, there is less codification about how to respond to negative studies that show no effect. The problem with negative studies is that the lack of significant results can be due to a number of different causes. One cause is that the recipients could not benefit from the nutrient because they were not deficient in that nutrient. Another cause is that the dose was too small to fulfill the need (6). Other causes include that the response was prevented because another essential nutrient was also deficient, the participants did not receive the nutrient due to a lack of fidelity in implementing the intervention, or because the pill was not ingested—or simply to statistical bad luck.

To demonstrate biological efficacy, 2 causes of lack of impact must be prevented by the research design and its implementation. The first is lack of potential to benefit from the doses delivered; the second is not delivering the doses. Unless these causes are rejected, one can infer nothing about biological efficacy from the trial. As an example, all of the biological efficacy trials included in meta-analyses for biological efficacy of vitamin A to prevent deaths were undertaken in populations with demonstrable vitamin A deficiency, and the studies ensured the dose ingestion by delivering the verified doses to the child by research staff. A lack of impact issuing from such a design should trigger in-depth investigations into reasons for a lack of benefit, and into the delivery procedures and their verifications (1). It is a tribute to the quality of past researchers, whose results were included in previous meta-analyses, that the meta-analytic results are so clear. Another interpretation of the uniformity of benefit is that the potential to benefit was high in these populations, and one would therefore expect the same benefits in other populations with similar evidence of vitamin A deficiency. This evidence of external validity is important both in determining generalizability and in setting the conditions for achieving impact elsewhere.

Dialogue about Delivering Nutrition

Proof of biological evidence is of little import if the efficacious agent cannot be delivered to those who can benefit from it. However, the dialogue about how to use such evidence is much less developed than it is for nutrient efficacy. There is not even an agreed-upon term that refers to "delivering nutrition to populations." This lack of agreement is reflected by the fact that we do not have agreement on whether there is a difference between "public nutrition" (2) and "public health nutrition." To make progress in this arena we need to define terms and codify their meanings. When a term refers to an outcome or a process we need to codify the terminology for how the process is described and how the outcome is achieved.

[3] Abbreviations used: CHW, community health worker; DEVTA, Deworming and Enhanced Vitamin A; GRADE, Grades of Recommendation, Assessment, Development, and Evaluation; PIP, program impact pathway; RCT, randomized controlled trial.

Although we lack adequate codification for dialogue, we do have a common understanding about certain principles. We agree that we should give priority to delivering substances that have proven biological efficacy. We have less consensus about the role of other factors, which must also be in place for nutrients to be ingested. These factors can be listed under the general headings of "social and economic resources" and "knowledge and motivation." In principle, we also agree that we should use nutrition delivery systems that have been shown to "work." However, we do not agree about what constitutes "work."

One way of determining that a nutrition delivery system "works" is that the steps to accomplish the delivery have occurred. Selecting indicators for some of these steps is often used in project management and monitoring. Nutrition interventions can be conceptualized as a set of flows. One flow delivers nutrients from a source to a beneficiary. Other flows deliver other kinds of resources and information to those involved in the nutrient flow.

The delivery flow of nutrients is the easiest to conceptualize. However, it should be noted that one does not actually track the nutrient itself, but only the container, which covers the nutrient. Examples of nutrient containers are food fortificant mixes, foods, pills, or other kinds of supplements. The nutrient container may also change along the flow. For example, fortificants are mixed into a food, and that food is then mixed with other foods. Following the sequence of nutrient containers is not always easy, particularly when fortified foods are delivered through the free market. For others, such as vitamin A capsules, it is easier to follow the container because the nutrient enters and exits the program in the same container. Even in this case following the container is less straightforward than one might think. The container is rarely transferred directly to the beneficiary from its entry into the program. It usually passes from 1 delivery subsystem to another (**Table 1**).

If the handovers between subsystems, and the movement between handovers, are efficient, the whole system is effective in delivering the containers. We find it useful to differentiate movement between handovers from the handover activities themselves, because the nutrient flow is often imperiled at the handovers. Some aspects of container movement within a subsystem, such as the movement of trucks, are usually well monitored. This is well understood for food relief logistics. However, it is at the handover steps after the truck arrives that the system often falters.

Using the container as a proxy to follow a nutrient flow is dangerous unless the container retains the nutrient. In our experience, some genetic biofortification programs do not adequately evaluate and track how the nutrient is affected by shelf life. On the other hand, well-run salt-fortification programs routinely check the salt for its fortificant (e.g., iodine) content.

TABLE 1 Activities that flow a vitamin A supplement from 1 subsystem to another in a community

1. Program receives the capsule to pass on to the CHW[1]
2. Program delivers the capsule to CHW
3. CHW gives the capsule to the mother
4. The mother or CHW puts capsule directly into mouth
5. Vitamin A is ingested, absorbed, and metabolized
6. Vitamin A produces biological outcomes (e.g., no death)

[1] CHW, community health worker.

Other resource flows that are part of the process of intervention delivery have similar characteristics to nutrient flows. Some aspects have concrete characteristics, such as money or firewood, which can be measured and monitored. Others, particularly flows of knowledge and motivation, are frequently less systematically monitored. They are part of the "black box" that is often referred to with respect to behavior change interventions.

Knowledge flows are certainly measurable, but they are more difficult to operationalize than tracking the flow of the nutrient container. For instance, the handover from the person who transfers the knowledge to the recipient depends on the amount of knowledge of the knowledge transferor and his/her efficacy of transfer to the recipient. Moreover, the efficacy of transfer depends not only on the transferor but also on the recipients' capacity to receive knowledge, as well as other factors that impinge on knowledge delivery and reception (7). Research on this handover is so poorly developed that, at present, it requires a research context to examine it.

In fact, research into the difference between delivery and acceptance for all handovers in the delivery of nutrition is so poorly developed that it is not clear what acceptance means nor what term to use for this concept.

Program Efficacy Studies

The need for systematic attention to the development of sustainable programs to deliver nutrition interventions is now widely recognized. In addition to sustainability, these programs have to be capable of wide coverage that will reach those in need. Among the factors that impede progress is the lack of an established framework for assessing the efficacy of delivery systems comparable to the one developed by WHO for establishing the biological efficacy of nutrients. There are many frameworks in the literature, but none are systematized by a respected body, such as WHO, which has the mandate and the authority to do this internationally.

In the absence of such a framework, a number of different approaches have been used, some of which are directed to the examination of program efficacy whereas others focus on program effectiveness. The success of the RCT model for establishing biological efficacy has led to its application as an approach to assess program efficacy (8). Some of these program efficacy RCTs (9–11), especially when they are augmented by other information (12–14), have provided useful information, not only for program development but also for policy. Our experience in undertaking program efficacy trials (10,11) is that certain prerequisites are necessary. These are shown in **Table 2**.

The reason for the first prerequisite is that judgments about efficacy depend entirely on the biological outcome to ascertain whether or not the agent was actually delivered

TABLE 2 Prerequisites for conducting a program efficacy trial

1. Proven efficacy of the agent that is to be delivered
2. Proven "potential to benefit" on the part of the recipients in both intervention and control groups
3. The benefits most likely outweigh the risk of harm
4. Very high likelihood that the delivery system can deliver the intervention under the conditions of the study
5. Rigorous methods to identify and correct program inefficiencies over the course of the study

to the intended beneficiary. Only if the agent's biological efficacy is already ascertained can one know that a lack of program impact is due to failures in the delivery system and/or the biological status of the recipients with respect to their potential to benefit. Without that certainty, one does not know if improving or changing the delivery system is worthwhile. In other words, a program efficacy trial should never be used to assess biological efficacy. The second prerequisite is a prerequisite for both biological efficacy trials and program efficacy trials. The third prerequisite relates to being sure that the program is worth studying. This prerequisite is analogous to the prerequisite of biological potential in biological efficacy trials. Why conduct an RCT of a delivery system that does not work? Obtaining evidence that a program is working well requires qualitative research to be sure that it is worth studying. The fourth prerequisite is also essential. We have found that no matter how good our training and supportive supervision is, failures in the flow occur [e.g., (12)]. Thus, information about program inefficiencies and the steps taken to correct them is essential for interpreting "negative" results and for generalizing the findings beyond the study site.

At present, program efficacy studies use the same criteria that are used for biological efficacy studies. The differences between different types of efficacy studies were described in 2004 (8), but the next steps, laying out the implications for how to evaluate them, have not been undertaken in the intervening decade. This absence leads to confusion and contradictions in interpretation. For example, many RCTs are purported to be successful because the results are significant ($P < 0.05$), but the reports lack the complementary information (15) that is essential for program development. Adherence to the criteria listed in Table 2 for program efficacy studies is even more important when the results of the studies are negative.

Program Impact Pathways

To facilitate communication about the analysis of program efficacy and intervention management, a construct we find to be particularly helpful is the idea of "program impact pathways" (PIPs). A key feature of this construct is the concept of flow, referred to above. This idea of "flow" is implicit in discussions about "value chains" (16) of products as they progress from original producer to consumer. It is more explicit in causal diagrams (17) and in the program theory of interventions (18). For nutrition, we suggest that the construct of a PIP be used to refer to a specific kind of flow: the flow from a nutrient's introduction into a program to its biological outcome. Table 1 depicts an abbreviated PIP for a vitamin A supplementation program that uses CHWs as the platform to deliver the capsule. This is the simplest of examples. The container, the capsule containing the vitamin A, is followed as it is handed over at crucial steps until its contents are released into the mouth of the beneficiary and is then transformed into performance, health, and survival outcomes. A complete PIP would also identify inefficiencies in the flows of the nutrient. It identifies other flows at each step that increase or decrease efficiencies.

A PIP is a prerequisite for designing an intervention because it allows the program planners to identify the resources and behaviors essential for each step. It is also essential for the management of a program because it identifies the steps where efficiency must be improved. It is necessary for evaluation because it increases the plausibility of the findings (8). Finally, it is essential for predicting generalizability about implementing a program elsewhere because it helps to identify and predict the inefficiencies that are likely to arise in the new environment, and how to prevent these failures (19). For these reasons we suggest that an explicit PIP should always be included in program efficacy trials because it provides a basis for generalizability of the results and thus for the trial's external validity.

A PIP provides the means not only for identifying inefficiencies in the flow of delivery, it is also useful for specifying the points in the delivery process where a "handover action" takes place. For example, in biological efficacy trials, the ultimate handover of the intervention is at the end of the flow when the vitamin A is fed to the child. In a program efficacy trial, there are more intervention handover locations, beginning with the point at which the funding to acquire vitamin A is allocated to a program level. The following section demonstrates why establishing and following the PIP is necessary not just for external validity but also for internal validity.

The DEVTA trial: an example of confusing biological efficacy research with program efficacy. The Deworming and Enhanced Vitamin A (DEVTA) trial was designed to deliver a vitamin A supplement to 1 million preschool children in north India every 6 mo through the governmental primary health care system (20). The capsules were delivered to the program. The CHWs were trained in capsule delivery and relevant reporting. According to the investigators, the monitoring and evaluation system they set up showed good compliance in the delivery of the capsules, and in the measurement of biological outcomes through sampling mothers and their children.

The study found no difference in mortality between the supplementation group and the control group. The researchers concluded that vitamin A was not biologically efficacious to reduce mortality. The investigators came to this conclusion because the only step in the PIP pathway they considered was the final outcome because they believed they had examined and ruled out earlier failures in the

flow. Is this the only possible explanation, or is it even plausible? A review of how well the DEVTA study met the prerequisites in Table 2 provides some insights to answer this question.

The study protocol presented the following 2 objectives relative to vitamin A: *1*) to determine how reliably, efficiently, and sustainably the Anganwadi system can deliver safe, simple health services (in this case, anthelminthics and micronutrient supplements) and *2*) to determine whether retinol supplementation at 6-mo intervals can somewhat improve child survival during ages 1–6 y (21).

The first prerequisite in Table 2 is "proven efficacy of the agent that is to be delivered." The investigators' first objective is a program efficacy study. Its goal is to test the quality of a specific delivery system as a vehicle for a specific nutrition intervention. This is correct, because vitamin A supplementation already has proven biological efficacy. However, the second objective implicitly calls into question whether vitamin A is biologically efficacious in this environment. If there was doubt about this efficacy, the study of the delivery system should not have been undertaken. A biological efficacy trial would be the only appropriate study to answer this question because it is the only design in which the researcher controls the delivery to the child. Calling into question the efficacy of vitamin A, as is done by the second objective, destroys the possibility of using biological impact to present convincing evidence for the first objective that the delivery system is efficacious.

The second prerequisite in Table 2 ("proven potential to benefit on the part of the recipients in both intervention and control groups") was plausibly met by the DEVTA study. The investigators showed that vitamin A status improved among the children in a small opportunistic sample who had received the capsules.

The investigators do not provide evidence related to the third prerequisite, "very high likelihood that the delivery system can deliver the intervention under the conditions of the study." Apparently the DEVTA trial introduced recording procedures and the appropriate training but did not make any other changes to the program delivery system. Nor does it appear that formative research was conducted to examine potential problems in the use of delivery system. They do not report data on CHW motivations and practices or how these might relate to the delivery and reporting of capsule ingestion. In fact, there are no reports about any blemishes either in the program or about how delivery failures were identified and addressed. Thus, it is a "black box" efficacy study with a negative outcome. Those who have experience with large-scale programs of the type used in this study should be very skeptical that the program and reporting worked as the authors claimed.

Appropriate skepticism is reinforced by the absence of attention to the fourth prerequisite in Table 2, "rigorous methods to identify program inefficiencies." Delivery records and biochemical response were verified only in a relatively small, nonrandom, opportunistic subsample of 2106 of the 1,000,000 children in the study. There is no information about how this small subsample was selected, but it is likely that, as in all opportunistic selection, the children who were included were likely to be more easily reached and therefore received both the supplement and the validation visits. There is no evidence that other, less easily reached children received the supplement. There is thus no evidence that the intervention was or was not given to those who could most benefit. The impression of inadequate identification of program deficiencies is reinforced by the absence of reports of any failures in the delivery system. One concludes that the delivery system was unlikely to have had enough fidelity to the study design to be effective.

The authors' interpret the findings as follows:

"Interpretation: DEVTA contradicts the expectation from other trials that vitamin A supplementation would reduce child mortality by 20–30%, but cannot rule out some more modest effect. Meta-analysis of DEVTA plus eight previous randomized trials of supplementation (in various different populations) yielded a weighted average mortality reduction of 11% (95% CI: 5%, 16%; $P = 0.00015$), reliably contradicting the hypothesis of no effect" (20).

The first sentence is a statement of fact, and the authors could have claimed that this was an effectiveness trial that showed no impact under these "real life" circumstances. However, the second sentence of the "interpretation" compares the DEVTA program efficacy trial to previous biological efficacy trials, which means that the authors think that the DEVTA trial was an efficacy trial. One would have expected that the fact of no effect would have provoked some thought about program fidelity, and some speculation as to why the delivery system failed. Speculation is all that could have been done because the study was not designed and conducted as a program efficacy trial or even as a program effectiveness trial, and therefore it did not collect the PIP evidence necessary to support any inferences about where the inefficiencies occurred.

Because the authors of the DEVTA study did not differentiate between biological and program efficacy, they concluded that their study should be included in a meta-analysis with previous biological efficacy results. However, this combining of biological and program efficacy trials is scientifically incorrect. The 2 types of research are conducted differently and produce different kinds of evidence. They have completely different purposes. One tests biology; the other tests the delivery system. Thus, we conclude that the DEVTA study adds nothing to our knowledge about the biological efficacy of vitamin A. It adds nothing methodologically to the scientific literature, except as a cautionary tale. The study has sown confusion by inappropriately undercutting the impeccable evidence of previous well-conducted biological efficacy studies.

Reactions to the DEVTA report have not focused on the underlying methodologic problem, namely the difference between biological and program efficacy. After stating that the "the trial authors propose that the DEVTA attenuates the global estimates of mortality reduction by half," an accompanying commentary to the DEVTA article claimed that the "DEVTA is a courageous study, and a watershed

for best practice in research to inform international development" (22). The commentator, like the authors, confuses biological and program efficacy. A letter to the editor that correctly identifies many weaknesses in the conduct of the study (23), written by the most eminent scientists in the conduct of vitamin A efficacy studies, challenges the appropriateness of including the DEVTA in meta-analyses of efficacy studies because of the poor conduct of the study. However, they did not use this as an opportunity to note the inappropriateness of including a program efficacy trial in a biological efficacy meta-analysis.

Addressing the Dialogue Gap

We perceive that mixing biological efficacy studies with program efficacy studies is illogical and scientifically incorrect. It is not just a matter of semantics because it has real consequences for program and policy decisions. Similarly, confusion about the difference between results of efficacy trials versus effectiveness studies (3) can lead to incorrect interpretation about the nature of the evidence base for policy decisions.

Confusion about terminology is inevitable in a period in which research is breaking new ground, but the consequences of not addressing it are particularly serious when the research is applied to practical problems, such as the delivery of nutrition. Decisions on planning programs and making policy depend on scientists and nonscientists having a common understanding about the concepts and terms for the facts and processes that are the basis for their decisions. At present, we are in an early stage in the process of translational research to move from science to action in the delivery of nutrition to populations. The lack of consensus about efficacy and effectiveness, as constructs, as terms, and particularly in relation to how to codify them, is so great that the basis for the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) itself is called into question, as evidenced by the confusion around the DEVTA trial, which threatens moving forward in delivering nutrition to populations. This confusion will be compounded by other confusions as we develop, refine, and test methods to deliver nutrition.

In the next section, we lay out some steps that should be taken to clarify the development of method and theory for delivering nutrition interventions.

Describe and codify concepts and vocabulary relevant to the delivery of nutrition. Description and codification of concepts are essential for establishing the environment for moving forward in the large task of translating biological knowledge in nutrition to public health action. The WHO has led the way with the development of GRADE. The GRADE approach is a major advance for several reasons: *1*) substantively, it differentiates between the biological and the behavioral (programmatic) aspects of nutrition delivery; *2*) it presents a widely accepted scientific approach to ascertaining biological efficacy (this acceptance is augmented because the approach was developed by a credible agency and the work was carried out collaboratively with key stakeholders); *3*) the approach is being widely disseminated with clear explanation for nonscientists whose acceptance of findings about biological efficacy requires that they understand them. Therefore, the confusion sown by the DEVTA trial is easily addressed within the GRADE approach by better conceptualization and codification of differences between different kinds of efficacy. A glossary of terms used by the GRADE and their uses, as well as their conceptual definitions, is needed to avoid future confusion.

Moving forward to ascertaining the validity of program interventions is much more difficult, and this will be a major long-term endeavor, particularly because there is no agreed-upon model about how this should be done. The GRADE needs to be expanded to address delivery issues, or complementary models need to be developed as discussed below under the section entitled "Develop a framework to guide program-relevant research." Equally important, this development needs to be accompanied by formulating agreed-upon vocabulary for central concepts. Among the concepts that require attention are as follows: evidence-based, policy instruments, public nutrition, delivery of nutrition, PIP, nutrient container, flow, efficiency, delivery, and acceptance. The tasks are clear. It would be wise to develop a glossary-dictionary in the format of another WHO-sponsored dictionary (3) concurrently with the intellectual development and acquisition of knowledge. This glossary would be constantly changing as our knowledge advances, but it ensures that communication is not lost as new words mutate to mean the same things or the same words mean different things. For instance, a recent WHO guideline document (24) gives the name "explanatory trials" to what epidemiologists call efficacy trials (3).

Develop methods to situate programs within a larger framework of nutritional context and nutrition-related policy instruments. Programs are one of a number of different ways in which policies result in better or worse nutrition in populations. Other policy instruments directly or indirectly translate policy into actions that affect nutrition. These include laws and regulations, subsidies, formal education, and so on. Within a framework of policy instruments, the role of nutrition-related programs must be defined conceptually, both in general terms and for planning specific programs. Through this effort it will become more apparent which policy instruments augment or diminish the effect of nutrition programs. For example, maternal schooling, which may not include a nutritional component, augments the efficiency of improving resources for nutrition to improve nutrition (25). A significant part of this framework needs to include a description of programmatic approaches to deliver nutrition. The pioneering work of Zeitlin and Austinn (26), now 30 y old, needs to be updated and expanded to include other policy instruments to deliver nutrition (e.g., fortification legislation, the WHO International Code on Marketing of Breast-milk Substitutes), as well as a discussion of the complementary and synergistic effects of these different policy instruments on nutrition.

This general framework would provide the concepts for modeling nutrition strategies to deliver nutrition within the context of the nutrition environment and its determinants. This modeling should include information about likely feasibility, effectiveness, and cost-effectiveness of different strategies and would therefore provide guidance about selecting the most appropriate policy instruments in specific situations.

Develop a framework to guide program-relevant research. Turning to the importance of the advancement of what is sometimes referred to as "implementation science" or "delivery science," there are already a number of resources that nutrition can draw on, including a recent excellent WHO guide to implementation research (24). More than a decade ago, WHO produced a larger framework to guide funding for the development and evaluation of public health interventions (27). This framework was based on a schema for identifying implementation bottlenecks and conducting research to address these. The framework provides a starting point on which one could build.

A nutrition-specific program framework could be used to identify inefficiencies in the program delivery process—the areas that require research to provide programs with essential information on how to address them to improve delivery. For example, the biological efficacy of calcium for maternal health in pregnancy has been established. However, the mode of delivery through economically affordable pills is not likely to be programmatically feasible unless some means are identified to deliver pills that are reduced in size. Another example is the strong empirical base that has been established concerning the importance of behavior change communication in nutrition interventions. However, the amount of research on the organization and management of this essential program feature, and methods for assessing knowledge flow in specific cultural and organizational contexts, is minimal. A program framework needs to include guidance on approaches and techniques for assessment and response to this potential bottleneck and source of inefficiency in the delivery process.

A framework to guide program-relevant research would also identify specific knowledge areas in which biological efficacy is sufficiently well established that attention should now be directed primarily to implementation issues. Examples of knowledge in this category are the importance of breastfeeding for infant health or the biological efficacy of the nutrients reviewed in (1). Further replicative research is no longer necessary to substantiate this knowledge. Of course, more research on these issues will be necessary if our present understanding is called into question by appropriate science, such as evidence that potential to benefit is impaired in some populations but not in others. For instance, carotene effectiveness in improving vitamin A nutrition is impaired by intestinal parasites or lack of fat in the diet, which was discovered through a biological efficacy trial (28). Another reason for performing efficacy trials might be to show that a population no longer needs vitamin A programs because it is no longer vitamin A deficient. This would be a novel indication for a new round of biological efficacy trials and deserves more reflection in the context of developing the framework.

Another part of the framework will need to address program implementation assessment and the components of feasibility of implementation, effectiveness/efficacy, cost, and other issues in the implementation context. This part requires formulation of an integrated evaluation methodology that includes adequacy and plausibility as well as the more usual outcome indicators (29). It would also identify when probability designs (e.g., RCTs) are needed and appropriate (30) and when other approaches will produce more meaningful information.

Conclusions

The WHO and the nutrition community have made major advances in establishing the evidence through the GRADE for the biological efficacy of critical minerals and vitamins for child and maternal health. This approach is so robust that the threats to its success will be easily addressed, as discussed above. The success is a good harbinger for the next difficult steps in delivering nutrition through programs and other policy instruments. In this article we suggest that one of the important requirements for bringing research on intervention delivery to the same level as for biological efficacy is the development of frameworks for investigating nutrition interventions within a changing nutrition environment, and for a systematic approach to implementation research. To be successful, there will need to be consensus. Establishing fora and systems for communication among scientists, policy makers, program planners, and implementers will be essential. Part of the challenge is institutionalizing the kind of successful outreach that WHO and others are doing for the GRADE. It will also require attention to a body of concepts and terms and a theoretical, as well as empirical, structure for intervention research in nutrition.

Acknowledgments

We thank Monika Bloessner for her review and insightful edits. Both authors read and approved the final manuscript.

Literature Cited

1. WHO. Nutrition evidence-informed guidelines [cited 2013 Jul 21]. Available from: http://www.who.int/nutrition/publications/micronutrients/guidelines/en/index.html.
2. Mason JB, Habicht JP, Greaves JP, Jonsson U, Kevany J, Martorell R, Rogers B. Public nutrition. Am J Clin Nutr. 1996;63:399–400.
3. International Epidemiological Association. Porta M, editor. A dictionary of epidemiology.5th ed. Oxford: Oxford University Press; 2008.
4. Peña-Rosas JP, De-Regil LM, Rogers LM, Bopardikar A, Panisset U. Translating research into action: WHO evidence-informed guidelines for safe and effective micronutrient interventions. J Nutr. 2012;142 Suppl:197S–204S.
5. Tovey D. The role of The Cochrane Collaboration in support of the WHO nutrition guidelines. Adv Nutr 2014;5:35–9.

6. Campos FA, Flores H, Underwood A. Effect of an infection on vitamin A status of children as measured by the relative dose response (RDR). Am J Clin Nutr. 1987;46:91–4.
7. Mbuya MNN, Menon P, Habicht J-P, Pelto G, Ruel MT. Maternal knowledge following nutrition behavior change communication is conditional upon both health workers' knowledge and knowledge sharing efficacy in rural Haiti. J Nutr. 2013;143:2022–8.
8. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health. 2004;94:400–5.
9. Habicht J-P, Martorell R, editors. The development and legacy of the INCAP Oriente studies 1969–2009. J Nutr. 2010;140:392–3.
10. Rivera JA, Sotres-Alvarez D, Habicht JP, Shamah T, Villalpando S. Impact of the Mexican Program for Education, Health, and Nutrition (Progresa) on rates of growth and anemia in infants and young children—a randomized effectiveness study. JAMA. 2004;291:2563–70.
11. Ruel MT, Menon P, Habicht JP, Loechl C, Bergeron G, Pelto G, Arimond M, Maluccio J, Michaud L, Hankebo B. Age-based preventive targeting of food assistance and behaviour change and communication for reduction of childhood undernutrition in Haiti: a cluster randomised trial. Lancet. 2008;371:588–95.
12. Bonvecchio A, Pelto GH, Escalante E, Monterrubio E, Habicht JP, Nava F, Villanueva MA, Safdie M, Rivera JA. Maternal knowledge and use of a micronutrient supplement was improved with a programmatically feasible intervention in Mexico. J Nutr. 2007;137:440–6.
13. Donegan S, Maluccio JA, Myers CK, Menon P, Ruel MT, Habicht J. Two food-assisted maternal and child health nutrition programs helped mitigate the impact of economic hardship on child stunting in Haiti. J Nutr. 2010;140:1139–45.
14. Arifeen SE, Hoque DME, Akter T, Rahman M, Hoque ME, Begum K, Chowdhury EK, Khan R, Blum LS, Ahmed S, et al. Effect of the integrated management of childhood illness strategy on childhood mortality and nutrition in a rural area in Bangladesh: a cluster randomised trial. Lancet. 2009;374:393–403.
15. Leroy JL, Ruel M, Verhofstadt E. The impact of conditional cash transfer programmes on child nutrition: a review of evidence using a programme theory framework. J Dev Eff. 2009;1:103–29.
16. Hawkes C, Ruel M. Value chains of nutrition. 2020 Conference: Leveraging Agriculture for Improving Nutrition and Health. New Delhi, India. Washington: International Food Policy Research Institute; 2012.
17. Pearl J. Causality: models, reasoning and inference. 2nd ed. Cambridge (UK): Cambridge University Press; 2009.
18. Funnel SC, Rogers PJ. Purposeful program theory: effective use of theories of change and logic models. San Francisco: Jossey-Bass (Wiley); 2011.
19. Stoltzfus RJ. How can the scientific community support the generation of the evidence needed to improve the quality of guidelines for micronutrient initiatives? Adv Nutr 2014:5;40–5.
20. Awasthi S, Peto R, Read S, Clark S, Pande V, Bundy D; DEVTA (Deworming and Enhanced Vitamin A) Team. Vitamin A supplementation every 6 months with retinol in 1 million pre-school children in North India: DEVTA, a cluster-randomised trial. Lancet. 2013;381: 1469–77.
21. Awasthi S, Peto R, Read S, Clark S, Pande V, Bundy D, and the DEVTA (Deworming and Enhanced Vitamin A) team. Vitamin A supplementation every 6 months with retinol in 1 million pre-school children in north India: Supplement to: DEVTA,a cluster-randomised trial. Lancet [cited 2013 Mar 14]. Available from: http://dx.doi.org/10.1016/S0140-6736(12)62125-4.
22. Garner P, Taylor-Robinson D, Sachdev HS. DEVTA: results from the biggest clinical trial ever. Lancet. 2013;381:1439–41.
23. Sommer A, West KP, Martorell R. Vitamin A supplementation in Indian children. Lancet. 2013;382:591.
24. Peters DH, Tran NT, Adam T. Implementation research in health: a practical guide. Geneva: World Health Organization; 2013.
25. Reed BA, Habicht JP, Niameogo C. The effects of maternal education on child nutritional status depend on socio-environmental conditions. Int J Epidemiol. 1996;25:585–92.
26. Zeitlin MF, Austin JE. Nutrition intervention in developing countries: an overview of nutrition programs. Oelgeschlager, Gunn & Hain; Boston:1981.
27. de Zoysa I, Habicht JP, Pelto G, Martines J. Research steps in the development and evaluation of public health interventions. Bull World Health Organ. 1998;76:127–33.
28. Jalal F, Nesheim MC, Agus Z, Sanjur D, Habicht JP. Serum retinol concentrations in children are affected by food sources of beta-carotene, fat intake, and anthelmintic drug treatment. Am J Clin Nutr. 1998;68:623–9.
29. Habicht J-P, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. Int J Epidemiol. 1999;28:10–8.
30. Methodologies to evaluate the impact of large scale nutrition programs. Doing Impact Evaluation Series, No. 13. Washington: World Bank; 2009.