collection (n = 130) and further sub categorized according to how data was collected. Primary data included: personal experiences, as described by the submission's author (n = 16); surveys conducted specifically for the submission (n=34); and new interviews of patients and family members on disease and drug experiences (n = 36). Half (forty-seven of ninety-three) of the patient input submissions included experiences of one or more patients who had received the drug under review. Secondary data included: published literature (n = 31); existing surveys (n = 27); past conversations with patients and family members (n = 36); experiences of patient group staff interacting with patients and family members (n = 19); and advice from clinical experts (n = 17). Many patient input submissions (sixty-eight out of ninety-three) reported multiple approaches to collect data. Use of two approaches was most common (thirty-seven out of ninety-three) with five or six approaches used in three of ninety-three submissions.

CONCLUSIONS:

Despite resource and timing challenges, many patient groups gather primary data to share with CADTH and find individuals with experience of the drug under review.

PD25 Principal Component Approximation: Medical Expenditure Panel Survey

AUTHORS:

Yi-Sheng Chao, Chao-Jung Wu (chaoyisheng@post.harvard.edu)

INTRODUCTION:

Principal component analysis (PCA) is important to summarize data or reduce dimensionality. However, one disadvantage of using PCA is the interpretability of the principal components (PCs), especially in a high-dimensional database. This study aims to analyze the patterns of variance accumulation according to PCA loadings and to approximate PCs with input variables from sample data sets.

METHODS:

There were three data sets of various sizes used to understand the performance of PC approximation: Hitters; SF-12v2 subset of the 2004 to 2011 Medical Expenditure Panel Survey (MEPS); and, the full set of 1996 to 2011 MEPS data. The variables in three data sets were first centered and scaled before PCA. PCs approximation was studied with two approaches. First, the PC loadings were squared to estimate the variance contribution by variables to PCs. The other method was to use forward-stepwise regression to approximate PCs with all input variables.

RESULTS:

The first few PCs represented large portions of total variances in each data set. Approximating PCs using stepwise regression could more efficiently identify the input variables that explain large portions of PC variances than approximating according to PCA loadings in three data sets. It required few numbers of variables to explain more than eighty percent of the PC variances.

CONCLUSIONS:

Approximating and interpreting PCs with stepwise regression is highly feasible. Approximating PCs can help i) interpret PCs with input variables, ii) understand the major sources of variances in data sets, iii) select unique sources of information and iv) search and rank input variables according to the proportions of PC variance explained. This is an approach to systematically understand databases and search for variables that are highly representative of databases.

PD26 Principal Component Approximation: Canadian Health Measures Survey

AUTHORS:

Yi-Sheng Chao (chaoyisheng@post.harvard.edu), Chao-Jung Wu

INTRODUCTION:

Principal component analysis (PCA) is used for dimension reduction and data summary. However, principal components (PCs) cannot be easily interpreted. To interpret PCs, this study compares two methods to approximate PCs. One uses the PCA loadings to understand how input variables are projected to PCs. The other uses forward-stepwise regression to determine the proportions of PC variances explained by input variables.

https://doi.org/10.1017/S0266462318003008 Published online by Cambridge University Press