Published in final edited form as:

Dev Psychol. 2010 November ; 46(6): 1767–1770. doi:10.1037/a0021293.

Complexity, usefulness, and optimality: A response to Foster (2010)

Keith F. Widaman¹, Shannon J. Dogan², Gary D. Stockdale³, and Rand D. Conger³

¹ Department of Psychology, University of California, Davis, Davis, CA, US

kfwidaman@ucdavis.edu

² University of California Cooperative Extension, CA, US

³ Family Research Group, Department of Human and Community Development, University of California, Davis, Davis, CA, US

Abstract

In his commentary, Foster (2010) made arguments at 2 levels, offering a broad critique of statistical or methodological approaches in developmental psychology in general together with critical comments that applied only to our recent article (Dogan, Stockdale, Widaman, & Conger, 2010). Certain criticisms by Foster aimed at the field as a whole appear to be justified, whereas others seem overly broad and of dubious validity. In addition, Foster ignored the full range of methodologies used by both developmental psychologists and economists to pursue the identification of causal processes. Other critical remarks by Foster were directed specifically at our article, and many of these are simply incorrect, reflecting Foster's failure to recognize the standards in developmental psychology or his failure to note specific comments or descriptions we provided in our article. Future exchanges regarding methodological innovations and priorities in developmental psychology and economics should enrich and inform one another, rather than taking the form of one field dictating to the other the correct way to pursue science.

Keywords

causal inference complexity developmental research usefulness causality methodology empirical work developmental psychology

American Psychological Association

In his commentary, Foster (2010) made arguments at 2 levels, offering a broad critique of statistical or methodological approaches in developmental psychology in general together

Copyright © 2014 ProQuest LLC. All rights reserved.

Keith F. Widaman, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616
kfwidaman@ucdavis.edu.

Full Text: Developmental Psychology 0012-1649 1939-0599 American Psychological Association dev_46_6_1767 10.1037/a0021293
2010-22712-008 Response to Commentary Complexity, Usefulness, and Optimality: A Response to Foster (2010) Cynthia García Coll
Editor Keith F. Widaman Shannon J. Dogan Gary D. Stockdale Rand D. Conger Department of Psychology, University of California,
Davis University of California Cooperative Extension Family Research Group, Department of Human and Community Development,
University of California, Davis

with critical comments that applied only to our recent article (Dogan, Stockdale, Widaman, & Conger, 2010). Certain criticisms by Foster aimed at the field as a whole appear to be justified, whereas others seem overly broad and of dubious validity. In addition, Foster ignored the full range of methodologies used by both developmental psychologists and economists to pursue the identification of causal processes. Other critical remarks by Foster were directed specifically at our article, and many of these are simply incorrect, reflecting Foster's failure to recognize the standards in developmental psychology or his failure to note specific comments or descriptions we provided in our article. Future exchanges regarding methodological innovations and priorities in developmental psychology and economics should enrich and inform one another, rather than taking the form of one field dictating to the other the correct way to pursue science.

developmental research complexity usefulness causal inference

In his commentary, Foster (2010) melded a critique of statistical and methodological techniques used in the field of developmental research in general with a critical review of methods used in our article (Dogan, Stockdale, Widaman, & Conger, 2010). We thank Foster (2010) for his supportive comments on our research article, such as “this article is among the best submitted to *Developmental Psychology* this year, and it is being published for that reason” (p. 1765). If the entire commentary had been as laudatory, we would have had no occasion to write a response. Instead, Foster also noted that our “article meets the standards of the field... [but] also illustrates some of the problems I have encountered over time, and these problems are my focus here” (p. 1760). He then argued that even good articles in the field of developmental psychology typically do not meet his standards for scientific usefulness.

In this response, we discuss several major points raised by Foster (2010), noting our level of agreement or disagreement with his contentions; unfortunately, space limitations do not allow us to discuss all of his criticisms in detail. In particular, we intend to correct several misconceptions Foster maintains about our article specifically and, in our opinion, about the field of developmental psychology as a whole. Foster admitted that, as an economist, he has blind spots about our field but stated that he felt these would enable him to identify more easily problems in developmental psychology. We aim to illuminate some of these blind spots so Foster can gain a more informed understanding of our field, even as we hope to learn from his portrayal of current developments in economics.

Causal Inference

Foster (2010) argued that empirical work in developmental psychology appeared to have three goals: (a) describing developmental phenomena, (b) testing developmental theories, and (c) evaluating interventions based on our theories. For Foster, an important component of the second goal is to generate causal inferences regarding relations among predictor and outcome variables with an eye toward designing more effective intervention programs. In economics, researchers usually attempt to identify causal relations so that manipulating some aspect of the economy might yield predicted gains. Thus, as an economist, Foster felt

that developmentalists should often focus on the framing and testing of causal hypotheses that might lead to more successful interventions.

Most, if not all, areas of science probe the causal structure of phenomena within their purview. Along with most commentators on the topic, we agree that correlational research will never support causal conclusions as strong as those that can be drawn from experimental research involving manipulation of variables (i.e., random assignment of participants to conditions). However, many topics in developmental psychology simply cannot be studied through random assignment of participants to conditions. We could not randomly assign our participants to specified levels of alcohol consumption and then wait 1 or 2 years to assess changes in the number of sexual partners that resulted from this manipulation. Instead, our investigation involved self-selection by participants of their levels of alcohol consumption and number of sexual partners. We then sought to understand how these choices unfolded year-by-year across a 13-year period, hoping to determine whether symmetric or asymmetric lead-lag relations, if not firm causal effects, held between the two variables of primary interest.

Our statistical approach used a version of lagged regression effects, or what is termed *Granger causality* in the economics literature. The core of our model would typically be described as a form of cross-lagged regression (cf. Cole & Maxwell, 2003) to enable us to test whether either process had lagged effects on the other. Granger causality may be somewhat passé in certain circles, as newer, more complex methods have been developed. However, cross-lagged effects are a step on a continuum toward more sophisticated evaluations. Our study was designed as an initial test of lead-lag relations between alcohol consumption and sexual behavior from adolescence into adulthood, with associated tests involving covariates to test whether the lagged effects we identified were only the biased or confounded effects of outside variables. It is especially important that we knew of no prior studies that collected data as many times (eight) across such an extended period (from 15 to 28 years of age) on the core variables for our study.

Various methods have been developed to adjust correlation or regression analyses of nonexperimental data to give a more accurate picture of the influence of potential causative effects. These methods include covariate adjustment, propensity score matching, and the use of instrumental variables, among others. Covariate adjustment is commonly used in psychology, propensity score matching has been developed primarily in statistics, and instrumental variable methods are commonly used in economics. Each of these methods has its strengths, but each has its limitations as well. None of these methods leads to strong causal inferences of the sort justified by experimental studies, but each can lead to reduced bias in estimation of effects.

The choice among adjustment methods can resemble opting for the flavor of the month (or decade). For example, a great deal of work on propensity score matching, an approach designed to improve causal inference from nonexperimental studies, has been conducted over the past 2 decades and more (Rosenbaum, 2002; Rubin, 2001). However, Shadish, Clark, and Steiner (2008) recently showed that covariate adjustment led to greater reduction in bias in a nonexperimental study than did propensity score matching. Although this

conclusion was not without its critics (see Hill, 2008; Little, Long, & Lin, 2008; Rubin, 2008), bias reduction methods designed to support stronger causal conclusions wax and wane in popularity. However, as Foster and Kalil (2007) noted in their observational study of children in low-income families, even contemporary statistical techniques designed to reduce bias in estimation do “not control for all potentially confounding factors” (p. 1664).

From our perspective, this comment is at the heart of a basic disciplinary divergence between economists and psychologists more generally. In Foster's (2010) Figure 1, he suggested that with representative samples and appropriate statistical complexity, a researcher may be able to draw causal inferences from observational data. Those publishing in psychology journals are well aware that attempts to draw causal conclusions from observational data are not well received in the discipline inasmuch as one can never know whether all possible confounds have been controlled in the analyses (see Shadish, Cook, & Campbell, 2002). Nevertheless, we do not believe this is reason for despair. In some instances, theoretical issues addressed in observational studies can also be investigated with randomized experiments, and replication across both forms of research provides an even stronger basis for causal inference. In fact, repeated replications of what Foster would *call predictive studies with observational data* would also increase confidence in a possible causal connection between a predictor and a developmental outcome and provide important information about possible preventive interventions, which also can be used as a theoretical test through randomized experiments (Conger, Lorenz, & Wickrama, 2004). We contend that, ultimately, causal inferences are weak at best without replication across different studies, different populations, and different research designs. No single study, no matter how representative or statistically complex, provides firm grounds for inferring causality.

In terms of our study that led to Foster's (2010) critique, at the behest of reviewers, we included a fairly wide range of covariates to see if these led to a reduction in the magnitude of the effect of alcohol consumption on later sexual behavior (Dogan et al., 2010). This method of bias reduction is similar to the covariate adjustment methods used by Shadish et al. (2008) and is commonly used in developmental psychology. Simply put, significant effects of alcohol consumption on later sexual behavior remained in the presence of estimated effects of covariates, so these effects appeared generally to survive tests involving alternative, third-variable explanations. We took care not to argue that we had arrived at strong causal conclusions from our research, but our results did show that lagged effects of alcohol consumption on later changes in sexual behavior were not reduced appreciably by the covariates we examined.

Complexity Versus Usefulness Versus Optimality

One primary criticism that Foster (2010) leveled at the entire field of developmental psychology—heralded by the title of his commentary—was that articles published in our leading journals were overly complex and distinctly lacking in usefulness, at least sufficient usefulness given the complexity of our statistical analyses. We have grudging agreement with aspects of this contention. However, rather than arguing for either simpler or more complex methods of analysis, we feel that researchers should use optimal methods of analysis. Optimal statistical methods are the simplest ones that answer the questions posed

in a research project, consistent with sentiments by the APA Task Force on Statistical Inference (Wilkinson & the Task Force on Statistical Inference, 1999). At times, the simplest methods to answer a question will seem complex to some arbiters. Whereas we agree with Foster that psychologists should strive for simplicity in statistical models and analytic techniques, we feel that optimality should trump a one-sided push toward simplicity.

With regard to our article (Dogan et al., 2010), we had a certain analytic complexity thrust upon us. The first submission of our manuscript reported results based on cross-lagged regression models. To meet suggestions for revision made by Foster and other reviewers of our manuscript, we increased the complexity of our analyses in two major ways: (a) switching to a dual latent change score model (McArdle & Hamagami, 2001) that could represent and test more directly changes in behavior from one time point to the next, including the ability to model mean changes as well as changes in interindividual differences, and (b) including a number of covariates that constituted *third variables*, or alternative explanations of the lagged effects we had found. Charging our study with too great a level of complexity seems disingenuous if the charge is made by someone who had a hand in augmenting that complexity.

On a related point, Foster (2010) complained that our “list of covariates used seems partial at best” (p. 1764), noting that we did not include socioeconomic indicators or measures of school or peers as covariates. We agree to some extent with this statement but note that we included all covariates that were suggested during three rounds of reviews of our article. In retrospect, we wish we had included additional covariates mentioned by Foster. However, including additional covariates would only have increased the complexity of our analyses, and a complaint that a study considered only a limited number of covariates often amounts to an offhand comment that can be directed toward any study that uses covariates to reduce bias. Such a criticism gains import if supplemented by a cogent argument for how a missing covariate might account for certain findings. We sought as covariates psychological third variables (e.g., excitement seeking) that were easily construed as potential common causes of both alcohol use and number of sexual partners. Economists might think that socioeconomic status accounts for all important behavioral phenomena. However, absent a reasoned account for how a third variable like socioeconomic status might function as a common cause, the force of this criticism is reduced. Indeed, in 30 years of controlling for socioeconomic status in studies of the association between social processes and individual adjustment, we have never found that it significantly changed results in any of our investigations.

Sample Selection

Foster (2010) argued that developmental researchers often do not attend closely to sample selection. This often leads to samples of convenience that are not representative of any specifiable population, rendering results of such studies of unclear external validity or generalizability. We agree strongly that issues of selection are paramount in the design and execution of a study and that use of unrepresentative samples can compromise conclusions. We agree, basically, with the contention by Foster that “needlessly complex analyses of unrepresentative data plumb the depths of usefulness” (pp. 1762).

On the other hand, we are unsure of the force of this criticism in relation to our article (Dogan et al., 2010). Our study was based on a sample that was carefully drawn to be representative of rural Iowa where the study took place. A high percentage of families eligible for the study participated, a remarkably low attrition rate has occurred during the 20 years since the study began, and the sample size was relatively large. Although not nationally representative, the study sample was representative of the population of persons living in rural Iowa. Given this representativeness, we presume that results of the study are likely to generalize to rural areas in many parts of the United States and certainly offer a strong basis for comparison with results that might be obtained in urban settings across our country. Indeed, findings from this research have tended to replicate with urban, rural, and ethnic minority populations (e.g., Conger, Conger, & Martin, 2010).

Linearity

Foster (2010) complained that much developmental research considers only linear models, eschewing consideration of nonlinear trends between variables. We agree, and one of us has argued that linear models for age effects are unlikely to be of much utility across broad swaths of the age continuum (Widaman, 2007). However, over more restricted age levels, local trends are likely to be quite linear, and replicable nonlinear trends might require extremely large sample sizes to estimate reliably. Because our analytic model estimated rather local relations among variables, linearity was likely. Yet, the model was able also to estimate the longer term nonlinearities in average levels of alcohol consumption and number of sexual partners. Again, the force of this criticism is unclear in relation to our article.

Model Specification and Estimation

Model specification and estimation of parameters was another topic on which Foster (2010) had pointed criticisms. We agree with much that Foster said, concurring that model specification should be carefully described and that estimation of parameters should consider the nature of the data to ensure optimal outcomes. Foster bemoaned the “low level of mathematical literacy” (p. 1763) in the field of developmental psychology; we also would like to see improvements in this area, although we would not describe the current state of our field in such pejorative terms. Foster argued that standard specifications, such as homogeneous error variances, should be tested, because error variances may increase with age in concert with increases in variance on true score variates. We agree with this as well, and current research by Grimm and Widaman (2010) documents how parameter estimates associated with intercept and slope latent variables in latent growth models can be greatly simplified with increased complexity of error structures.

However, we dispute several contentions by Foster (2010) with regard to our models and potential alternatives. Foster chided us because we provided figures of our models, not a series of equations. This was our choice, a common one in our field. However, researchers have understood for over 30 years that, if figural models are sufficiently detailed, figural models and mathematical equations are functionally isomorphic—the figural model implies a set of equations and vice versa. Foster also claimed that we had assumed away “any shared determinants of the two processes at a point in time except in the last period” (p. 1764)

because we allowed a covariance between error terms only at the last time of measurement. This is incorrect. We showed such a covariance at the last time of measurement but stated that analogous covariances at the other times of measurement were estimated but were omitted so the figure would not be too cluttered.

In complaining about complexity, Foster (2010) often lumped together latent variable and latent group analyses. We agree that latent group or mixture models rely heavily on assumptions that may not hold (cf. Bauer & Curran, 2003), but latent variable models are much less suspect. Given levels of error variance in manifest variables in many research studies, latent variable modeling should not be disdained, because it provides the best hedge against biasing effects of unreliability on parameters relating to theoretical constructs. Indeed, as microeconomists become increasingly interested in psychological and biological phenomena in attempts to understand the underpinnings of economic behavior, they may find great utility in latent variable approaches. Foster also maligned developmental research by stating that our field still relies on “statistical significance as an either–or alternative” (p. 1764). Conversely, we feel that our field is moving away from such a shallow interpretation of results, supplanting either–or decisions based on significance tests with a consideration of the magnitude of parameter estimates, their standard errors, and their resulting confidence intervals (see Wilkinson & the Task Force on Statistical Inference, 1999). Further, a careful reading of our article (Dogan et al., 2010) shows that we emphasized the magnitude of estimated effects before and after controlling for third variables, clearly not using the either–or approach to evaluating statistical significance.

Finally, Foster (2010) argued that “a set of separate correlation matrices involving all the key constructs at each time point” (p. 1765) would be a simpler and more direct way to represent and test our hypotheses about relations between alcohol consumption and sexual behavior. However, our hypotheses dealt most importantly with lagged effects between the two key behaviors, with these lagged effects estimated as regression coefficients. We cannot conceive of any way in which inspection of correlations among constructs could ever offer compelling support for cross-time lagged effects. Indeed, Rogosa (1980) offered a trenchant criticism of cross-lagged correlations if lagged regression effects in panel models generated the data; in such situations, lagged regression procedures captured the process generating the data, whereas cross-lagged correlations often led to incorrect conclusions. In pushing for maximal simplicity, we think Foster lost all focus on the theoretical conjectures that motivated our study and how they had to be estimated and tested in an optimal fashion.

Summary

In closing, we argue that optimality in research design and analysis should be the goal, rather than a single-minded push for greater simplicity. Optimality applies to all aspects of a research study—selection of participants, selection of materials and procedures, collection of data, and design and execution of analyses. The greater the optimality on all these dimensions, the more useful the results of a study are likely to be. Usefulness should be the ultimate goal, and optimality—not simplicity—appears to be the best way to ensure maximal usefulness.

Empirical research in psychology, economics, and other social and behavioral sciences would be improved if disciplinary boundaries were erased and greater communication of state-of-the-art developments in each area of study occurred. We agree that developmental psychology can benefit from current methodological developments in economics, but we feel just as ardently that the field of economics could benefit from a closer, more careful reading of methodological developments in psychology and in developmental psychology in particular. We look forward to a time and place when the free and informative exchange of ideas across disciplines can happen in a forum in which more light than heat is generated.

Acknowledgments

Support for this work was provided by grants from the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, and the National Institute of Mental Health (DAO17902, HD047573, HD051746, and MH051361).

Grant/sponsorship: Sponsor: Eunice Kennedy Shriver National Institute of Child Health and Human Development; Recipient: No recipient indicated; Sponsor: National Institute on Drug Abuse; Recipient: No recipient indicated; Sponsor: National Institute of Mental Health; Grant number: DAO17902, HD047573, HD051746, MH051361; Recipient: No recipient indicated;

References

- Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*. 2003; 8:338–363. doi: 10.1037/1082-989X.8.3.338. [PubMed: 14596495]
- Cole DA, Maxwell SE. Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*. 2003; 112:558–577. doi: 10.1037/0021-843X.112.4.558. [PubMed: 14674869]
- Conger RD, Conger KJ, Martin MJ. Socioeconomic status, family processes, and individual development. *Journal of Marriage and Family*. 2010; 72:685–704. [PubMed: 20676350]
- Conger, RD.; Lorenz, FO.; Wickrama, KAS., editors. *Continuity and change in family relations: Theory, methods, and empirical findings*. Erlbaum; Mahwah, NJ: 2004.
- Dogan SJ, Stockdale GD, Widaman KF, Conger RD. Developmental relations and patterns of change between alcohol use and number of sexual partners from adolescence through adulthood. *Developmental Psychology*. 2010; 46:1747–1759. doi: 10.1037/a0019655. [PubMed: 20677862]
- Foster EM. The U-shaped relationship between complexity and usefulness: A commentary. *Developmental Psychology*. 2010; 46:1760–1766. doi:10.1037/a0020180. [PubMed: 21058835]
- Foster EM, Kalil A. Living arrangements and children's development in low-income White, Black, and Latino families. *Child Development*. 2007; 78:1657–1674. doi:10.1111/j.1467-8624.2007.01091.x. [PubMed: 17988313]
- Grimm KJ, Widaman KF. Residual structures in latent growth curve modeling. *Structural Equation Modeling*. 2010; 17:424–442.
- Hill J. Comment. *Journal of the American Statistical Association*. 2008; 103:1346–1350. doi: 10.1198/016214508000001002.
- Little RJ, Long Q, Lin X. Comment. *Journal of the American Statistical Association*. 2008; 103:1344–1346. doi:10.1198/016214508000000995.
- McArdle, JJ.; Hamagami, F. Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data.. In: Collins, LM.; Sayer, AG., editors. *New methods for the analysis of change*. American Psychological Association; Washington, DC: 2001. p. 139-175. doi: 10.1037/10409-005
- Rogosa D. A critique of cross-lagged correlation. *Psychological Bulletin*. 1980; 88:245–258. doi: 10.1037/0033-2909.88.2.245.
- Rosenbaum, PR. *Observational studies*. 2nd ed.. Springer-Verlag; New York, NY: 2002.

- Rubin DB. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*. 2001; 2:169–188. doi: 10.1023/A:1020363010465.
- Rubin DB. Comment: The design and analysis of gold standard randomized experiments. *Journal of the American Statistical Association*. 2008; 103:1350–1353. doi: 10.1198/016214508000001011.
- Shadish WR, Clark MH, Steiner PM. Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*. 2008; 103:1334–1343. doi: 10.1198/016214508000000733.
- Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin; New York, NY: 2002.
- Widaman, KF. Intrauterine environment affects infant and child intellectual outcomes: Environment as direct effect.. In: Little, TD.; Bovaird, JA.; Card, NA., editors. *Modeling contextual effects in longitudinal studies*. Erlbaum; Mahwah, NJ: 2007. p. 387-436.
- Wilkinson L, the Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*. 1999; 54:594–604. doi:10.1037/0003-066X.54.8.594.