Models for analysis of longitudinal data

To the Editor:

In his editorial comment on the paper by SHERRILL et al. [1] WARE [2] discusses the use of efficient statistical models for longitudinal data, and how improved efficiency and validity results from proper model specification.

He fails to point out, however, that these models are a special case of a more general class of "multilevel" models, which explicitly take account of hierarchically structured data. In longitudinal repeated measures designs we have a basic two-level hierarchy, where measurement occasions are nested within subjects. In fact, however, the paper by SHERRILL et al. [1] analyses data which come from a three-level hierarchy, since individuals are further nested within households. Since the SHERRILL et al. [1] model ignores this third level, it is incomplete and is potentially invalid if there is a sizeable between-household variation in respiratory symptoms.

General models for hierarchically structured data have been studied by a number of authors, not mentioned by WARE [2], and general purpose software packages are widely available. In particular, the model studied by SHERRILL et al. [1] together with the extension to incorporate between-household variation as well as general continuous time autoregressive structures can be fitted using one of these packages, ML3 [3]. This package has been successfully used to describe the development of maximum oxygen consumption in young athletes [4]. The hierarchical model was used to assess the effects of training by separating this effect from those of normal growth and development.

References

A. Baxter-Jones, P. Helms, Dept of Child Health, University of Aberdeen, Scotland, UK.

From the author:

The comments of Goldstein and colleagues provide a useful addition to the discussion of the paper by SHERRILL et al. [1]. Those working to develop methods for the analysis of longitudinal data have tended to ignore other types of clustering in the data, in part because within-subject correlation tends to be much larger than correlation from other types of clustering. For example, intrafamily correlation of pulmonary function measurements will be much smaller than within-subject correlation. Nevertheless, it would be helpful to have the capacity to explore these issues empirically. In the past, efforts to do so have been handicapped by the limitations of available software. Perhaps the next generation of software, including ML3, will be more flexible.

J.H. Ware
Harvard School of Public Health, Boston, USA.