

Combining information from multiple data sources to create multi-variable risk models: Illustration and preliminary assessment of a new method. *Journal of Biomedicine & Biotechnology*, 2:113-123 (2005).

ABSTRACT

A common practice of meta-analysis is combining the results of numerous studies on the effects of a risk factor on a disease outcome. If several of these composite relative risks are estimated from the medical literature for a specific disease, they cannot be combined in a multivariate risk model, as is often done in individual studies, because methods are not available to overcome the issues of risk factor colinearity and heterogeneity of the different cohorts. We propose a solution to these problems for general linear regression of continuous outcomes using a simple example of combining two independent variables from two sources in estimating a joint outcome. We demonstrate that when explicitly modifying the underlying data characteristics (correlation coefficients, standard deviations, and univariate betas) over a wide range, the predicted outcomes remain reasonable estimates of empirically derived outcomes (gold standard). This method shows the most promise in situations where the primary interest is in generating predicted values as when identifying a high-risk group of individuals. The resulting partial regression coefficients are less robust than the predicted values.