Calculating Measures of Comorbidity Using Administrative Data

(Vicki Stagg, Dr. Robert Hilsden, Dr. Hude Quan)

BACKGROUND REFERENCES

Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis*. 1987;40(5):373-383.

Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epi*demiol. 1992;45(6):613-619

Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Medical Care 1998;36(1):8-27

Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Medical Care, 2005;43(11):1130-1139

PRESENTATION NOTES

Slide #1 -
I would like to share with you 2 programs – charlson & elixhauser - I've been writing under the supervision of **Dr. Robert Hilsden**, which incorporate the work of **Dr. Hude Quan**, both of the **University of Calgary.**
These programs are applicable to the medical field for calculating measures of comorbidity using administrative data.

Slide #2 –
I would first like to clarify some terms -
Medical *administrative data* are records of inpatient hospital visit information, which in Canada are used by provincial health care departments.

A ***comorbidity*** is an illness or disease which is NOT directly related to the patient's primary reason for admittance to hospital, but which has the potential to increase the likelihood of a poor outcome.

The significance of enumerating comorbidities and summarizing them into some kind of ***comorbidity index*** *or* simply *the total number of comorbidities* pertains to their role as risk factors in predicting mortality, measuring the burden of disease, and adjusting for the severity of disease overall (called "case-mix adjustment".)
So comorbidity measures are often included in statistical models for the purpose of stratification and adjustment.
Two *tools* commonly used to classify and define comorbidities were developed by Mary **Charlson** in 1987 at Cornell and later in 1998 by Anne **Elixhauser** in Maryland.

Finally, ***clinical conditions*** are summarised in the **International Classification of Disease** which assigns codes to every diagnosis. Modifications have resulted in what are called ICD-9-CM codes and ICD-10 codes.
Comorbidity coding algorithms associate the comorbid disease with the relevant clinical diagnostic codes.

Slide #3 –
I have developed 2 ado programs – **charlson** and **elixhauser**
The **charlson** method incorporates 17 comorbidity definitions, is weighted and contains 3 algorithms – the original ICD-9-CM developed by Dr. Richard **Deyo** at the University of Washington, and the Enhanced ICD-9-CM and ICD-10, both developed by Dr. Hude **Quan** of the University of Calgary.

The **elixhauser** method includes 30 comorbidites, is not weighted and includes the 2 algorithms, Enhanced ICD-9-CM and ICD-10, developed by Dr. Quan.

Dr. Quan has programmed all three algorithms into the SAS programming language (*Medical Care,* 2005;43(11):1130-1139).

Slide #4 -
3 international research groups were involved in the development of the ICD-10 coding algorithms; I am using the Canadian version.
The "Enhanced" ICD-9-CM algorithms were developed to improve earlier versions.

Slide #5 –
This table was taken from **Dr. Quan's** 2005 paper, which illustrates the 3 different coding algorithms for Charlson comorbidities.

Slide #6 –
This is an enlarged view of the previous table showing more clearly the 3 different definitions (algorithms) for 2 Charlson comorbidities, with the associated ICD codes.

Slide #7 –
This next table shows the weights assigned to each of the Charlson comorbidities, which reflect the seriousness of the disease. The weights were determined from the associated relative risk of death in one year.
Notice some comorbidities have 2 forms – mild and severe with different weights – in particular Diabetes, Liver disease and Cancer.

Slide #8 -
The input database can contain patient demographic data as well as the visit diagnostic codes or only the codes for each patient-visit, which must be alphanumeric. Additional information may also be included for subsequent modeling.

Slide #9 –

This slide gives the syntax for the 2 ado programs, charlson and elixhauser, which are clearly very similar, with the exception that charlson has an additional option called **assign0**, which I will be explaining.

Slide #10 –
The input options specify –
-the **algorithm** being run (i.e. **index c, e or 10**), which implies the type of input data (ICD-9-CM or ICD-10),
-the **id variable** if patients have multiple visits (so patient is the observational unit rather than visit)
-the root (**diagprfx**) of the name of the diagnostic code variables which have a suffix ranging from 1 to the maximum number of diagnoses. This option is not necessary if a *varlist* is provided.
-and **assign0** is the flag to apply the hierarchical method which assigns a weight of 0 to mild comorbidities when a patient also exhibits a more severe form of the diagnosis.

Slide #11 –
Output options control what summaries will be displayed, if any. A summary table of the calculated comorbidity measure can be displayed with or without summaries of the individual comorbidities.
noshow controls the messages output as the program runs.

Slide #12 –
I have constructed a sample dataset of ICD-9-CM codes to help illustrate the features of the charlson program. The highlighted lines are patient-visits with both the mild and severe form of certain diseases.
Eg: i*d3 has both uncomplicated(wt 1) & complicated diabetes(wt 2).*
id9 has both degrees of liver disease (wt 1 & w t 3)
id10 has two severities of types of cancer (wt 2 and wt 6)
id7 has both mod/sev liver dis (wt 3) and metastatic cancer(wt 6)

Slide # 13 –
This is the command to run the Enhanced algorithm on the sample ICD-9-CM data.

Slide # 14 –
I am showing here the messages output as the program runs, when the option **noshow** is omitted.

Slide #15 –
This is sample output. The top table is a summary of the Charlson index, which essentially is the frequency summary of the weighted sums of Charlson comorbidites represented in the data.
Notice 2 patient-visits have an index (sum) of 9 – (due to the weighting done)
These are id10 who had both severities of cancer as well as mild liver disease & id7 who presented with both moderate/severe liver disease & metastatic cancer.

Slide # 16 –
This illustrates the summaries of individual comorbidities, showing here the frequency of both severities of diabetes

Note there are 3 patient-visits with mild diabetes.

Slide #17 –
This describes the contents of the newly-created output dataset which contains not only the input data, but also the comorbidity indicator variables, the ….

Slide #18 –
….corresponding weight variables,
and the calculated Charlson index variable (weighted sum of comorbidities)
and a grouped form of the index.
It is important to note that when patient, rather than visit, is the observational unit only the data from the final hospital visit will be retained, if the visits have been ordered by date.

Slide #19 –
I selected a few variables to list – the comorbidity indicator variables for both forms of diabetes (mild & severe – ynch10 & ynch11), as well as for moderate/severe liver disease (ynch15), along with the corresponding weight variables.
Notice uncomplicated diabetes has a wt of 1, complicated diabetes has wt of 2 and moderate/severe liver disease has wt of 3 (as we saw before).

Slide #20 –
The calculated Charlson index is listed here, along with the grouped version, either of which can now be entered into a statistical model like logistic or multiple linear regression.
Note id3 has an index or weighted sum of 3, id9 has an index of 4 and id10 of 9.

Slide #21 –
The charlson program was rerun with the **assign0** option, which reduced the Charlson index of the patients who had both a mild and more severe form of a disease
– id3 (wt 3 to 2), id9 (wt 4 to 3), id10 (wt 9 to 7).
The Charlson index summary has frequencies which are different from before.
Eg.: there is only 1 patient with an index of 9; other changes occurred as well due to the new weights for id3, id9 and id10.

Slide #22 –
Notice the frequency of uncomplicated diabetes has changed from 3 to 2, because the uncomplicated diabetes diagnostic code of id3 was assigned a wt of 0.

Slide #23 –
I also ran the ICD-10 algorithm of the elixhauser program on real ICD-10 inpatient data with almost 3000 patients and with diagnostic code variables dx1 through to a maximum of dx25 clinical codes, along with other demographic and medical data.
Slide #24 –
This is the command to run the elixhauser ICD-10 algorithm.

Slide #25 –
This displays the summary of the calculated sum of elixhauser comorbidities, ranging from 0 to 10.

Slide #26 & 27 –

Finally, the individual Elixhauser comorbidities are listed here along with the percentage of their occurrence among the visitn see the Elixhauser approach encompasses several more diseases than the Charlson approach.

Slide #28 –

I have many people to thank, including my 2 sons Andrew & Malcolm who are present here today.

Slide #29 –

I welcome any comments and suggestions for improvement.
I will be posting improved versions of the programs on the SSC (Statistical Software Components) Archive (repository of RePEc), after incorporating some additional features. An older version of my charlson program is currently available on the SSC Archive.

Thank you very much!