Prediction of outcomes in critically ill infants

N. Ambalavanan MD
Professor, Department of Pediatrics
University of Alabama at Birmingham
ambal@uab.edu

Disclosure statement: Nothing to disclose

Outline of talk
- Why predict outcomes?
- How do we predict outcomes?
- Examples from recent studies:
  - Prediction of death or handicap in neonates with hypoxic-ischemic encephalopathy
  - Outcome trajectories
- Summary

"Prediction is hard, especially about the future"
Physicist Niels Bohr (1885-1962)

"Medicine is a science of uncertainty and an art of probability"
Sir William Osler

Basis of clinical practice

<table>
<thead>
<tr>
<th>Basis for clinical decisions</th>
<th>Marker</th>
<th>Measuring device</th>
<th>Unit of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence</td>
<td>Randomized controlled trial</td>
<td>Meta-analysis</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>Vehemence</td>
<td>Radiance of white blood cells</td>
<td>Luminometer</td>
<td>Optical density</td>
</tr>
<tr>
<td>Elusion (or elegance)</td>
<td>Smoothness of tongue or nap of sit</td>
<td>Tonometer</td>
<td>Diastolic</td>
</tr>
<tr>
<td>Phrenia</td>
<td>Level of religious fervor</td>
<td>Sentient to measure angle of perjuudication</td>
<td>International units of potency</td>
</tr>
<tr>
<td>Difference</td>
<td>Level of gilson</td>
<td>Naloximeter</td>
<td>Sights</td>
</tr>
<tr>
<td>Nerveousness</td>
<td>Litigation profile level</td>
<td>Every conceivable test</td>
<td>Bank balance</td>
</tr>
<tr>
<td>Confidence</td>
<td>Braden</td>
<td>Sust test</td>
<td>No sweat</td>
</tr>
</tbody>
</table>

* Applies only to surgeons.

Isaacs D, Fitzgerald D. Seven alternatives to evidence based medicine. BMJ 1999:319: 1618

Medical uncertainty

- Distinguish information from intrinsic uncertainty
- Recognize that barriers to acknowledge exist
- Recognize uncertainty in your learners

- Uncertainty can lead to delays and harms in patient care
- Uncertainty can lead to increased costs of care

- Acknowledge the "certainty of uncertainty" in medical practice
- Create multiscope plans with built in contingencies
- Utilize shared-decision making with the patient when uncertainty arises

Wray CM and Loo UK. J Graduate Medical Education Dec 2015
Intrinsic Uncertainty

| Acknowledge when personal uncertainty arises, recognizing its inherent nature in the practice of medicine.1,2 |
| Realize that failure to acknowledge uncertainty can lead to excess testing, increased costs, higher referral rates, unnecessary hospitalizations, delays in patient care, and patient harm.3,13,14 |
| Become familiar with and practice using the SNAPPS method for case presentations.15 |
| Assess one’s own attitude to and increase awareness of uncertainty by taking a self-assessment using the PRU scale.5 |
| Always keep a broad differential in mind no matter how certain one is of the diagnosis, and build contingency plans for the workup and treatment of patients.16 |

SNAPPS

- Summarize relevant patient history and findings;
- Narrow the differential;
- Analyze the differential;
- Probe the preceptor about uncertainties;
- Plan management; and
- Select case-related issues for self-study.

Disclosure

Namasivayam Ambalavanan MD, has no relationships with commercial companies to disclose.

Learning Objectives

At the end of this presentation the participant will be able to:

- Understand why prediction of clinical outcomes is important
- Understand how we can predict outcomes
- Know what we should look for in a model used to predict outcomes

Why predict outcomes?

Why predict outcomes? – Clinical uses

- Determining prognosis in individual patients: "One size does not fit all": A combination of clinical characteristics better predicts outcome than one variable E.g.
  - Overall survival at 25 weeks gestation is approximately 60%.
  - 25° white male neonate, no antenatal steroids, born by SVD, Apgars 1 & 2, birth weight 520 g, survival =20%
  - 25° black female neonate, antenatal steroids +, C/section, Apgars 6&7, birth weight 680 g, survival =80%
- Adults: Staging and grading of cancer + other morbidity

Why predict outcomes? – Clinical uses

- Why is it necessary to accurately estimate prognosis in individual patients?
  - “Self-fulfilling prophecies”
    - OBs and pediatricians who underestimate neonatal survival are less likely to resuscitate or use mechanical ventilation, inotropes, or other standard therapies
  - Adults:
    - Withholding care in intracerebral hemorrhage: Realistic compassion or self-fulfilling prophecy? 
      - Alejandro A. Ramirez and Michael N. Drainier
      - Neurology 2007;68:1647-1648
      - DOI: 10.1212/wnl.0000000000000069
  - Controlling for population differences in clinical trials
Why predict outcomes? – Administrative uses

- Resource allocation
- Benchmarking
- Quality improvement

How do we predict outcomes?

1. Intuitive prediction based on past experience and training
2. Statistical prediction model developed in similar scenarios
3. Combination of (1) and (2): Modification of statistical prediction model based on professional experience and training

Which is better? Intuition or Statistical models?

- Grove and Meehl:
  - Clinical vs. Statistical Prediction:
    - n=135 studies.
    - Clinical Prediction was superior in < 5%
- Swets, Monahan, and Dawes:
  - Compared all three methods.
  - In medicine, Statistical and sometimes Combination (Clinical + Statistical) models were superior.
  - In no case was Clinical Prediction superior.

Why are statistical models better?

- Humans have the unique capacity to observe, but not an unique capacity to predict on the basis of integration of observations
- Virtually all observations can be coded quantitatively
- Possible factors:
  - Statistical models
    - Same conclusion for a given data set (no fatigue, fluctuations in judgment)
    - Variables contribute to conclusions based on their actual predictive power (e.g. multiple regression uses only predictive variables, weighted in accordance to their independent contribution)
  - People
    - Have difficulty distinguishing valid vs. invalid variables, and develop false beliefs about associations between variables
    - Indulge in self-fulfilling prophecies
    - Over-confidence in clinical judgment - Known outcomes more predictable than in advance, past predictions mistakenly recalled as consistent with actual outcomes


---

Footnotes:
Scoring systems in the ICU

- Glasgow Coma Scale
- Adult:
  - Acute Physiology and Chronic Health Evaluation (APACHE) (I, II, III)
  - Simplified Acute Physiology Score (SAPS) II
  - Multiple Organ Dysfunction (MOD) score
  - Mortality Prediction Model (MPM)
  - Sequential Organ Failure Assessment (SOFA)
  - Logistic Organ Dysfunction Score (LODS)
  - Therapeutic Intervention Scoring System (TISS)
- Pediatric:
  - Paediatric Risk of Mortality (PRISM) (Pre-ICU, II, III)
  - Paediatric Index of Mortality (PIM)
- Neonatal:
  - Apgar score
  - CRIB
  - SNAP, SNAP-PE

AUC of ROC curve about 0.86-0.87; Wide variation in assessment


What do we look for in a model?

- Physicians reluctant to use a model regardless of accuracy unless they believe in it
- Clinical credibility of a model:
  - All clinically relevant data should have been evaluated (e.g. smoking for MI)
  - Simple to obtain required data, which is reliable
  - Avoidance of arbitrary thresholds
  - Model structure apparent, and predictions should make sense
  - Simple to use model – printed tree or clinical algorithm more likely to be used than a computer for complex calculations


Show me…

- Evidence of accuracy:
  - Correct classification rates (vs. Physicians)
  - Low false (+) rate
  - Low false (-) rate

- Evidence of generalizability:
  - Separate testing on new test set
  - Definitions should be accepted
  - Not derived from biased retrospective databases, but prospective studies

- Evidence of clinical effectiveness:
  - Effects of providing prognosis needs to be evaluated
  - Studies should eliminate biases such as the checklist effect, Hawthorne effect, and contamination (each provides 10-20% improvement)


Imperfect prediction is still useful

- Probability of mortality = 10% vs. 90%
- Probability of handicap = 20% vs. 90%

Statistical Prediction Models

- Aim is to build model using observed patient data and outcomes, then test in new patients / different subset of data

- Standard:
  - Regression analysis: Logistic, Linear etc
  - Cox proportional hazards model

- Non-standard:
  - Neural networks
  - Decision trees e.g. Categorization and regression trees (CART), Chi Square Automatic Interaction Detection (CHAID)
  - Generalized additive models
  - Hybrids
How well do these models work?

- **Performance measures:**
  - **Accuracy Measures:**
    - Discrimination-based Measures: Error-rate; Sensitivity, Specificity, **ROC analysis**
    - Generalized measures: Mean absolute error; Brier Score; \( R^2 \); Log-likelihood
  - **Precision**
    - Calibration: global; local and Hosmer-Lemeshow
    - Indirect measures of precision
    - Integration: the Cox approach

- **Identifying thresholds and ROC curves**

  Eye pressure vs. Glaucoma

  ROC Curve for CRP vs. BPD/Death

  Better decisions through science. Scientific American, 283, 70-76.

- **Comparing ROC curves**

  Better decisions through science. Scientific American, 283, 70-76.

- **Statistical Prediction Models**

  - **Standard Models:**
    - **Multi-variable Regression Models**
      - Identification of relationships between many independent variables and a dependent variable.
      - Set of independent variables explains a proportion of variance (\( R^2 \)), and establishes relative predictive importance (by comparing beta weights)
    - Relative predictive importance can be established by dropping single independent variables, and checking \( R^2 \) change
Common types of regression analyses

- Multiple Linear Regression
- Stepwise Multiple Regression (regression calculated either adding or eliminating variables one at a time, until minimal change in R²)
- Logistic Regression:
  - Dependent variable is dichotomous and independents are any type
  - Estimation of the probability of a certain event occurring – maximum likelihood estimation after transforming dependent into a logit variable (natural log of the odds of the dependent occurring or not)
  - Standard: Binary logistic regression; Variants: Multinomial, Ordinal

Example of a Regression Equation

- Logit P (survival 0, death 1) =
  \[ 7.42 - (0.006 \times \text{Birthweight in g}) - (0.14 \times \text{Apgar 5'}) - (0.67 \times \text{Race W0 B1}) - (0.12 \times \text{Gest Age in wks}) - (0.73 \times \text{Antenatal Steroids 0 or 1}) - (0.49 \times \text{RDS 0 or 1}) + (0.81 \times \text{Multiple births 1,2,3 etc}) \]
- Items with a "-" sign improve survival and those with a "+" sign decrease it

Statistical Prediction Models

- Non-Standard Models:
  - Neural Networks
  - Classification Trees

Conventional vs. Neural Network

Neural networks

- Non-parametric pattern recognition techniques (software algorithms) that can recognize complex or non-linear relationships between independent and dependent variables.
- This is achieved by "training" the network with a training set consisting of the independent variables and the known dependent outcomes.
- Once trained, the neural network can be used for prediction of outcomes in a separate test or validation set.
**Decision Trees**

- Classification and Regression Tree (CART) analysis
  - decision trees to predict outcomes by recursive partitioning and automatic selection of optimal cut-points of variables.
  - Rather than a prediction equation similar to that produced by regression models, recursive partitioning produces a classification tree with a series of binary splits

**Advantages of CART analysis**

- No predesigned/empirical cut-points
  - optimal cut-point for each variable is identified
- More important variables are higher on the decision tree, facilitating the identification of the relative importance of variables.
- Decision trees are also easier for clinicians to use
  - do not require equations or calculations
  - follow the tree from beginning to end, with decisions being made at each node based on available clinical data

**Example of a CART model**

![Example of a CART model](image_url)

**Which babies should be cooled?**

![Which babies should be cooled?](image_url)

**Questions:**

- In infants diagnosed with HIE,
  - Which early perinatal factors determine prognosis?
  - Do early perinatal factors help determine which infants would benefit the most from hypothermia?

**Does cooling work?  Yes!**

- Whole Body Hypothermia (Shankaran et al. NEJM 2005)
  - Death/mild-severe disability ↓ 18% (62 to 44%, p=0.01)
- Head cooling (Gluckman et al. Lancet 2005)
  - Death/severe disability ↓ 11% (66 to 55%, p=0.1)
  - Benefit if less severe aEEG (OR 0.42, 0.22-0.8, p<0.01)

**Early neurologic evaluation**

<table>
<thead>
<tr>
<th>Category</th>
<th>Signs of Encephalopathy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moderate</td>
</tr>
<tr>
<td>Consciousness</td>
<td>Lethargic</td>
</tr>
<tr>
<td>Spontaneous activity</td>
<td>Decreased</td>
</tr>
<tr>
<td>Posture</td>
<td>Distal flexion, complete extension</td>
</tr>
<tr>
<td>Tone</td>
<td>Hypotonia</td>
</tr>
<tr>
<td>Primitive reflexes</td>
<td>Suck</td>
</tr>
<tr>
<td></td>
<td>Moro</td>
</tr>
<tr>
<td>Autonomic system</td>
<td>Pupils</td>
</tr>
<tr>
<td></td>
<td>Heart rate</td>
</tr>
<tr>
<td></td>
<td>Respiration</td>
</tr>
</tbody>
</table>

**Death/Disability Scoring**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level of variable</th>
<th>Odds Ratio</th>
<th>Score (Range 5-65)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posture</td>
<td>Normal</td>
<td>0.937</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Distal flexion</td>
<td>0.401</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Decerebrate</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Spontaneous activity</td>
<td>Normal/Decreased</td>
<td>0.147</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Base deficit of first postnatal blood gas</td>
<td>&lt;15</td>
<td>0.073</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15-22</td>
<td>0.304</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>&gt;22</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Apgar score at 5 minutes</td>
<td>7-10</td>
<td>0.082</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4-6</td>
<td>0.676</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>0-3</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Chronic hypertension/ pre-e/Eclampsia</td>
<td>Yes</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

N=172; 22 infants deleted due to missing values, c 0.873; R-square 0.399

**Score ranges**

- **Control group**
  - Maximum Score: 65
  - Minimum Score: 5

- **Hypothermia group**
  - Maximum Score: 52
  - Minimum Score: 23

**Interpretation of score ranges**

- **Control group**
  - Maximum Score: 65
  - Minimum Score: 5

- **Hypothermia group**
  - Maximum Score: 52
  - Minimum Score: 23

**How accurate are these models?**

<table>
<thead>
<tr>
<th>Death / Disability scoring</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Correct classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe encephalopathy</td>
<td>80</td>
<td>75</td>
<td>48</td>
<td>93</td>
<td>76</td>
</tr>
<tr>
<td>Death</td>
<td>63</td>
<td>78</td>
<td>47</td>
<td>87</td>
<td>74</td>
</tr>
</tbody>
</table>

In infants receiving hypothermia, at optimal cutoff of 33 for Death / Disability scoring and 18 for Death scoring.

**Early neurologic evaluation**

<table>
<thead>
<tr>
<th>Category</th>
<th>Signs of Encephalopathy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moderate</td>
</tr>
<tr>
<td>Consciousness</td>
<td>Lethargic</td>
</tr>
<tr>
<td>Spontaneous activity</td>
<td>Decreased</td>
</tr>
<tr>
<td>Posture</td>
<td>Distal flexion, complete extension</td>
</tr>
<tr>
<td>Tone</td>
<td>Hypotonia</td>
</tr>
<tr>
<td>Primitive reflexes</td>
<td>Suck</td>
</tr>
<tr>
<td></td>
<td>Moro</td>
</tr>
<tr>
<td>Autonomic system</td>
<td>Pupils</td>
</tr>
<tr>
<td></td>
<td>Heart rate</td>
</tr>
<tr>
<td></td>
<td>Respiration</td>
</tr>
</tbody>
</table>
Classification trees

Outcome Trajectories in Extremely Low Birth Weight (ELBW) Infants

Background

- The probability of death or survival with handicap is better estimated by consideration of:
  - Gestational age
  - Gender
  - Exposure to antenatal steroids
  - Multiple birth
  - Birth weight


Rationale

- Consider two infants:
  - 23 wk, 450 g, male, one of triplets, no antenatal steroids
    - Predicted outcome using NRN outcome estimator: 93% death; 98% death or moderate-severe NDI
  - 25 wk, 750 g, female, singleton, antenatal steroids +
    - Predicted outcome using NRN outcome estimator: 15% death; 40% death or moderate-severe NDI

Prediction at Birth of Death/NDI

Is this probability constant over the hospital course?
Concept

- Probability of an outcome is not constant but is a dynamic "outcome trajectory" with modification of the initial probability estimate by subsequent variables

- Outcome predictions at different time points after birth may be improved by considering additional events (e.g. IVH, BPD, NEC, ROP) that occur before that point in time

Available at: https://neonatal.rti.org/OTEstimator/
Summary

- Prognosis may be determined with reasonable accuracy by statistical models
- No model is 100% accurate
- Avoid empirical cut-points
- Development of these models occasionally yields insight into underlying predisposing factors
- The only useful data are data that are used – if data are collected and not used, are they really data?

“It is not certain that everything is uncertain”
- Blaise Pascal (1623-1662)