A R T I C L E

RESPONSE TO COMMENTARY: ETHICAL DIMENSIONS OF THE BEIP

CHARLES H. ZEANAH
Tulane University Health Sciences Center

SEBASTIAN F. KOGA
University of Virginia School of Medicine

BOGDAN SIMION
SERA Romania

ALIN STANESCU
Institute of Maternal and Child Health, Bucharest

CRISTIAN L. TABACARU
Formerly, Secretary of State for Child Protection, Government of Romania

NATHAN A. FOX
University of Maryland

CHARLES A. NELSON
Boston Children's Hospital/Harvard Medical School

BEIP CORE GROUP

We appreciate the thoughtful commentary on our paper that Dr. Wassenaar provided and welcome the opportunity to address several concerns raised in it. Because of space limitations, we did not cover all aspects of the study in the article, and we are pleased to have an opportunity to clarify several points here.

Although noting that randomized controlled trials (RCTs) are the gold standard of intervention studies, Dr. Wassenaar raised two concerns about the scientific validity of the BEIP study. The first concerned "lack of" blind ratings, and the second, questions about the sample size. In case other readers share this concern, we will describe the steps that we took to ensure objective assessments for each of the measures that we obtained, and we will address the issue of sample size.

First, all measures of brain functioning, including electroencephalograms (EEGs) and

The BEIP Core Group consists of: Charles H. Zeanah, Anna T. Smyke, and Sebastian F. Koga (Tulane University); Charles A. Nelson (University of Minnesota); Nathan A. Fox (University of Maryland), Peter J. Marshall (Temple University); Susan Parker (Randolph Macon College), and Hermi R. Woodward (University of Pittsburgh/MacArthur Research Networks). Direct correspondence to: Charles H. Zeanah, M.D., Department of Psychiatry, Tulane University Health Sciences Center, Tidewater Building TB-52, 1440 Canal Street, New Orleans, LA 70112.

INFANT MENTAL HEALTH JOURNAL, Vol. 27(6), 581–583 (2006)

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/imhj.20117

event related potentials (ERPs), had no opportunity for bias, as these electrical recordings were obtained completely independent of any human observer; the data were then transmitted to the University of Maryland, where they were analyzed. Second, all ratings of observed behavior during laboratory procedures, including assessments of attachment, emotional expression, language, and interactional behavior, were coded by raters who were unaware of group membership of the children, and in some cases, unaware of the study design. They were as blind as any coders of any study using behavioral assessments could be. Most susceptible to bias are ratings using the Observational Record of the Caregiving Environment (ORCE) of children in their naturalistic environments because we could not disguise that some children were in homes and some children were in group settings (though it is far from clear that they would have been able to determine that it was an institution rather than a child care setting). The trained raters who coded these videotapes were told only that they would be coding young children in homes and in group care settings. Third, height, weight, and head circumferences, as well as Bayley exams, were conducted by research assistants (RAs) who were not blind to group assignment. On the other hand, this is equally true of all studies using such measures in which children's status is apparent (e.g., children with Down Syndrome or Fetal Alcohol Syndrome). Fourth, caregiver report data do not lend themselves to expert coder bias, but caregiver bias is a widely recognized issue that we have considered in all publications involving those data, endeavoring where possible to include objective as well as subjective assessments of constructs and examining their convergence and divergence.

When bias is a possibility in coding data, it is important to consider the pattern of findings across all measures and to determine whether those completely removed from the possibility of bias yielded results similar to those that have more possibility of bias. We are satisfied that the results obtained so far from the BEIP are similar across all types of measures.

As for sample size, we consistently demonstrate in all our publications that we have adequate numbers of participants for the questions being addressed. We would have welcomed more participants, of course, but perhaps it is worth noting that the proper metric is not the biomedical studies to which Dr. Wassenaar alludes, but studies of the behavior of young children in institutions like this one. In fact, as best we have been able to determine, this is the largest study ever conducted on young children living in institutions.

With regard to the risk benefit ratio, Dr. Wassenaar raised several concerns. First, he appreciates that limited resources made it impossible to provide the intervention to all participants in the event that the intervention proved effective, but he expressed concerns that the investigators may not have advocated on behalf of the children and that "…67 children, subjected to batteries of physical and psychosocial evaluations for over four years, languish in these institutions to this day."

In fact, this is not the case. Of the 67 children originally in our institutional group, only 12 remain in institutions as of this writing. The other 55 have been adopted, returned to their families of origin, or placed in government sponsored foster care that did not exist at the outset of the study. Romania has embraced the use of foster care far more than they did when we began our work in 1999, and we followed scrupulously the dictates of Romanian law and the agreements we made originally with child protection officials about not interfering with decisions made about individual cases. This did not preclude us from advocating vigorously for children's safety, as in some cases in which return of children to potentially risky settings was contemplated.

Furthermore, our investigative team and Romanian partners planned, organized, and ob-

tained funding to support two national conferences in Bucharest. In addition to providing continuing professional education, we also presented preliminary findings from our study to the biomedical, neuroscience, and social services communities of Romania. Over 300 participants from all over Romania attended these conferences, as did the Romanian press. Finally, as a legacy of the project, we obtained additional support for 3 years of pilot funding to initiate the Institute of Child Development in Bucharest. This entity supports research, training, and services for young children with developmental delays, behavior problems, and/or social disadvantages.

Dr. Wassenaar also raised questions about the adequacy of screening of foster parents. The screening of foster parents in Romania is similar to that in the U.S. Prospective foster parents apply, document employment and income, complete interviews and home visits, and undergo criminal background checks. Following this, they receive training that was designed by Romanians but based upon training of foster parents in the U.S. Furthermore, they are carefully and systematically monitored by our social work staff. A difference from foster care in the U.S. is that foster parents are full-time employees who are paid a salary rather than a subsidy per child.

Finally, Dr. Wassenaar speculated that staff morale in institutions might have deteriorated as a result of the study, and that this might have contributed to poorer quality caregiving and lower performance in the control (institutionalized) group on the measures of interest. Since we did not assess morale of institution staff, we cannot directly refute this concern, although it is worth noting that with a nearly 50% drop in census of children on the units following foster care placement, staff morale might well have increased. In fact, the trend across all of the measures in the study among children still institutionalized is towards more favorable scores in follow-up assessments after baseline—exactly the opposite of what Dr. Wassenaar's prediction would lead us to anticipate.

For us, the major unresolved second guessing about the BEIP from an ethical standpoint concerns whether the study should have attempted an intervention on the institution side. Comparing foster care to enhanced institutional care was not the question posed to us by Romanian government officials originally, however, nor was it the question most relevant from a Romanian policy perspective. Nevertheless, comparing our model of foster care to some form of enhanced institutional care would have likely reduced the ethical concerns about studying "business as usual" institutional care, an intervention for young children in which the investigators had little confidence.