The Trauma Mortality Prediction Model
(-tmpm.ado-)
is Robust to the AIS, ICD-9, and
the ICD-10 Lexicons

Alan Cook, MD, FACS
Turner Osler, MD, MS (Biostatistics), FACS

Stata Conference, San Diego
July 26-27, 2012
• Objectives
  – Briefly Discuss:
    • The importance of mortality prediction in trauma
    • The foregoing mortality prediction models in trauma
  – Introduce the Trauma Mortality Prediction Model (TMPM)
  – Highlight the ~tmpm.ado~ user-written command for Stata
TMPM: Trauma Mortality Prediction Model
# 10 Leading Causes of Death by Age Group, United States – 2009

<table>
<thead>
<tr>
<th>Age Groups</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td></td>
</tr>
<tr>
<td>15-24</td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Cause</th>
<th>&lt;1</th>
<th>1-4</th>
<th>5-9</th>
<th>10-14</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Congenital Anomalies</td>
<td>5,319</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Short Gestation</td>
<td>4,518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SIDS</td>
<td>2,226</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Malignant Neoplasms</td>
<td>1,608</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Unintentional Injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Placenta Cord, Membranes</td>
<td>1,064</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Septicemia</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Respiratory Disease</td>
<td>581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Circulatory System Disease</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Neonatal Hemorrhage</td>
<td>517</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data source: National Vital Statistics System, National Center for Health Statistics, CDC.
Produced by: Office of Statistics and Programming, National Center for Injury Prevention and Control, CDC using WISQARS™.
Why estimate mortality in populations of injured patients?

- The quantification of trauma is required for:
  - The scientific study of injury
  - The objective evaluation of trauma care
- We compromise our research and our patient care if we fail to develop and use accurate measures of trauma outcomes
  - Specifically mortality
• If a human being goes through a windshield at 80 mph, it might seem that there are an infinite number of different possible injuries could result.

• A different spectrum of injuries could result from a stab wound or gunshot wound to the chest or abdomen.
TMPM: Trauma Mortality Prediction Model
• Human beings are structurally similar to one another
  – Human anatomy tends to “fail” in structurally similar ways
  – There are only ~1,000 different injuries a person could have
    • Fortunately, these injuries have already been extravagantly detailed in injury coding lexicons such as the AIS and ICD-9 coding systems
• Unfortunately, a person could have up to, say 10 injuries

• So the number of injury patterns one could have is $10^{1000}$
  – This is more than the number of atoms in the universe: $10^{80}$
Mortality estimation is not new in the field of trauma surgery

- Injury Severity Score (ISS)
  - Based on the Abbreviated Injury Score

- Trauma Score & ISS (TRISS)

- International Classification of Disease-9 (ICD-9) based injury severity score (ICISS)
TMPM: Trauma Mortality Prediction Model

• Most mortality prediction models are based on two administrative coding systems
  – International Classification of Diseases, 9th Revision, Clinical Modification, (ICD-9 CM or ICD-9)
    • Primarily a taxonomy of ALL human disease, including traumatic injuries
    • The basis of billing transactions between providers and payers
  – Abbreviated Injury Score
    • A taxonomy of traumatic injury and includes a severity score
    • Used in trauma data registries and mortality risk models
ICD-9 Lexicon for Coding Liver Injuries

864 Injury to liver

The following fifth-digit subclassification is for use with category 864:

0 unspecified injury
1 hematoma and contusion
2 laceration, minor
   Laceration involving capsule only, or without significant involvement of hepatic parenchyma [i.e., less than 1 cm deep]
3 laceration, moderate
   Laceration involving parenchyma but without major disruption of parenchyma [i.e., less than 10 cm long and less than 3 cm deep]
4 laceration, major
   Laceration with significant disruption of hepatic parenchyma [i.e., 10 cm long and 3 cm deep]
   Multiple moderate lacerations, with or without hematoma
   Stellate lacerations of liver
5 laceration, unspecified
9 other
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>541899.2</td>
<td>Liver NFS contusion (hematoma) NFS subcapsular, ≤ 50% surface area, nonexpanding or intraparenchymal ≤ 10 cm in diameter; minor; superficial; (OIS Grade I or II)</td>
</tr>
<tr>
<td>541810.2</td>
<td></td>
</tr>
<tr>
<td>541812.2</td>
<td></td>
</tr>
<tr>
<td>541814.3</td>
<td>&gt; 50% surface area or expanding; ruptured subcapsular or parenchymal; intraparenchymal &gt; 10 cm or expanding; blood loss &gt; 20% by volume; major; subcapsular; (OIS Grade III)</td>
</tr>
<tr>
<td>541820.2</td>
<td>laceration NFS simple capsular tears, ≤ 3 cm parenchymal depth, ≤ 10 cm in length; blood loss ≤ 20% by volume; minor; superficial (OIS Grade I or II)</td>
</tr>
<tr>
<td>541822.2</td>
<td></td>
</tr>
<tr>
<td>541824.3</td>
<td>&gt; 3 cm parenchymal depth; major duct involvement; blood loss &gt; 20% by volume; moderate (OIS Grade III)</td>
</tr>
<tr>
<td>541826.4</td>
<td>parenchymal disruption of ≤ 75% of hepatic lobe or 1-3 Couinaud’s segments within a single lobe; multiple lacerations &gt; 3 cm deep; &quot;burst&quot; injury; major (OIS Grade IV)</td>
</tr>
<tr>
<td>541828.5</td>
<td>parenchymal disruption of &gt; 75% of hepatic lobe or involving &gt; 3 Couinaud’s segments within a single lobe or involving retrohepatic vena cava/central hepatic veins; massive; complex (OIS Grade V)</td>
</tr>
<tr>
<td>541830.6</td>
<td>hepatic avulsion (total separation of all vascular attachments) (OIS Grade VI)</td>
</tr>
<tr>
<td>641840.4</td>
<td>rupture (&quot;fracture&quot;) NFS Use this code only when a more detailed description is not available.</td>
</tr>
</tbody>
</table>
A third administrative coding system is on the horizon for US healthcare
  - ICD-10
    • Industry standard for most of the world
    • Like the metric system, not implemented in the US yet
      - US Department of Health and Human services delayed compliance to October 1, 2013, then delayed again to October 2014
TMPM: Trauma Mortality Prediction Model

ICD-10 Lexicon for Coding Liver Injuries

- S36.11 Injury of liver
- S36.112 Contusion of liver
- S36.113 Laceration of liver, unspecified degree
- S36.114 Minor laceration of liver
  Laceration involving capsule only, or, without significant involvement of hepatic parenchyma [i.e., less than 1 cm deep]
- S36.115 Moderate laceration of liver
  Laceration involving parenchyma but without major disruption of parenchyma [i.e., less than 10 cm long and less than 3 cm deep]
- S36.116 Major laceration of liver
  Laceration with significant disruption of hepatic parenchyma [i.e., greater than 10 cm long and 3 cm deep]
  Multiple moderate lacerations, with or without hematoma
  Stellate laceration of liver
- S36.118 Other injury of liver
- S36.119 Unspecified injury of liver
• The Trauma Mortality Prediction Model (TMPM)
  – We wanted to be able to provide an estimated probability of mortality for any of these $10^{1000}$ injury patterns
  – How are we to proceed?
  – Given a large enough dataset, one might estimate the probability of producing death for each of the 1,000 possible injuries
    • We’ll call that “severity” or the “dose of trauma”
TMPM: Trauma Mortality Prediction Model

• **TMPM Development**
  – The basic idea was to fit a probit model with death as the outcome with 1,000 possible injuries as binary predictors
    • The cohort was the National Trauma Data Bank (NTDB)
      – N=1,000,000
    • Design matrix was 1,000 variables wide and 1,000,000 cases long
      – Mostly zeros – Most patients don’t have most injuries
      – The sum of the dataset provided a rich mosaic of injuries with the associated patient survival
TMPM: Trauma Mortality Prediction Model

- **TMPM Development**
  - Two separate probit models were created using the injuries described in the coding system
    - Model 1 used all possible injuries as binary predictors and death as a binary outcome
    - Model 2 was based on body region severity indicators
  - The empiric injury severity for each injury was estimated by taking a weighted average of the coefficients of the 2 regression models
Why two probit models?

- Even in the NTDB dataset, the prevalence of specific injuries was very uneven.
  - A few injuries were very common (minor skin laceration of the face)
  - Many injuries were very rare
    - Twenty-four (2%) AIS codes occurred only once in the entire dataset.
    - 684 codes (52%) occurred fewer than 100 times
- To estimate the severity of these rare injuries, they were “lumped” into larger groups
  - Now we have 2 estimates for each injury – so now what?
  - Average the 2 models weighted on the inverse of the variance of each of the 2 estimates!
TMPM: Trauma Mortality Prediction Model

- **MARC: Model Averaged Regression Coefficients**
  - How the models were averaged

\[
\frac{1}{\text{Var}(\text{Model 1})} \cdot \text{Coef}_{\text{Model 1}} + \frac{1}{\text{Var}(\text{Model 2})} \cdot \text{Coef}_{\text{Model 2}}
\]

\[
\text{Var}(R\text{Scoef}_m) = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} N_i \times W_j \times \left[ (AIS\text{coef}_j) - E(AIS\text{coef}) \right]^2
\]

Where:
- \(N_i\) is the number of AIS codes in RS region \(m\)
- \(W_j\) is the weight of the contribution of \(AIS\text{coef}_j\) to the variance of \(R\text{Scoef}_m\)

\[
W_j = \frac{1}{\text{Var}(\text{AIScoef}_j)} \sum_{r=1}^{N_i} \frac{1}{\text{Var}(\text{AIScoef}_r)}
\]

\[
E(\text{AIScoef}) = \sum_{j=1}^{N_i} W_j \times \text{AIScoef}_j
\]
• Ultimately the probability of death or \( p(\text{Death}) \) is computed as follows:

\[
p(\text{Death}) = \Phi \left[ C_0 + C_1 \times I_1 + C_2 \times I_2 + C_3 \times I_3 + C_4 \times I_4 + C_5 \times I_5 + C_6 \times S + C_7 \times I_1 \times I_2 \right]
\]

Where \( \Phi \) is the inverse normal function

- \( C_x \) are the coefficients from the model(s)
- \( I_x \) are the marc values of the 5 worst injuries
- \( S \) is an indicator variable indicating whether the worst 2 injuries are in the same body region
TMPM: Trauma Mortality Prediction Model

### TMPM vs. Other ICD-9 Based Mortality Prediction Scores

<table>
<thead>
<tr>
<th>Model</th>
<th>ROC</th>
<th>HL Stat</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICISS</td>
<td>0.846</td>
<td>432</td>
<td>44,071</td>
</tr>
<tr>
<td>SWI: SRR</td>
<td>0.861</td>
<td>833</td>
<td>42,813</td>
</tr>
<tr>
<td>SWI: MARC</td>
<td>0.872</td>
<td>112</td>
<td>42,716</td>
</tr>
<tr>
<td>All Injury: MARC</td>
<td>0.878</td>
<td>357</td>
<td>42,688</td>
</tr>
<tr>
<td>TMPM-ICD9</td>
<td>0.880</td>
<td>19</td>
<td>41,251</td>
</tr>
</tbody>
</table>
## TMPM: Trauma Mortality Prediction Model

### TMPM vs. Other AIS Based Mortality Prediction Scores

<table>
<thead>
<tr>
<th>Model</th>
<th>ROC**</th>
<th>HL Stat*</th>
<th>AIC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISS</td>
<td>0.872</td>
<td>296</td>
<td>37225</td>
</tr>
<tr>
<td>SWI</td>
<td>0.891</td>
<td>314</td>
<td>34059</td>
</tr>
<tr>
<td>TMPM</td>
<td>0.902</td>
<td>58</td>
<td>32003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>ROC**</th>
<th>HL Stat*</th>
<th>AIC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISS age+gender+mech</td>
<td>0.915</td>
<td>54</td>
<td>33773</td>
</tr>
<tr>
<td>SWI age+gender+mech</td>
<td>0.921</td>
<td>128</td>
<td>31770</td>
</tr>
<tr>
<td>TMPM age+gender+mech</td>
<td>0.928</td>
<td>19</td>
<td>29645</td>
</tr>
</tbody>
</table>
TMPM: Trauma Mortality Prediction Model
TMPM: Trauma Mortality Prediction Model

- We used Stata to understand the problem and create a prediction model.
- Unfortunately, the prediction model was complicated enough that we feared users might not be able to implement it easily and accurately.
- So, we used Stata in a second context, to create a simple command that would compute our model correctly, no matter who was using our model...."
TMPM: Trauma Mortality Prediction Model

• `~tmpm.ado~`
  – Goals
    • Execute the TMPM calculation of $p(\text{Death})$ correctly
    • Design `~tmpm.ado~` to accommodate AIS, ICD-9 or ICD-10 lexicons from a single command line
      – One at a time, though
    • Streamline the command itself
      – Can recognize if the data are formatted as wide or long without the user specifying such in the command
      – Give the user the option of receiving information about their dataset while the program is running
TMPM: Trauma Mortality Prediction Model

• `~tmpm.ado~`
  – Three modules to the program
    • ICD-9
    • AIS
    • ICD-10
  – `syntax [varlist], [ idvar(varname) aispfx(string) icd9pfx(string) icd10pfx(string) NOREPORT ]`
  – The appropriate module is called based on which `~pfx` is identified in the command line
  – The option “NOREPORT” suppresses the output of reports about the user’s dataset
TMPM: Trauma Mortality Prediction Model

tmpm, idvar(string) icd9pfx(string)

Display:
"Trauma Mortality Prediction Model p(Death) Estimation Module"
"Input Lexicon: ICD-9"

Preserve user’s dataset

Determine shape of dataset
(count number of columns with icd9pfx*)

If user’s dataset wide, make it long
TMPM: Trauma Mortality Prediction Model

- icd9 check -

Display Report: "Proportion of invalid ICD-9 codes"

Remove invalid ICD-9 codes
-drop if check != 0-

Check for codes outside of the range 800.00 - 959.9
(This is the range of ICD-9 codes for traumatic injury)

Display Report: "Proportion of ICD-9 codes unrelated to trauma"

Match ICD-9 codes to MARC values
-merge m:1 using ...

Display Report: "Proportion of codes matched with marc values"

Display list of unmatched ICD-9 codes
TMPM: Trauma Mortality Prediction Model

Keep two worst injuries and identify if they are from same body region (this is a variable from the MARC table)

Keep five worst (highest) marc values

Generate interaction term for two worst injuries

Calculate the model ("xBeta")
```
gen xBeta = (1.406958*marc1) + 
            (1.409992*marc2) + 
            (0.5205343*marc3) + 
            (0.4150946*marc4) + 
            (0.8883929*marc5) + 
            -(0.0890527)*same_region) + 
            -(0.7782696)*Imarc) + 
            -(2.217565)
```
Generate the p(Death) from the xBeta term
```
gen pDeathicd9=normal(xBeta)
```

Save this result as a temporary file, sorted by "idpfx"

Restore user’s dataset

Merge the temporary file with the user’s file matched on "idpfx"
This places the p(Death) variable into the user’s dataset.

Display:
"p(Death) estimation from ICD-9 lexicon is complete"

Erase temporary files

End
TMPM: Trauma Mortality Prediction Model

~tmpm.ado output from data using the AIS injury coding system

```
. tmpm, idvar(id) aispfx(predot_)

Trauma Mortality Prediction Model p(Death) Estimation Module
Input Lexicon: AIS

Proportion of codes matched with marc values: 98.75%
Your data contain 434 unique AIS codes
Your unmatched AIS codes:

<table>
<thead>
<tr>
<th>Unmatched</th>
<th>AIS Codes</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>115099</td>
<td>42</td>
<td>65.63</td>
<td>65.63</td>
</tr>
<tr>
<td></td>
<td>115299</td>
<td>5</td>
<td>7.81</td>
<td>73.44</td>
</tr>
<tr>
<td></td>
<td>131806</td>
<td>1</td>
<td>1.56</td>
<td>75.00</td>
</tr>
<tr>
<td></td>
<td>515099</td>
<td>1</td>
<td>1.56</td>
<td>76.56</td>
</tr>
<tr>
<td></td>
<td>919602</td>
<td>3</td>
<td>4.69</td>
<td>81.25</td>
</tr>
<tr>
<td></td>
<td>999999</td>
<td>12</td>
<td>18.75</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 64 | 100.00

p(Death) estimation from AIS lexicon is complete
```
TMPM: Trauma Mortality Prediction Model

```
.tmpm, idvar(id) icd9pfx(icdininj)
```

Trauma Mortality Prediction Model p(Death) Estimation Module
Input Lexicon: ICD-9

Proportion of invalid ICD-9 codes: 1.07%
Proportion of ICD-9 codes unrelated to trauma: 9.64%
Proportion of codes matched with marc values: 98.57%

Your data contain 543 unique ICD-9 codes

Your unmatched ICD-9 codes:

<table>
<thead>
<tr>
<th>ICD-9 Codes</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>905.4</td>
<td>1</td>
<td>2.04</td>
<td>2.04</td>
</tr>
<tr>
<td>920.0</td>
<td>38</td>
<td>77.55</td>
<td>79.59</td>
</tr>
<tr>
<td>930.9</td>
<td>1</td>
<td>2.04</td>
<td>81.63</td>
</tr>
<tr>
<td>941.28</td>
<td>1</td>
<td>2.04</td>
<td>83.67</td>
</tr>
<tr>
<td>946.0</td>
<td>1</td>
<td>2.04</td>
<td>85.71</td>
</tr>
<tr>
<td>948.00</td>
<td>1</td>
<td>2.04</td>
<td>87.76</td>
</tr>
<tr>
<td>958.4</td>
<td>3</td>
<td>6.12</td>
<td>93.88</td>
</tr>
<tr>
<td>958.8</td>
<td>2</td>
<td>4.08</td>
<td>97.96</td>
</tr>
<tr>
<td>959.11</td>
<td>1</td>
<td>2.04</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 49 | 100.00

p(Death) estimation from ICD-9 lexicon is complete
TMPM: Trauma Mortality Prediction Model

• ~tmpm.ado~ is structured to accommodate the ICD-10 lexicon
  – Currently it uses a translation from ICD-9 marc values
  • A warning of this appears in the Results window in place of the journal citation
  • This warning is made because, as yet, there are no ICD-10 data to test our predictions against.
  • Since our model is untested, we suggest that it only be used for research purposes until we have some idea of how well it works.
TMPM: Trauma Mortality Prediction Model

~tmpm.ado output from data using the ICD-10 injury coding system

```
.tmpm, idvar(id) icd10pfx(icd10) noreport

Trauma Mortality Prediction Model p(Death) Estimation Module

WARNING: The TMPM ICD-10 module is based on the NIH/CDC mapping algorithm and has not been evaluated empirically. The authors advise that it not be used for actual risk stratification.

Input Lexicon: ICD-10

p(Death) estimation from ICD-10 lexicon is complete
```
Future Directions

- `~tmpm.ado~` works on Stata 11 and 12
  - We are seeking to retrofit it to earlier versions of Stata.
- Implement a (click to run) feature in the Examples portion of the help file
Thank you

The Trauma Mortality Prediction Model

Thanks to
Kreshna Gopal, Ph.D.
Senior Computer Scientist, StataCorp

Nicholas Cox, Ph.D.
Department of Geography, Durham University
Editor, The Stata Journal
Frequent contributor to Statalist