Validation Of The Integrated Medical Model Using Historical Space Flight Data

Eric Kerstman M.D., M.P.H.

Advanced Projects

Wyle Integrated Science and Engineering Group

2010 Winter Simulation Conference
Baltimore, MD
December 7, 2010

Validation of the Integrated Medical Model Using Historical Space Flight Data

Eric L. Kerstman[1], Charles Minard[2], Mary H. Freire de Carvalho[2], Marlei E. Walton[2], Jerry G. Myers. Jr.[3] , Lynn G. Saile[2], Vilma Lopez[4], Douglas J. Butler[2], Kathy A. Johnson-Throop[5]

[1]University of Texas Medical Branch, Galveston, TX; [2]Wyle, Houston, TX, [3]NASA Glenn Research Center, Cleveland, OH, [4]JesTech, Houston, TX, [5]NASA Johnson Space Center, Houston, TX

Background

- The IMM is expected to be a significant contributor to medical decision making in operational and planning processes for space flight missions
- NASA Standard 7009 requires that real world events be accurately represented by the model results to reach sufficient levels of validation
- For the IMM, this requirement is partially fulfilled by comparing the model's predicted outcomes with observed mission data that has not been included in the model

Validation

- Model Validation
  - “Substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model”
    - Schlesinger et al. Terminology for model credibility. *Simulation.* 32 (3): 103-104
- Historical Data Validation
  - “If historical data exist, part of the data is used to build the model and the remaining data are used to determine (test) whether the model behaves as the system does”
    - Sargent. Verification and Validation of Simulation Models. *Proceedings of the 2007 Winter Simulation Conference*

Data Analysis

- Data on historical space flight missions were collected from mission medical records

- Data available for comparison included
  - Total number of medical events
  - The number of occurrences of each medical event
  - Medical resource utilization

Validation Approach

- Qualitative and quantitative approaches were used to compare historical data to model output
- Qualitative Approach
  - Plots were created to visualize the differences between the model and historical data
- Quantitative Approach
  - Goodness of Fit (GoF) testing was chosen to test the null hypothesis that the predicted outcomes are statistically equivalent to the observed data

Methods

Data Collection

- International Space Station (ISS) missions
  - Increment medical debriefs by ISS crew surgeons
  - ISS Private Medical Conference (PMC) Tool
- Space Shuttle Missions
  - Mission medical debriefs by Shuttle crew surgeons
  - Crew medical debriefs
  - Surgeon logs

Methods

Simulation

- Model was run for seven ISS missions and fourteen Shuttle missions*
- Mission and crew profile was matched to historical mission data [# of crew, sex, mission length, and number of extravehicular activities (EVAs)]
- Each simulation was executed for 20,000 trials

* Data from these missions have not been used as input for the model

Analysis

Qualitative Approach

- Spider Plots
  - Qualitatively assess the accuracy of IMM predictions for the total number of medical events
  - Simultaneously present the predicted and observed data for multiple missions
  - Primarily for face validation
  - Lacks formal statistical testing procedures
  - Useful in identifying potential discrepancies between the IMM and real-world events

Quantitative Approaches

- *Chi-squared Goodness of Fit (GoF)*
  - May be utilized when the expected number of medical events is five or more (*e.g.* skin rashes, headaches)
  - Reasonable test for the total number of medical events, specific medical conditions that occur frequently, and medical resource utilization
  - The test statistic is calculated as: $$X^2_{ts} = \sum_{i=1}^{N} \frac{(E_i - O_i)^2}{E_i} \sim X^2_{df,1-\alpha}$$
  - If the test statistic is greater than the critical value $\left(X^2_{df,1-\alpha}\right)$, then the null hypothesis that the predicted outcomes are statistically equivalent to observed data will be rejected
  - An $\alpha = 0.05$ level of significance was assumed for IMM GoF testing

Quantitative Approaches (Cont'd)

- *Exact Probability Calculations*
  - When expected values for medical events are less than five, *goodness of fit* tests may be done using exact probability calculations
  - The p-value is equal to the proportion of simulated trials where the number of events that occurred is equal to or more than the observed number

Analysis

- Multiple Comparisons
  - The alpha-level for statistical significance was determined using Bonferroni's correction method
  - If *N* statistical tests were performed and the overall alpha was set at 0.05, then the final alpha level for any individual test was 0.05/*N*
- Example
  - A single ISS mission with one crew member and 83 medical conditions
  - The alpha was 0.05/83 (0.0006)
  - Therefore, p-values less than or equal to 0.006 would be statistically significantly different

Results

Total Medical Events - ISS Missions

Mission	Expected	Observed	Difference
1	12	7	5
2	18	14	4
3	18	13	5
4	14	10	4
5	15	14	1
6	17	16	1
7	19	23	-4
Average	16	14	2

Results

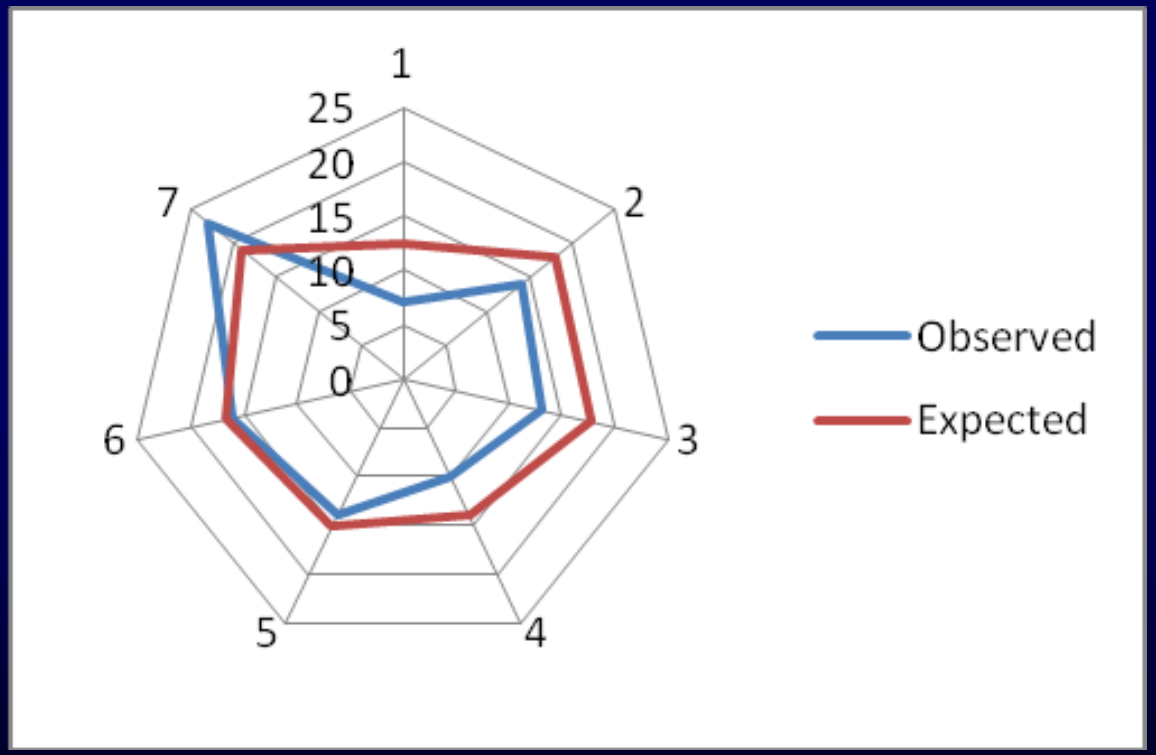
Total Medical Events

- ISS Missions
  - Expected values overestimated the number of medical events for six of the seven missions
  - The difference was not statistically significant ($p = 0.36$)
  - The shape of the expected values is similar to the observed values

Results

Spider Plot for ISS Missions

Total Number of Medical Events by Mission



Results – Total Medical Events – Shuttle Missions

Mission	# of Crew	Expected	Observed	Difference
1	6	24	26	-2
2	6	24	25	-1
3	6	24	22	2
4	7	28	27	1
5	6	25	31	-6
6	5	20	23	-3
7	6	26	28	-2
8	6	25	21	4
9	5	21	20	1
10	6	26	19	7
11	6	24	23	1
12	6	23	19	4
13	6	25	32	-8
14	6	24	21	3
Average	6	24	24	0

Results

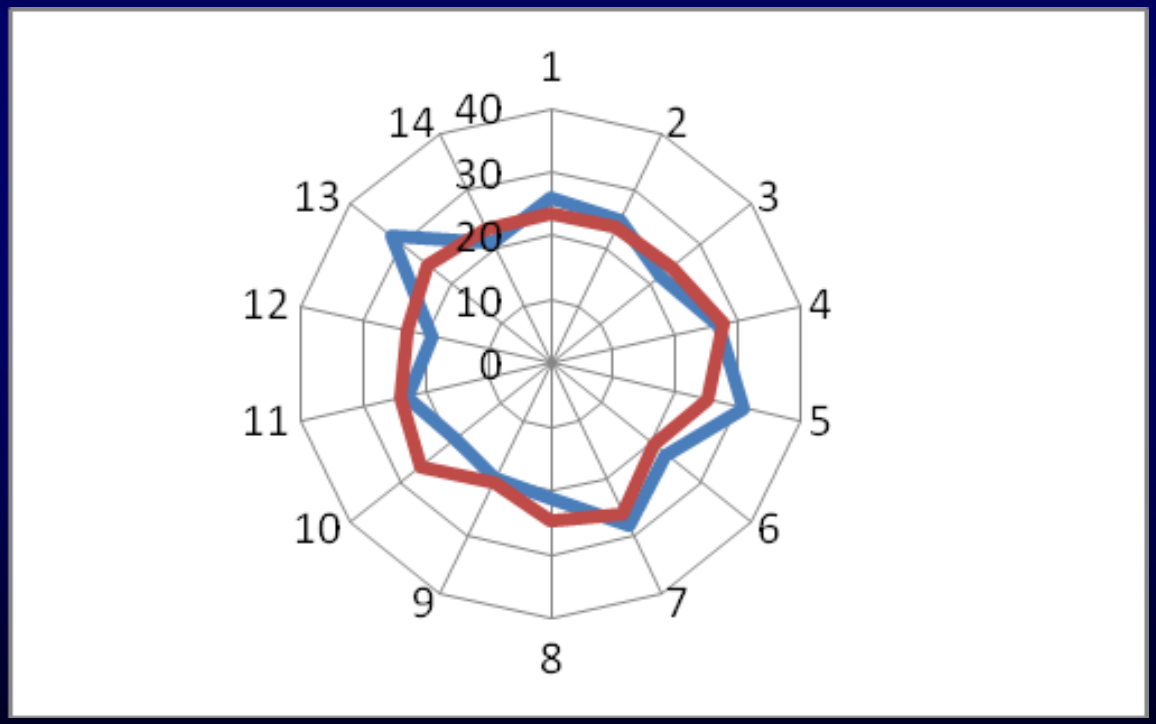
Total Medical Events

- Shuttle Missions
  - Expected values overestimated the number of medical events for eight missions
  - Expected values underestimated the number of medical events for six missions
  - The difference was not statistically different ($p = 0.83$)
  - The shape of the expected values is similar to the observed values

Results

Spider Plot for Shuttle Missions

Total Number of Medical Events by Mission



Results

- Specific Medical Events
  - 83 Medical conditions in the model
  - GoF testing performed for all conditions individually
  - Alpha level set at 0.0006 (0.05/83)

Results

Specific Medical Events

- ISS Missions
  - One medical condition was underestimated by the model (skin abrasion/laceration)
  - Three medical conditions were overestimated by the model (hip sprain/strain, paresthesias, and CO_2 headache)

Results

Specific Medical Events

- Shuttle Missions
  - Five medical conditions were underestimated by the model (nasal congestion, hip sprain/strain, constipation, early insomnia, and CO_2 headache)
  - One condition was overestimated by the model (space motion sickness)
  - Space adaptation headache and paresthesias were underestimated in some missions and overestimated in others

Results

- Resource Utilization
  - Only available for Shuttle missions
  - Only pharmaceutical usage on Shuttle missions was reliably tracked
  - There are 204 resources in the model
  - The alpha level was set at 0.0002 (0.05/204)
  - Ten pharmaceutical resources were underestimated by the model
  - Eleven pharmaceutical resources were overestimated by the model

Results – Resource Utilization

Pharmaceutical Resources Underestimated on one or more simulations

Resource	Overall p-value
Afrin	0
Ambien	0
Double Antibiotic Ointment	0
Dulcolax Suppository	0
Sonata	0
Bacitracin	1.51E-08
Triamcinolone Cream	4.5E-08
Aspirin	8.75E-07
Dulcolax Tablet	7.8E-06
Claritin	2.01E05

Results – Resource Utilization

Pharmaceutical Resources Overestimated on one or more simulations

Resource	Overall p-value
Phenergan Tablet	1E-44
Tylenol	5.69E-32
Sudafed	7.43E-23
Phenergan Injectable	4.76E-21
Afrin	1.95E-17
Ibuprofen	2.17E-13
Milk of Magnesia	7.49E-07
Ambien	7.49E-07
Dulcolax Tablet	1.01E-05
Benadryl Capsule	1.42E-05
Povidone Iodine Swabs	1.71E-05

Discussion

- For both ISS and Shuttle missions, the total number of medical events expected was accurately predicted by the model
- For both ISS and Shuttle missions, specific medical events were forecast well by the model
- Shuttle medical resource utilization was well predicted by the model

Limitations

- Limited number of ISS missions
- Missing or incomplete historical mission data
- Model baselined to ISS medical resources when analyzing Shuttle pharmaceutical utilization

Conclusions

- This analysis provides strong evidence for the validity of the IMM in predicting medical event occurrences and resource utilization for ISS and Shuttle missions
- The model results were validated by historical mission data that have not been used in the model
- A small percentage of medical conditions and medical resource utilization were under or over predicted by the model
- These differences between model output and historical mission data can be used to improve model input data and the accuracy of predicted outcomes

Questions?