SANJAY SHETE, PhD
Department of Biostatistics
University of Texas MD Anderson Cancer Center

Identifying SNPs associated with Mediators in Genomewide Association Studies:
Application to Smoking Behavior and Lung Cancer

A mediation model explores the direct and indirect effects between an independent variable and a
dependent variable by including other variables (or mediators). Mediation analysis has recently been used
to dissect the direct and indirect effects of genetic variants on complex diseases using case-control
studies. However, bias could arise in the estimations of the genetic variant-mediator association because
the presence or absence of the mediator in the study samples is not sampled following the principles of
case-control study design. In this case, the mediation analysis using data from case-control studies might
lead to biased estimates of coefficients and indirect effects. In this article, we investigated a multiple-
mediation model involving a three-path mediating effect through two mediators using case-control study
data. We propose an approach to correct bias in coefficients and provide accurate estimates of the
specific indirect effects. Our approach can also be used when the original case-control study is frequency
matched on one of the mediators. We conducted simulation studies to investigate the performance of the
proposed approach, and showed that it provides more accurate estimates of the indirect effects as well as
the percent mediated than standard regressions. We then applied this approach to the multiple-mediation
study of the mediating effects of both smoking and chronic obstructive pulmonary disease (COPD) on the
association between the CHRNA5-A3 gene locus and lung cancer risk using data from a lung cancer
case-control study. The results showed that the genetic variant influences lung cancer risk indirectly
through all three different pathways. The percent of genetic association mediated was 18.3% through
smoking alone, 30.2% through COPD alone, and 20.6% through the path including both smoking and
COPD, and the total genetic variant-lung cancer association explained by the two mediators was 69.1%.

Wednesday, November 28, 2012
Time: 3:30-4:30 PM
Place: 1301 McGavran-Greenberg