Web-based Linear Regression Analysis of Remote Datasets
Mark G. Weiner, M.D.,1,2 Abigail Cohen, Ph.D.,1 Meena Seshamani,
Alice Livshits,1 Alan Hillman, M.D., M.B.A.1,2
1Division of General Internal Medicine, 2Leonard Davis Institute of Health Economics
University of Pennsylvania School of Medicine, Philadelphia, PA

To promote improved communication of research results, novel methods of information dissemination must be developed which leverage the accessibility of the World Wide Web. This poster describes the development and application of a modular, Web-based tool which enables investigators to easily access raw data from remote sites, and perform their own interactive statistical analyses to corroborate the original investigators' findings, or to stimulate new hypothesis generation and further research. The current analyses enabled by the resource include descriptive statistics of selected variables, and multiple linear regression. We describe the development of this resource and its successful application on the results of a clinical trial designed to correlate a functional status assessment of patients using the Short Form-12 with productivity and financial outcomes.

METHODS

A. Clinical Trial
The Short Form-12 (SF-12) is a validated twelve-question survey instrument that measures current health-related quality of life. In general, higher scores are more desirable, and indicate a better state of health. Although work productivity is often assumed to be associated with better health, we sought to demonstrate the ability of an SF-12 score to predict different degrees of productivity loss. We administered the SF-12 survey, augmented with 4 questions pertaining to days lost or cut down from usual work activities to patients presenting to a busy general medicine clinic.

B. Database/Interface Development
The data from this clinical trial were integrated with a growing warehouse of information covering the population served by the University of Pennsylvania Health System. All the data exist centrally on a DEC AlphaServer, on a UNIX platform running Oracle 7.3. Web access to the data is achieved using an Apache Web Server. The graphical user interface for both data querying and analysis is constructed entirely in Java. To mask the underlying database complexity from the user during the query process, a modifiable metadata table supplies the interface with information regarding the database under analysis. The linear regression algorithms were programmed based upon matrix formulas supplied by Neter, et al.13

RESULTS

Utilizing an iterative process enabled by the interface, an investigator can quickly designate any number of independent variables such as SF-12 score, gender, or age and determine the relationship to productivity and other dependent variables using linear regression. The output of the linear regression includes the calculation of the beta coefficients for the variables, including their standard error and p-values. Additionally, an Analysis of Variance is performed. The user can also view scatterplots and histograms of the data to assess its adherence to the conditions of the linear model.

With easier access, and no requirement for proprietary software, the web resource produces results similar to statistical packages. The relationship between SF-12 scores and decreased productivity days is statistically significant. However, the analysis suggests there are clearly additional unmeasured factors that contribute to the number of days a patient misses work.4

DISCUSSION

This paper proposes a Web-based, user friendly mechanism to access databases and perform meaningful statistical analysis on any desired combination of variables in the data set. The metadata table can be modified to generalize the resource to query and analyze the results of other large-scale clinical trials, or prospectively collected data sets.

Acknowledgement
The clinical research project described in this paper was supported by a grant from the University of Pennsylvania Research Foundation.

References
2 http://www.med.upenn.edu/weiner/newreg/reg.html.