Probability, Plausibility, and Adequacy Evaluations of the Oriente Study Demonstrate That Supplementation Improved Child Growth 1,2

Jean-Pierre Habicht 3* and Reynaldo Martorell 4

3Division of Nutrition Sciences, Cornell University, Ithaca, NY 14853 and 4Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA 30033

Abstract

This article presents evidence that the high-nutrient supplement in the Oriente study (Atole) improved child growth. The evidence is presented at 4 levels. There was a causal effect of the intervention on child length, as assessed by probability analyses of the randomized, controlled trial (P < 0.05). The plausibility analyses, which included an examination of wasting, showed that the nutritional impact was due to the Atole, especially in those who were <3 y old and who suffered from diarrhea. The adequacy analyses revealed excellent biological efficacy of the Atole at the individual level. At the level of the whole population, the efficacy of impact was much less, because many children did not participate fully in the supplementation program. The external validity of the biological impact is likely to be good for populations with similar diets and medical care. J. Nutr. doi: 10.3945/jn.109.114496.

Introduction

Over the years, from the initial publications of the Institute of Nutrition of Central America and Panama’s Oriente Longitudinal Study (1969–77) to recent analyses (1), researchers have based their conclusions on the inference that the differences in outcomes between the villages that received a high-nutrient supplement (Atole) and those that received a lower-nutrient supplement (Fresco) were due to nutrition. The composition, nutrient content, and the distribution of these supplements are given in Ramirez-Zea et al. (2). The purpose of this paper is to summarize the evidence to support this inference from the perspective of a larger framework of research concerns (3). The evidence is discussed at 4 levels: 1) the probability analyses of a randomized control trial; 2) plausibility analyses; 3) adequacy analyses; and 4) external validity.

The structure of this review follows recommendations for evaluating interventions (4–6) in a fashion that permits one to take advantage of the complementarities of randomized controlled trial (RCT) analyses, which result in probability statements, and of analyses that appeal to plausibility for their inferences, as well as analyses that address adequacy and external validity issues. In this paper, we limit the application of this framework to the results of the Oriente nutrition intervention on physical growth in children. Throughout, it is important to keep in mind that the objective of the Oriente study was to determine the efficacy of protein supplementation on child development and growth (7). It was not intended, nor designed, to be a public health effectiveness trial of a nutrition intervention.

Probability analysis: the RCT

Previously, we have shown (8) that in 1973 there was a 2.45-cm (P < 0.05) difference in length at 3 y of age between the Atole and Fresco children who had lived in the villages from birth to 3 y of age compared with the initial difference between the villages in 1968 before supplementation began. This analysis took advantage of the cluster RCT design, with a replicate of 2 villages per treatment, as presented in Ramirez-Zea et al. (2). The randomization in an RCT incorporates into the statistical test of impact all possible confounding that may have been present between the individuals at baseline or accumulated throughout the trial. Thus, the probability statement (e.g. P < 0.05) of significance is not just a statement of association but a statement of causality that the intervention caused the outcome. It is the so-called golden standard of evidence-based decision-making (9).

The idea of randomizing clusters, such as villages, to the nutrition supplement intervention, rather than individuals, was proposed for the Oriente study in the late 1960s by John E. Gordon, who was then Professor Emeritus of Epidemiology,
Harvard School of Public Health. Since then, in the intervening 40 y, we have learned about many of the pitfalls of using this design (10).

The most serious pitfall is the potential for differences in response to the intervention across comparison clusters as a consequence of initial differences. This results in a failure to achieve significance. In the Oriente study, this pitfall was avoided by the care given to choosing the village pairs from many candidate pairs (9). However, the SD, \( \pm 1.7 \) cm, of the initial mean lengths of 3-y-old children across the 4 villages (8) was still large enough that we would have had to have 2-fold more pairs of villages to meet today’s requirements for statistical power (power \( P \geq 0.80; \) \( P < 0.05 \)) and 5-fold more to meet the recommendation of a power of 0.90 (\( P < 0.05 \)). We were, in fact, incredibly lucky that the response between Atole and Fresco consumers across the large village pair (2.55 cm) and the small village pair (2.35) (8) were so similar that we had the statistical power to state with high probability (\( P < 0.05 \)) that the differences between the Atole and Fresco villages were due to the intervention, which included the supplementation.

Typically, in an RCT, investigators stop the analysis when causal impact has been demonstrated (10).

**Plausibility analyses**

The probability analyses of the RCT findings demonstrated that the intervention improved growth. Thus, there is no need for plausibility analyses to bolster the claim of causality, which is the major objective of usual epidemiological analyses (11) that cannot establish a probability of causality (12). However, the probability analysis demonstrating impact does not give any information about which of the several different components of the intervention were responsible for the impact. Some of the components, medical care, weekly home interview visits, and cognitive and anthropometric measuring sessions, can be excluded from consideration, because they were controlled by the design. Furthermore, care was taken to ensure that their implementation was applied equally across the Fresco and Atole villages. Identical procedures were followed and the same personnel were rotated on a fixed schedule. Also, the impact of Atole compared with Fresco cannot be due to maternal supplementation, because there was no differential impact between the 2 treatments in birth weight (13).

**The role of plausibility analyses within a successful RTC.**

Other aspects of the intervention, however, were outside the control of the project. Some of these, particularly the decision to participate in the intervention, the frequency of attendance at the centers where the supplements were given, and the volume ingested, are directly linked to the potential impact of the intervention. The RCT probability analyses excluded the possibility that the impact was due to self-selection for participation or for ingestion of the supplements. However, these probability analyses say nothing about the pathways that led from the intervention to the impact, so one cannot conclude that the growth impact was due to the supplement. To shed light on what goes on inside the black box of an RCT, one needs to supplement probability analyses with plausibility analyses (5) of the pathways. Analyses of these pathways are sometimes referred to as Program Impact Pathway (PIP) analyses. Statistical methodologies to facilitate such analysis are being newly applied (14).

The last step in the PIP between intervention and the impact on growth is the efficacy of the supplement to improve growth. Children who did not consume any supplement from 1 to 3 y of age grew 15.7 cm in that interval compared with 18.0 cm for children who consumed 1 serving/d [163 kcal/d (682 kJ/d)] over the same period (15). Eighteen centimeters is only 1 cm below the WHO growth standard of 19 cm (16), which indicates that the Atole was associated with an efficacious response of 70% in that the supplement improved growth 2.5- of a 3.3-cm potential for improvement. Thus, the dose-response analysis provides plausible evidence that the impact of the intervention was due to the supplement.

However, one cannot know whether the impact was due to energy, protein, or the micronutrients. In pregnant women (13), there was a substantial overlap in energy intake so that one could compare the impact of Atole to Fresco for the same energy intakes, but very different protein and micronutrient intakes, to demonstrate that neither protein nor the micronutrients had any impact beyond the impact of energy. However, there was little overlap in energy intake from 1 to 3 y of age in children between the Fresco and Atole groups (17). Thus, one cannot use a plausibility analysis to compare the differential impact of energy, protein, and micronutrients separately as was done for pregnant women (13). In the case of children, the impact on growth could be due to energy and/or protein and/or micronutrients.

**Who benefited biologically?** PIP analysis can also be used to identify those within an RCT sample who account for the impact found, because they were the ones who benefited from the intervention. Differential impact can be assessed through analyses of effect modifications (synergisms and antagonisms).

Analyses of data from a 1973–80 supplementation program in Colombia reported that supplementation offset the negative effect of diarrheal disease on length (18). We also found an offset so that impact of the Oriente intervention on growth from 3 to 36 mo was 3 times greater in children with high diarrhea rates. The difference in growth over this period between Atole and Fresco children was 4.3 cm at high levels of diarrhea (defined as 50% of time ill) compared with 1.4 cm in children with no diarrhea (19). Similarly, wasting associated with diarrhea was more responsive to the intervention for both prevention (20) and recovery (21) than was wasting not associated with diarrhea. The implication of this finding is that the benefits of improved nutrition depend on the underlying diarrhea levels in a population. This conclusion affects the generalizability (external validity) of nutritional impact studies.

Another biological effect modification is the age of the child. The Oriente study results (20–23) have had a major influence on directing attention to the fact that the ill effects of malnutrition on growth cease earlier than was previously recognized. The effect is limited primarily to the first 3 y of life.

**The importance of the PIP to highlight behavioral effect modification.** The primary effect modifier in the Oriente study was the gap between the delivery of the supplement at the centers and the amount that individuals ingested. The statement that those who did not ingest it did not benefit is so obvious that it appears to be a truism. However, the implications for the implementation and interpretation of the results of an RCT are often ignored. Interpreting and extrapolating the impact results of the RCT requires understanding the behavioral PIP of ingestion, the pathway from participation in the intervention, to attendance of those who participated, to the ingestion of the supplement by those who attended. Analyses of the pathway showed that participation and ingestion of the supplement were not consistently associated with any determinants on which the study collected data. The major consistent influences on attendance were distance to the village feeding center and socioeco-
conomic status, for both mothers (24) and children (17). Because the centers were located in the middle of compact villages, the fact that distance predicted attendance means that one cannot rely on supplementation centers for the delivery of supplementation unless they are located close to the targeted beneficiaries. On the other hand, the analyses of who benefited show that the supplements were well targeted in that they reached those who needed them most.

**Adequacy analyses**

Adequacy analyses complement the RCT probability analysis for impact and the plausibility analyses for the PIP. Adequacy analyses ascertain whether the biological (clinical) and the population impact are adequate. The supplement was clinically 70\% efficacious for the long-term responses. Clinical efficacy to cure short-term malnutrition (wasting) was even higher (close to 100\%) (25). Thus, we conclude that the intervention was adequate relative to its biological efficacy.

The adequacy of the intervention to prevent stunting in the population was moderate. In the full sample, at 3 y of age after 3 y of exposure, the *Atole* children were 83.7 cm tall, which is 8.0 cm less than the WHO growth standard. The impact of the *Atole* relative to the *Fresco* of 2.45 cm was about one-third of the 7.95-cm deficit between the growth of the *Fresco* children and the WHO growth standard.

The discrepancy between 70\% adequacy of biological efficacy and the 31\% population adequacy was due to the low volume of supplement ingestion (17) even though the supplement was accessible to all. Addressing the behavioral steps that lead to ingestion when delivery of the supplement is assured remains a major challenge to developing “delivery efficacy” trials such as the Oriente study.

**External validity**

The evaluation procedure described above identified factors that need to be taken into account in assessing the generalizability of the findings to other situations, or the external validity of the study. In the consideration of external validity, one also needs to examine factors that do not vary across treatment groups but that affect the potential to respond to the intervention. The most important one in the Oriente study is diet. The amount and quality of protein in this maize-bean diet was higher than one finds in, e.g. poor cassava-eating populations for whom the higher protein supplement might be more beneficial.

Other issues that need to be considered for external validity are intervention components that were not envisaged in the original design and that can therefore escape attention in ascertaining external validity. In the case of the Oriente study, there were 2 major components that fall into this category: medical care and the nutritional content of the *Fresco*.

**Treatment of human participants in the Oriente study.** In 1969 when the Oriente project was being developed, concerns about the treatment of human participants were focused on “causing no harm.” The intervention itself and the measurement methods used met these concerns. However, 2 issues disturbed us, albeit intuitively, because ethical principles of human studies (26) were not yet laid out. One was that the quality of medical care that was provided initially to participants was much less than one could have done at about the same cost. Therefore, we decided to establish an effective medical system (26,27). The funding agency opposed our improving the quality of medical care, because it would destroy the external validity of the results for most developing countries. We argued that internal validity was more important and that uncontrolled illnesses would blur and bias the impacts that were the objective of the study. We were probably wrong about the blurring, at least to some degree, because the impact of better nutrition turned out to be greater for ill children (at least those ill with diarrhea). However, we were certainly correct about biases due to preventable epidemics, one of which (whooping cough) occurred in some, but not other, villages just before the initiation of the study.

The second issue was that we touted the supplements as something that was “good for you.” Originally, *Fresco* contained only nonnutritive energy. Equilibrating to the level of *Atole* the amounts of micronutrients that were thought to be limiting in the diet improved the nutritional benefit of the *Fresco*. We argued, correctly, that this would increase the external validity of the study, because it would permit a clearer inference about the impact of protein. In the case of protein, we were correct, at least with respect to pregnant women. This was especially important, because the study was designed on the presumption that protein was the limiting nutrient in the diet. As discussed above, the overlap between protein and energy in children’s intake makes it impossible to make this determination with respect to children.

The external validity of implementing our procedures in a public health setting is of interest. Our total control over the delivery of the supplement means that this study is classifiable as a “delivery efficacy” trial (5). It is not an effectiveness trial of a similar delivery system within a usual public health program. It is sobering that the conceptualization of public health efficacy and effectiveness trials in nutrition requires much more work to stage them in an appropriate sequence and to design them properly.

**Conclusions**

In conclusion, the combination of the results from the probability analysis and the plausibility analyses lead to the conclusion that the impact on child growth of the *Atole* (high supplementation) intervention is not only causal, it is also nutritional. Paradoxically, the successful RCT, which delivered the strong inference that the intervention had an impact on child growth, could not identify the nutrient(s) responsible for the impact. This is in contrast to the findings reported for supplementation during pregnancy and birthweight (13), where the RCT failed in that there were no differences between the intervention villages, but where the evidence for a causal effect of energy is persuasive.

**Acknowledgments**

We thank Gretel Pelto for discussions about and improvements to this paper. J-P. Habicht wrote the paper and had primary responsibility for its final content. R. Martorell provided critical review of the paper. Both authors read and approved the manuscript.

**Literature Cited**


