Astrology and health outcomes – lessons for clinical and epidemiological research

Peter C Austin

Institute for Clinical Evaluative Sciences

Toronto, Ontario

June 11, 2008

Summary of talk

1. Overview of prior research on astrology and health.

2. Astrology and health care outcomes in Ontario, Canada.

3. Implications for the conduct and interpretation of clinical and epidemiological research.

An overview of research on astrology and health

- ISIS-2 (Second International Study of Infarct Survival) randomized 17,187 patients with suspected acute myocardial infarction.
- Included patients entering 417 hospitals in 16 countries.
- Streptokinase alone and aspirin alone produced a highly significant reduction in 5-week vascular mortality.

  – Lancet 1988;2(8607):349-360.

- A subgroup analysis indicated that there was a slight adverse effect of aspirin on mortality for patients born under Gemini or Libra.

- For patients born under all other astrological signs there was a strikingly beneficial effect.

Why did the authors report the results of this subgroup analysis?

- “even in a trial as large as ISIS-2, reliable identification of subgroups of patients among whom treatment is particularly advantageous is unlikely to be possible. When in a trial with a clear positive overall result many subgroup analyses are considered, *false* negative results in some particular subgroups must be expected” (ISIS-2 authors).

- “it is of course, clear that the best estimate of the real size of the treatment effect in each astrological subgroup is given not by the results in that subgroup alone but by the overall results in all subgroups combined” (ISIS-2 authors).

Disclaimer

Warning: taking this subject matter too seriously can be hazardous to your health.

Psychology and survival

- Chinese-Americans, but not whites, die significantly earlier than normal if they have a combination of disease and birth year which Chinese astrology and medicine consider ill-fated.
- The more strongly a group is attached to Chinese traditions, the more years of life are lost.
- Authors concluded that reduction in survival was a result, at least in part, from psychosomatic processes.

Source: Lancet 1993;342(8880):1142-5.

Astrological Signs and health outcomes in Ontario

Ontario

Canada's most populous province

(population 12,686,952 in 2006)

Data sources

- Registered Person's Database (RPDB): Database maintained by the Ontario Ministry of Health and Long Term Care. Contains basic demographic data on all residents of Ontario that are eligible for provincial health care insurance.

- Canadian Institute for Health Information (CIHI) Discharge abstract database (DAD): Records demographic and clinical detail on every hospitalization in Ontario.

- Each database contains an encrypted version of each resident's health insurance number.

Study population

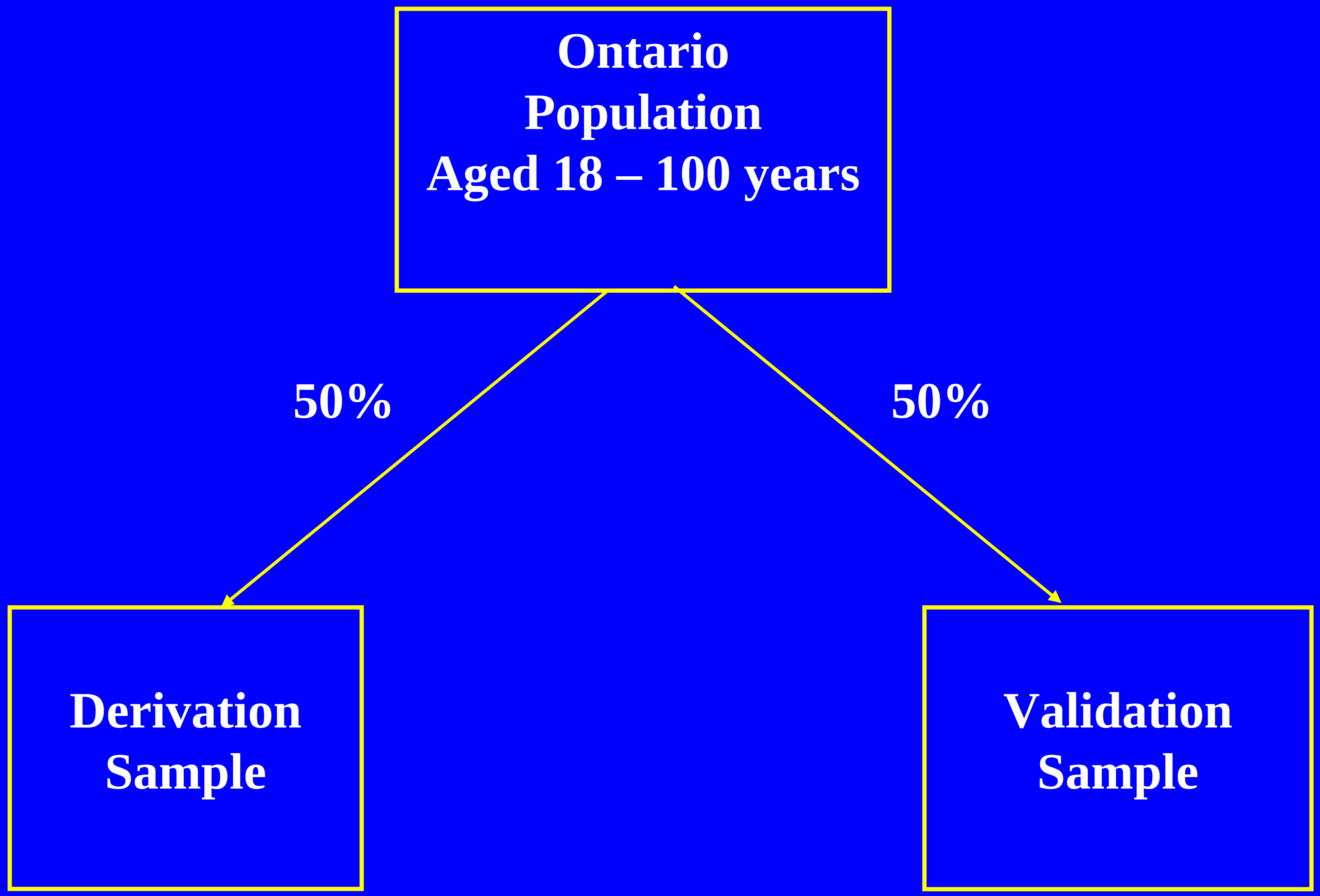
The RPDB was used to identify residents of Ontario aged 18-100.

We identified 10,674,945 residents of Ontario aged 18 to 100 years in 2000, who were alive on their birthday in 2000.

We determined the astrological sign under which each resident of Ontario was born using their birth date recorded in the RPDB.

Residents were randomly divided into equally sized derivation and validation samples.

Figure 1


Ontario
Population
Aged 18 – 100 years
50%
50%
Derivation
Sample
Validation
Sample

Astrological signs

- Aries March 21-April 19
- Taurus April 20-May 20
- Gemini May 21-June 21
- Cancer June 22-July 22
- Leo July 23-August 22
- Virgo August 23-September 22
- Libra September 23-October 22
- Scorpio October 23 - November 21
- Sagittarius November 22-December 21
- Capricorn December 22-January 19
- Aquarius January 20-February 18
- Pisces February 19-March 20

(www.astrology.com)

Diagnoses for hospitalizations

We examined the CIHI discharge abstract database for all hospital admission among subjects aged 18 to 100 years between January 1, 2000 to December 31, 2001.

Only admissions that were classified as urgent or emergent were selected. Elective or planned admissions were excluded.

Each admission was classified according to the most responsible diagnosis, using the first three digits of the ICD-9 coding scheme.

Diagnoses were then ranked from most frequent to least frequent.

Identifying zodiac signs at increased risk of hospitalization

Beginning with the most frequent cause of hospitalization, we:

- Determined which subjects in the derivation sample had been hospitalized with this diagnosis in the year following their birthday in 2000.
- Determined the proportion of residents born under each sign that were hospitalized within a year of their birthday in 2000.
- Identified the astrological sign with the highest probability of hospitalization.
- Tested whether the probability of hospitalization was statistically significantly different in this sign than in the other signs combined using a Chi-squared test.

This process was repeated for all diagnoses until two diagnoses were identified for each astrological sign.

Results

- We searched through 223 out of 895 possible urgent or emergent diagnoses.
- Of these 223 diagnoses, there were 72 (32.3%) for which residents born under one sign had a significantly higher probability of hospitalization compared to residents born under the remaining 11 signs combined.
- The number of significant diagnoses ranged from a low of 2 (Scorpio) to a high of 10 (Taurus), with a mean of 6 diagnoses for each astrological sign.

Two most frequent significant causes of hospitalizations per sign

Astrological Sign	ICD-9 Code	Diagnosis	Relative Risk	P-Value
Aries	733	Other disease of bone and cartilage	1.27	0.0402
	008	Intestinal infections due to other organisms	1.41	0.0058
Taurus	820	Fracture of neck of femur	1.11	0.0368
	562	Diverticula of intestine	1.27	0.0006
Gemini	998	Other complications of procedures, NEC	1.15	0.0330
	303	Alcohol dependence syndrome	1.30	0.0154
Cancer	560	Intestinal obstruction without mention of hernia	1.12	0.0475
	285	Other and unspecified anemias	1.27	0.0388

Two most frequent significant causes of hospitalizations per sign

Astrological Sign	ICD-9 Code	Diagnosis	Relative Risk	P-Value
Leo	578	Gastrointestinal hemorrhage	1.23	0.0041
	V58	Encounter for other and unspecified procedure and aftercare	1.17	0.0397
Virgo	823	Fracture of tibia and fibula	1.26	0.0355
	643	Excessive vomiting in pregnancy	1.40	0.0344
Libra	808	Fracture of pelvis	1.37	0.0108
	430	Subarachnoid hemorrhage	1.44	0.0377
Scorpio	566	Abscess of anal and rectal region	1.57	0.0123
	204	Lymphoid leukemia	1.80	0.0395

Two most frequent significant causes of hospitalizations per sign

Astrological Sign	ICD-9 Code	Diagnosis	Relative Risk	P-Value
Sagittarius	784	Symptoms involving head and neck	1.30	0.0376
	812	Fracture of humerus (no laughing matter)	1.28	0.0458
Capricorn	799	Other ill-defined and unknown in causes or morbidity and mortality	1.29	0.0105
	634	Abortion	1.28	0.0242
Aquarius	413	Angina pectoris	1.23	0.0071
	481	Other bacterial pneumonia	1.33	0.0375
Pisces	428	Heart failure	1.13	0.0013
	411	Other acute and subacute forms of ischemic heart disease	1.10	0.0182

Validation sample

- The above results were generated in the derivation sample.
- We tested each of the above 24 associations in the independent validation sample.
- Only 2 of the 24 associations were significant in the validation sample.
  - Leos had a significantly higher probability of hospitalization due to gastrointestinal hemorrhage, with a relative risk of 1.15 ($P = 0.0483$).
  - Sagittarius had a significantly increased risk of hospitalization due to fracture of the humerus, with a relative risk of 1.38 ($P = 0.0125$).
- The remaining 22 associations were no longer significant ($0.0743 \le P \le 0.9574$).

Implications for clinical and epidemiological research

- Multiple significance testing
- Data-driven statistical analyses
- Importance of biological plausibility
- Subgroup analyses
- Validation studies
- Measures of effect
- Data mining

Multiple significance testing

- In the validation sample we tested 24 distinct hypotheses.
- Under the null hypothesis, P-values are uniformly distributed between 0 and 1 – the probability of a Type I error is 0.05, when using a 0.05 significance level.
- If all 24 null hypotheses were true, then the probability of correctly concluding that all 24 were true would be $(1 - 0.05)^{24} = 0.292$

Multiple testing (2)

- The probability of making at least one Type I error is 0.708.
- To account for 24 statistical tests, one could use a test-wise significance level of 0.00213 to preserve an overall Type I error rate of 5%.
- Using a significance level of 0.00213, none of the 24 associations would be significant in the validation sample.

Multiple testing (3)

Unstructured multiple hypothesis testing should account for the increased risk of the Type I error rate.

Statistical methods to adjust for multiple comparisons are well described in the statistical literature.

Data-driven analyses

- In the derivation sample, we compared the astrological sign with the highest probability of the outcome with all other signs combined.
- Our dichotomization of astrological signs was data-driven, and not driven by theory or prior experience.
- Significance-testing was not used – no need to adjust for inferential multiple comparisons.

Data-driven analyses (2)

- The Chi-squared test assumes that the comparison was pre-specified and not selected according to the data.
- When the data under analysis influences how variables are analyzed then statistical tests may not perform as advertised.
- We used a data-driven approach to generate hypotheses.

Data-driven analyses (3)

These methods are frequently used in statistical analyses in the medical and epidemiological literature

Automated variable selection methods are commonly used in biomedical research.

- Forward, backwards, and stepwise variable selection use repeated significance testing to determine the variables to include in the regression model.

Data-driven analyses (4)

Automated variable selection methods have been shown to result in:

- P-values that are biased low.
- Regression coefficients that are biased high in absolute value.
- Models that contain a high proportion of ‘noise’ variables.
- Confidence intervals that have low coverage probabilities.
- Non-reproducible models.

Data-driven analyses (5)

Data-driven analyses in observational and experimental studies can result in mis-leading conclusions.

Selecting variables for inclusion or thresholds for categorization or dichotomization of variables based on significance testing can lead to studies that are biased towards finding a significant association.

Data-driven analyses: instability of automated variable selection methods

- Data were collected on a sample of 4,911 patients hospitalized with an acute myocardial infarction (AMI) between April 1, 1999 and March 31, 2001 at 57 Ontario hospitals. The data were collected as part of the EFFECT study.
- Data on patient history, cardiac risk factors, comorbid conditions and vascular history, vital signs on admission, and laboratory tests were collected from the patients' medical records using retrospective chart review.
- We selected variables whose univariate association with 30-day mortality had a significance level of $P < 0.25$ and whose prevalence was at least 1%.

Reference: Journal of Clinical Epidemiology. 2004;57:1138-1146.

Case study - Data

Demographic	age and gender
Presenting characteristics	acute pulmonary edema; cardiogenic shock.
Cardiac risk factors	diabetes; smoking history; history of CVA/TIA; hyperlipidemia; family history of CAD.
Comorbid conditions and vascular history	angina; cancer; dementia; previous AMI; depression; peripheral arterial disease; previous PTCA; congestive heart failure (chronic); aortic stenosis.
Vital signs on admission	systolic BP; diastolic BP; heart rate; respiratory rate.
Laboratory tests	hemoglobin; white blood count; sodium; potassium; glucose; urea; creatinine.
Outcome	30-day mortality

Case study - Methods

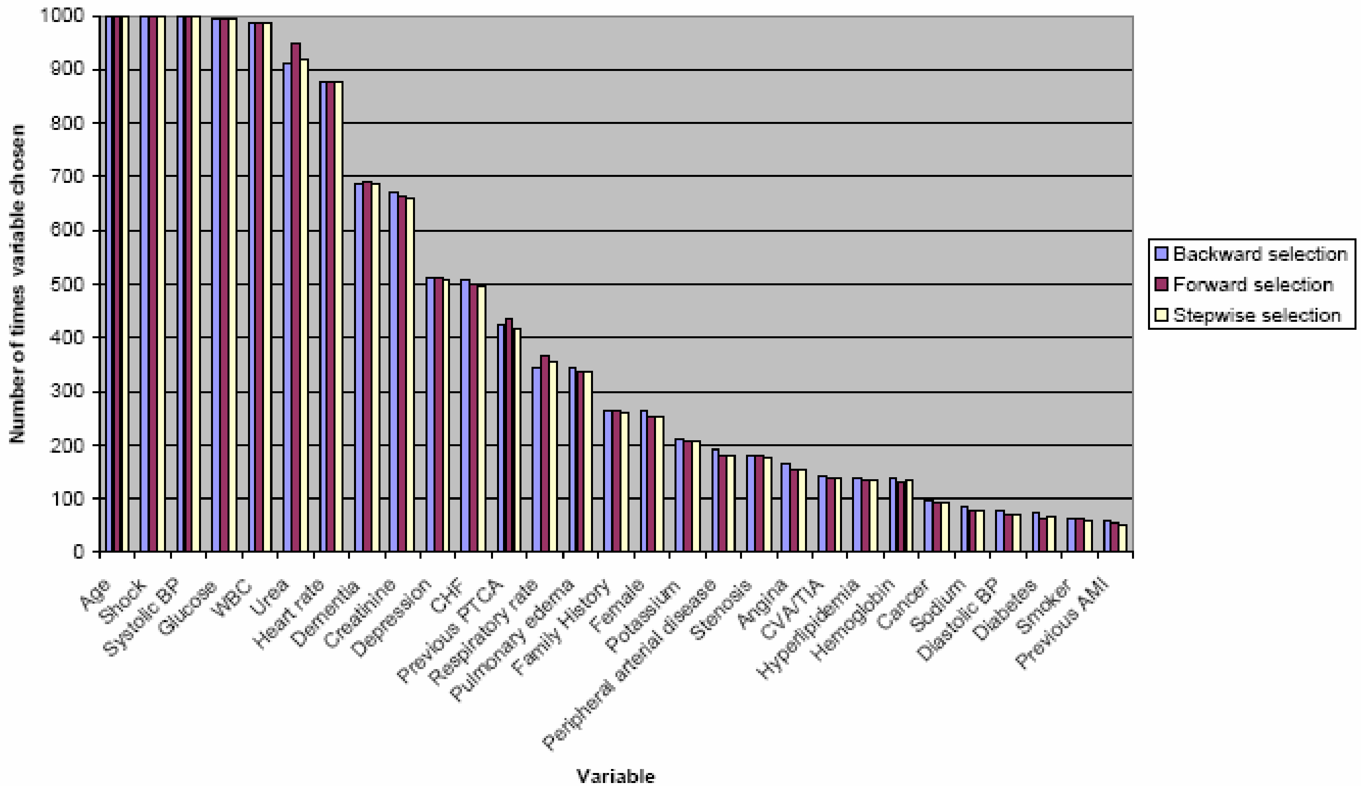
- From the initial sample we drew 1,000 bootstrap samples (samples of the same size as the initial sample, each drawn with replacement from the initial sample).
- In each bootstrap sample, we used forward selection, backward elimination, and stepwise selection using significance levels of 0.05 to identify independent predictors of 30-day AMI mortality.
- We then determined the frequency with which each of the 29 individual predictors were identified as statistically significant predictors of 30-day AMI mortality across the 1,000 bootstrap samples.

Case study - Results

- Backwards model selection resulted in 940 unique regression models in the 1,000 bootstrap samples.
- 889 models were selected only once, 45 models were selected twice, 3 models were selected three times, and 3 models were chosen four times.
- Forward and stepwise variable selection produced similar results.

Case study - Results

Figure 1. Number of times each variable was selected by automatated variable selection methods



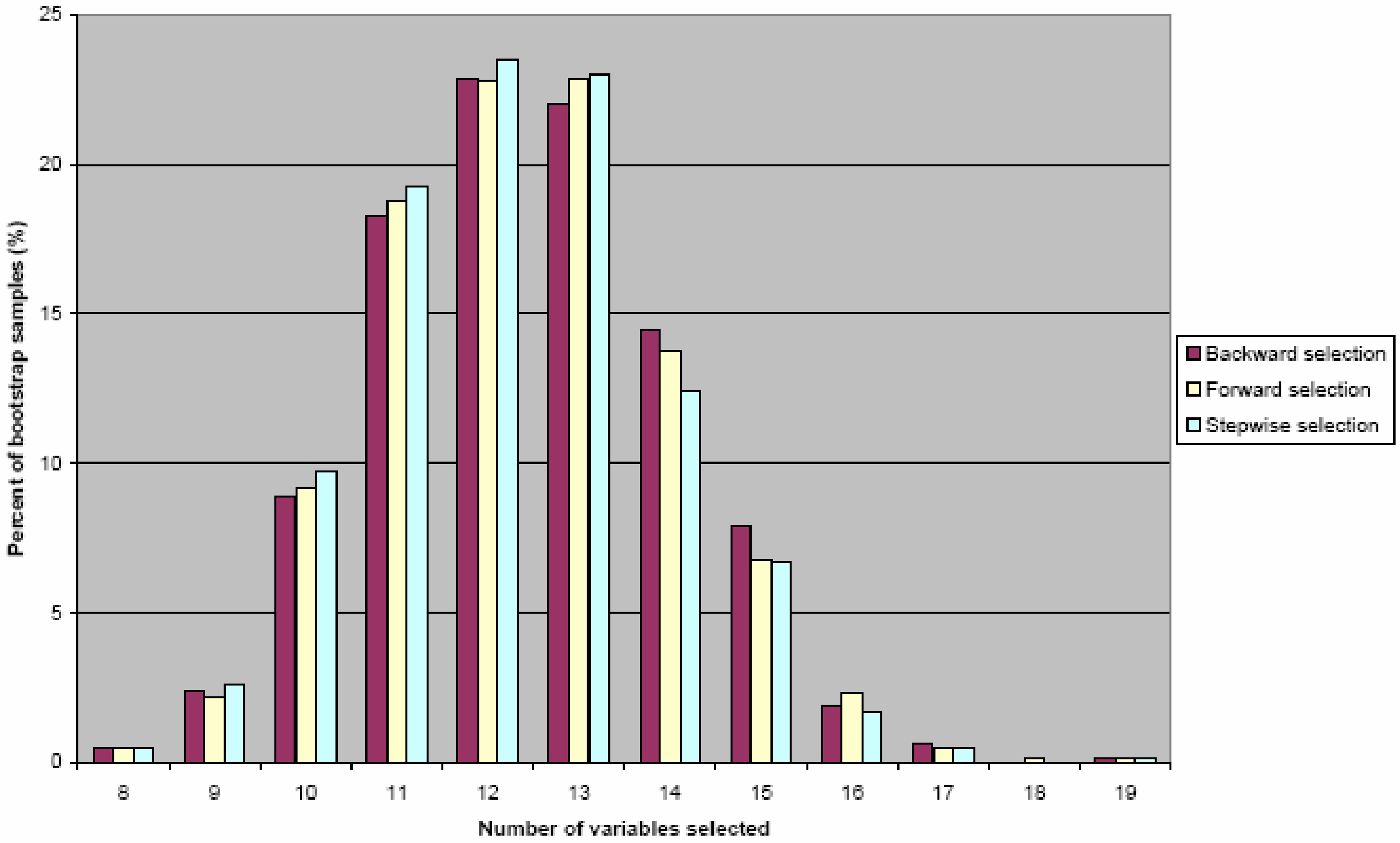
- 3 variables (age, systolic BP, and cardiogenic shock) were identified as significant predictors of AMI mortality in 100% of the bootstrap samples using each method.

- 3 additional variables (glucose, white blood count, and urea) were identified as significant predictors of AMI mortality in at least 90% of the samples using each method.

- 6 variables (cancer, sodium, diastolic BP, diabetes, smoking status, and history of previous MI) were selected in fewer than 10% of the bootstrap samples. However, at least one of these six variables was identified as a significant predictor in 37.3% of the samples using backwards elimination.

- 12 variables were identified as independent predictors in fewer than 20% of the bootstrap samples. However, at least one of these 12 variables was identified as a significant predictor in over 75% of the bootstrap samples using backwards selection.

Figure 2. Number of variables in final model



Biological Plausibility

- None of our derived hypotheses had any apparent biologic plausibility.

- There is no currently plausible mechanism by Leos might be predisposed to gastrointestinal hemorrhage or Sagittarians to humeral fractures.

Biological Plausibility (2)

- The examples in our case-study were intended to be humorous.

- We speculate that, had we used different biological or socio-demographic categorizations, then post-hoc explanations could have been constructed for many of the observed associations.

Biological Plausibility (3)

- Hypothesized associations should be pre-specified and should usually have biological plausibility.
- Caution is required in interpreting results that do not have biological plausibility.
- Non biologically plausible results should be replicated in independent studies.

Subgroup analyses in clinical trials

- Subgroup analyses and multiple safety and efficacy endpoints are common in RCTs.
- We examined 131 RCTs published in the *Journal of the American Medical Association*, *New England Journal of Medicine*, *The Lancet*, and the *BMJ* between January 1, 2004 and June 30, 2004.
  - Mean and median number of subgroups were 5.1 and 2.
  - Mean and median number of endpoints were 26.5 and 19.
  - Maximum number of subgroups and endpoints were 68 and 185, respectively.

Subgroup analyses (2)

Authors have suggested guidelines for subgroup analyses:

- Subgroup analyses should be pre-specified.
- Subgroup analyses should have biological plausibility.
- Subgroup analyses and secondary outcomes should only be examined if primary endpoint is significant.
- One should be guided by trends and consistency, rather than statistical significance.

Validation Studies

- The current study used independent derivation and validation samples.
- The use of derivation/validation samples has frequently been advocated in the statistical literature.
- The use of validation sample allows one to assess the reproducibility of findings generated in the derivation sample.

Validation studies (2)

- The PRAISE study examined the effect of amlodipine in patients with congestive heart failure and found no benefit in the primary analysis.
- A subgroup analysis demonstrated that amlodipine reduced the risk of fatal and non-fatal events in patients with severe non-ischemic heart failure ($P = 0.04$).
- Amlodipine helped prevent a secondary outcome (mortality) in the same patients ($P < 0.0001$).

N Engl J Med 1996;335:1107-14.

Validation studies (3)

- The PRAISE-2 was designed to examine the effect of amlodipine in non-ischemic heart failure patients.
- There was no effect on mortality or cardiac events.
- Trial never reported in detail – Clinical Trials Update: OPTIME-CHF, PRAISE-2, ALL-HAT. Eur J Heart Fail 2000;2:209-212.

Validation studies (4)

- ELITE trial suggested a survival benefit in elderly heart failure patients treated with losartan compared to captopril (Lancet 1997;349:747-752).

- This finding was not replicated in the ELITE II trial (Lancet 2000;355:1582-7).

Measures of Effect

- We reported the relative risk to compare the risk of hospitalization for the astrological sign with the highest rate of hospitalization compared to the other signs combined.
- Relative risks ranged from 1.10 to 1.80.
- Absolute risk of hospitalization ranged from a low of 0.002% to a high of 0.160%.

Measures of effect (2)

- Relative risk reductions, which are commonly reported in clinical research, can make exposure effects appear more striking.
- Relative risk does not convey information about the baseline risk of the event.

Measures of effect (3)

Multiple measures of effect should be conveyed in clinical research.

Researchers should report

- baseline risk
- absolute risk reduction
- relative risk reduction
- number needed to treat

Data Mining

Data mining has been described as “*the nontrivial extraction of implicit, previously unknown, and potentially useful information from data*” (Cambridge Dictionary of Statistics).

Or as a “*semi-automatic extraction of patterns, changes, associations, anomalies, and other statistically significant structures from large datasets*” (www.rgrossman.com/dm.htm).

Data mining (2)

- We began with no pre-specified hypotheses
- We used automated methods to detect significant associations.
- We used an independent validation sample to test our associations.

Data mining (3)

Our study demonstrates that findings obtained using data mining should be interpreted with some degree of skepticism.

Acknowledgements

- This was joint work with Drs. Muhammad Mamdani, David Juurlink, and Janet Hux.

- Drs. Austin, Mamdani, Juurlink were supported by New Investigators Awards from the Canadian Institutes for Health Research (CIHR).

- Study published in the *Journal of Clinical Epidemiology* 2006;59:964-969.

- peter.austin@ices.on.ca