Evaluation of Dietetic Product Innovations: The Relative Role of Preclinical and Clinical Studies

Maria Makrides, Robert A. Gibson

Women's and Children's Health Research Institute and University of Adelaide, Adelaide, S.A., Australia

Abstract

A variety of systems are used to establish efficacy of food ingredients. Immortal human cell lines have the advantage of rapid throughput and often have the ability to point to mechanisms of action. Transgenic and natural variants of animals (usually rats and mice) have proven to be extremely useful in elucidating effects in vivo, although extrapolation of results to humans has risks. Animal models are also useful in establishing safety and toxic levels of ingredients. Human trials have the most relevance to society. Types of evidence for efficacy rise from improved status level in subjects as a result of eating food (long-chain polyunsaturated fatty acid, levels in erythrocytes), change in surrogate markers as a result of eating food (plasma cholesterol or glutathione peroxidase activity), change in a physiological outcome (such as visual evoked potential acuity or heart rate variability) through to the highest level of evidence, a change in a clinical outcome (improved global development, reduction in infections) established in randomized controlled trials. Ultimately, there is a need for tests of pragmatic interventions that can easily be incorporated into usual dietary practices of the culture in which it is tested.

There is great potential for nutritional interventions in early life to result in improved health outcomes. This is based on the large body of evidence of both experimental and epidemiological studies showing that good nutrition during pregnancy and early life may enhance neurodevelopmental outcomes, reduce the prevalence of allergies, improve body composition and may ultimately reduce the prevalence of chronic diseases [1–3]. With such promise, the evaluation of nutritional interventions, which often take the form of specialized products or food innovations, is of paramount
importance. The scope of this paper is to review the relative role of preclinical and clinical studies in the assessment of both safety and efficacy for new food innovations.

**Role of Preclinical Studies**

There is a range of preclinical studies that are relevant to the assessment of food innovations, and these extend from cellular studies to studies with experimental animals. There are immortal cell lines available to screen bioactive molecules or fractions that are dietary components. Such assays can happen quickly, and it is possible to investigate or screen multiple bioactive compounds in specific cell types and gain insight into possible mechanisms of action. However, if it difficult to assess the relevance of such studies to women and children when the putative bioactive compounds are included as part of a dietary regimen. Such cellular studies provide an important first-pass evaluation to select bioactive molecules (for example, protein or lipid fractions) worthy of further investigation.

Animal models offer greater diversity and specificity of effect than is possible in cell studies, although they are more time consuming and resource intensive. Numerous models are available to assess the effects of dietary components in situations relevant to the human target group. For example, the (rat) pup-in-a-cup model aligns well with the neural and gastrointestinal maturity of a very preterm infant [4]. The intricacies and complexities of the model in many ways are not surprising as the rat pup requires some of the extra supports (thermoregulation) that would also be required in a neonatal intensive care unit. More commonly, however, animal models are based on genetic predisposition, such as Brown Norway rats, which are allergy prone, and offer a model of an allergy-sensitive human [5]. Animal models are useful in identifying target outcomes for human trials. It is possible to harvest organs, and so provide information about how nutrients or bioactive ingredients are acting. There are, however, some dangers in over-extrapolation to humans as effects in animal models are not always translated to the human situation. For example, conjugated linolenic acid has a long history of improving growth and body composition in animal studies and is widely used in the pig industry for this reason. However, human studies have not consistently demonstrated the positive effects observed in other animals [6, 7].

One of the most important roles of animal studies is safety evaluation. Safety in experimental animals is commonly assessed using a toxicological approach where the innovative ingredient is fed at concentrations well beyond what would normally be expected in typical dietary patterns. This allows the determination of tolerable safe levels and gives an indication of the safety buffer in relation to usual dietary intake.
Role of Clinical Studies

The ultimate evaluation of safety and efficacy of new food innovations is through well-designed and appropriately powered randomized controlled trials (RCTs). Such trials are complex and expensive and involve a large investment from all involved including the participants, the researchers and the industry. RCTs are therefore the final studies in the pathway to new, innovative products with proven clinical efficacy. However, before arriving at large-scale RCTs, different types of clinical studies are often undertaken to answer questions of bioavailability or tolerance in order to ensure that the product tested in the large-scale RCTs has an optimized composition and a maximum chance of resulting in the desired clinical benefit. The following section uses the addition of long-chain polyunsaturated fatty acids (LC-PUFA) to infant formula as a case study of the pathway from small, focused biochemical human studies to large-scale RCTs.

Bioavailability and Biochemical Status Studies

The earliest intervention studies to show that adding the LC-PUFA, docosahexaenoic acid (DHA, 22:6n-3) and eicosapentaenoic acid (EPA, 20:5n-3) as fish oil, to infant formula resulted in increased plasma and erythrocyte DHA and EPA concentrations were reported in the late 1980s [8]. This change in biochemical status was evidence of bioavailability, although more intricate studies of absorption [9] and dose response using different LC-PUFA sources followed [10, 11]. The value of these studies was the demonstration that the key LC-PUFA are efficiently absorbed from triglyceride and phospholipid sources even by preterm infants with an immature gastrointestinal tract [9]. In addition, the biochemical response in plasma and erythrocytes was dependent on the concentration in the diet/product, and the degree of response was equivalent between different sources [12].

Safety and Tolerance Studies

Although measures of tolerance and safety are often included in many study designs, it is also considered mandatory to conduct clinical studies with the aim of demonstrating that the new product containing the innovation is equivalent to the standard product. For studies involving infants and young children, growth is most often used as a surrogate for safety. As nutrition and growth are inextricably linked, nutritional interventions that have a negative influence on growth represent an undesirable change that in many cases is associated with negative shorter or longer term clinical outcomes. It was for this reason that significant concern was raised in the LC-PUFA field when the results of some of the earliest intervention studies involving preterm infants suggested that supplementation of infant formula with n-3 LC-PUFA was related to poorer weight and length gains compared with unsupplemented formulae [13], while conversely resulting in improved visual function [14, 15].
It was this paradox that was of concern as poor growth of premature infants is well known to be associated with poor neurodevelopmental outcomes [16], the very domains that LC-PUFA supplementation was postulated to support. Nevertheless, the published intervention trials currently available suggest little or no effect of LC-PUFA supplementation of infant formula on growth of preterm infants [17–19]. However, it should be noted that the majority of studies that have investigated the effect of LC-PUFA supplementation on infant growth were not designed as trials of equivalence, that is, clinical trials specifically designed and powered to demonstrate equivalence of growth between the supplemented and unsupplemented, control infant formula [20]. Such trials require the investigators to decide on the smallest difference that would be considered clinically significant and set the confidence interval that would include the mean of groups, treatment and control, to claim equivalence. It is therefore not uncommon for equivalence trials to require larger sample sizes than trials which are designed to test hypotheses of difference.

**Studies Designed to Show Changes in Surrogate Markers or Physiological Responses**

Surrogate markers or physiological responses are used as outcomes in clinical intervention trials to provide an indication of a likely effect in an associated clinical outcome. Trials with surrogate or physiological outcomes are often smaller (fewer participants) and have a faster turnaround time than trials with clinical outcomes. In the LC-PUFA field, different measures of infant visual acuity have been used as short-term assessments that may be indicative of longer term neurological maturity [21]. The visual acuity studies in the LC-PUFA field have been useful in clarifying the dose-response for preterm infants [22] and have also helped to elucidate the more subtle response of term infants [23]. However, such surrogate or intermediate outcomes are often less complex and focus on specific developmental domains and therefore do not consistently predict global neurodevelopmental outcome.

**Studies Designed to Show Changes in Clinical Outcome**

Although often complex, time consuming and expensive, RCTs with clinical outcomes provide the most robust and directly relevant answers regarding the efficacy (and safety) of new food or supplement innovations. It is for this reason that major RCTs are not generally undertaken without a body of congruent evidence from preclinical studies and other human biochemical or physiological studies that all point towards a safe and efficacious dietary intervention. It has been relatively uncommon for large scale RCTs to be undertaken in early life nutrition, and the cases of successful large-scale RCTs have best been achieved with a combination of government and industry funding. This underscores the large investment required. The coinvestment by government is particularly noteworthy because it highlights the acceptance that nutritional interventions during early life have the potential to change longer term outcomes that are
important to the functioning of the individual as well as to the community more generally. The most recent and relevant example from the LC-PUFA field is the DINO (DHA for the Improvement in Neurodevelopmental Outcome) trial in preterm infants born before 33 weeks’ gestation [24].

The significance of the DINO trial comes from the fact that the developmental quotient of children born preterm is 11 points (95% CI: 9–13) lower than term-born controls [25]. In addition, preterm children have a higher incidence of attentional problems [26], impaired executive functioning [27], reduced memory and learning capacity [28], and visual-spatial perceptual deficits [29]. Collectively, these cognitive impairments compound so that preterm children have higher rates of learning disability, a greater need for integration assistance, and an increased likelihood of repeating a grade at school compared with their term-born counterparts [30, 31]. Therefore, any intervention with potential to enhance cognitive development for preterm children, and hence improve quality of life and decrease the burden on families and society, is considered a priority and worthy of investment.

The DINO trial included all preterm infants born less than 33 weeks’ gestation regardless of whether infants were fed expressed breast milk or infant formula. DINO demonstrated that DHA given at a dose designed to approximate the in utero accumulation rate (three times the standard dietary dose) resulted in fewer preterm children with significant cognitive delay at 18 months corrected age compared with control (5.2 vs. 10.5%; p = 0.03), although there was no overall difference in the mean developmental quotient [24]. This was explained by two significant interactions (diet by sex and diet by birthweight strata). The effect of DHA supplementation was most pronounced in girls born <33 weeks’ gestation and in infants born weighing <1,250 g [24]. Despite the complex results, the importance of the DINO trial is that of all the neonatal interventions tested in children born preterm (drugs, nutrients, environmental) only caffeine and increased dietary DHA have shown promise as strategies to improve cognitive outcomes [24, 32].

Three key lessons to come from the DINO trial that are important for other large-scale nutrition interventions are, first, the importance of having a pragmatic intervention that can be easily incorporated into usual dietary practices, second, the need for an appropriate (often large) sample size with minimal attrition to underscore the robustness of the outcomes, and finally the role of the independent scientific researcher is vital to ensure potential outcomes of true public health importance and secure the funding relationship between industry and the competitive government funding.

References

27 Anderson PJ, Doyle LW: Executive functioning in school-aged children who were born very preterm or with extremely low birth weight in the 1990s. Pediatrics 2004;114:50–57.
Discussion

Dr. S. Koletzko: Dr. Makrides, did you find in the reviewed studies and also in your own DINO study any hints concerning other beneficial outcomes for LC-PUFA supplementation like immunological or infectious parameters, immune response to vaccination, or incidence of atopic diseases?

Dr. Makrides: We did measure many other clinical outcomes [1]. There were no differences except for chronic lung disease. Fewer babies in the high DHA group required oxygen therapy at 36 weeks postmenstrual age, and again there was an interaction. The effect was driven by the smallest babies and the boys, so neonatal clinicians that are interested in respiratory outcomes have been more focused on this result rather than the developmental results, which was the primary reason for doing the trial. We have also measured parental report of allergies through to 18 months. There were no differences in the medical diagnosis of asthma, which you would expect at such a young age, eczema, food allergy, but there was a lower prevalence of doctor diagnosis or medication for hay fever in children in the high DHA group. Allergy outcomes are part of the 7-year follow-up.

Dr. B. Koletzko: I greatly appreciated your presentation and particularly the conclusions that you have drawn. I couldn’t agree more with your recommendation to do clinical trials properly with adequate power. You described very nicely the systematic review of the trials on PUFA and visual acuity. You concluded that in term infants half of the trials describe a benefit, others describe no difference. Based on this analysis, you conclude that only a trial in preterm infants is justified. I must admit I do not understand the basis for the conclusion. There are differences in results, heterogeneous studies, different interventions, different outcomes, outcome measures, so further studies may well be justified to add clarity.

Dr. Makrides: I was actually saying that when you stand back and consider all the data as a researcher you focus on the research question that is likely to have an important public health impact and be worthy of a large financial and human resource investment. For this reason, we focussed on preterm infants. Regarding term infants, I think there are different issues here. The issue of LC-PUFA and term infants implies infant formula, whereas for preterm infants we are considering both breastfed and formula-fed babies.

Dr. B. Koletzko: I would fully agree then if you reword the conclusion to say it is more likely to find something in preterms. The second question I have relates to your categorization of visual acuity measures as a surrogate marker and the DQ measure as an end point marker. I am just wondering how you define a surrogate marker and an end point marker. Is DQ really an endpoint or are not both markers of function, even though potentially of different predictive value? If we count the number of deaths or the number of patients that have a remission of leukemia then we have an endpoint. Is DQ really an endpoint?
**Dr. Makrides:** In the LC-PUFA field, visual function has been used largely as a surrogate or physiological outcome rather than a clinical outcome. For example, VEP is largely a physiological measure. It is possible to be blind and still have a very normal VEP response. The LC-PUFA studies using VEP as an outcome were designed to exclude children with visual abnormalities so that the VEP response could be used as a marker of the maturity of the visual pathway and what that might tell us about neural maturation because it’s easy to measure during early life. The issue of DQ is a more complex one. When children are less than 2 years of age, tests like the Bayley Scales give a good indication of global developmental delay indicated by whether children are falling 1 or 2 standard deviations below the mean. For more subtle, clinically relevant changes, testing at older ages is needed. The clinical relevance of VEP function is debatable as we do not know of robust or consistent associations between VEP acuity and later outcome. This is why I said that DQ was more clinically relevant than VEP acuity.

**Dr. B. Koletzko:** Well that’s debatable. If you would take the same effect size, for example your intervention changes visual acuity by 3 standard deviations, I would predict that has effects on the perception of the environment and learning. But let me move to my last and third question with respect to the really outstanding DINO trial that you have performed. You said the follow-up, which is admirable, will tell us whether there are important effects, in other words you regard effects only as important if they persist until 7 years of age. As a pediatrician I find that a rather unfair assessment. For example, would you consider iron supply in infancy as irrelevant if it improves iron status in early childhood but not permanently into school age and later? Or if you take Anneli’s example, do you consider it irrelevant to diagnose celiac disease at 1 year rather than 3 years, which will reduce suffering of the child during 2 years, but it’s a transient effect and probably will not change outcomes in adulthood. So why is an effect irrelevant if it is transient?

**Dr. Makrides:** With regard to the DINO trial, the 7-year follow-up will actually give us the conclusive outcome data in terms of impact into adulthood because what we can measure at 7 is much more likely to be predictive of adult IQ than what we can measure at 18 months. I am not denigrating the 18 months data, I think it’s incredibly important, but the 7-year outcome data will be more robust in terms of understanding the full public health impact.

**Dr. Solomons:** It’s a comment directly related to the last example. I want to point out that Dr. B. Koletzko probably is unaware of the studies by Betsy Lozof published in *Pediatrics* which show the transient situation. Anemia in infancy related to iron deficiency has a permanent long-term effect on cognition. They have studied the children until their adolescence, so transiency of the syndrome has nothing to do with long-term outcome of a dependent functional outcome, and I think you should apologize to everyone for forgetting that you knew that.

**Dr. Greer:** I was just going to comment on the visual acuity versus the IQ. Personally, I would be much more in favor of using the IQ as a primary outcome. If you look at what I consider the best studies in visual acuity with infants randomized to LC-PUFAs or control, the difference between groups amounts to one line on the standard Snelling eye chart. Does it really make a difference if your visual acuity is 20-20 rather than 20-15? It is a quantitative assessment, but still it’s not really as important as an IQ difference.

**Reference**