Appendix B – Search space analysis examples

Questions = Number of questions at issue = outcomes × predictors × lags
Models = Number of models = 2^k where k = number of covariates
Search space = Approximation of analysis search space (number of statistical tests) = Questions × Models

Assume the following conditions in the three examples below, used differently in each example:

- prospective cohort studies of air pollution and respiratory health with sufficiently large sample sizes of children enrolled in elementary grade 4 and followed for 8 years
- one outcome is of interest – asthma diagnosis later in life (surveyed annually during grade 5 until high school graduation – 8 years) due to outdoor air quality predictors
- three outdoor air quality predictors are of interest – PM2.5, NO2 and O3
- an 'asthma diagnosis' is based on responses to annual questionnaires where a subject in the cohort answers yes to the following question… "*Has a doctor ever said you had asthma?*"
- questionnaires are administered by trained professionals once each year to each family of an enlisted child for up to 8 years of follow-up until high school graduation (i.e., 8 lags)
- covariate confounders are considered related to four socioeconomic measures – median household income, proportion of subjects with no high school diploma, percent of males unemployed, percent living in poverty
- covariate confounders are considered related to two meteorological variables – humidity and temperature
- air quality parameters (PM2.5, NO2 and O3) are also be treated as covariate confounders

Example 1

Assume a simple epidemiological investigation of childhood asthma diagnosis using the 3 air quality predictors – daily average levels of PM2.5, NO2 and O3, 8 years of follow-up and no covariate confounders:

- Questions = 1 outcome × 3 predictors × 8 lags = 24
- Models = 1 (i.e., no consideration of covariate confounding)
- Search space = 24 × 1 = 24

Example 2

Assume an epidemiological investigation of the same 3 predictors, 8 years of follow-up, 2 weather variables treated as covariate confounders (annual average temperature and annual average humidity), 4 socioeconomic measures also treated as covariate confounders:

- Questions = 1 outcome × 3 predictors × 8 lags = 24
- Models* = 2^{2+4} = 64
- Search space = 24 × 64 = 1,536

* Note: there are 6 covariates in Models – 2 weather variables and 4 socioeconomic measures.

Example 3

Assume an epidemiological analysis of the same 3 predictors, 8 years of follow-up now using 4 weather variables treated as covariate confounders (average temperature in winter and summer months, average humidity in winter and summer months), plus the 4 socioeconomic measures treated as covariate confounders, and using air quality parameters also treated as covariate confounders:

- Questions = 1 outcome × 3 predictors × 8 lags = 24
- Models^ = 2^{4+4+2} = 1,024
- Search space = 24 × 1,024 = 24,576

^ Note: there are 10 covariates in Models – 4 weather variables, 4 socioeconomic measures; there is 1 air quality variable is treated as a predictor which is adjusted with the other 2 air quality variables as covariates.