USING QALYS IN CRITICAL CARE TRIALS

Niall D. Ferguson, MD, FRCPC, MSc
Director, Critical Care Medicine
University Health Network & Mount Sinai Hospital
Associate Professor of Medicine & Physiology
Interdepartmental Division of Critical Care Medicine
University of Toronto
Time of Mortality Reporting:
All primary-endpoint mortality studies during past decade in 5 major medical journals & 5 CCM journals

- 72% reported follow-up up to 30 days
- 11% reported follow-up up to 60 days
- 8% reported follow-up up to 90 days
- 2% reported follow-up up to 180 days
- 2% reported follow-up beyond 180 days

- 7 studies reported follow-up at/beyond 180 days
- Longest follow-up: 365 days (n=3)

* Scales, Friedrich, Kiss, Sibbald, Redelmeier – unpublished results
Choosing a primary endpoint: Why focus on mortality?

- Mortality is common in critically ill patients, so it is *feasible* to measure
- Survival is unambiguous
  - Convincing
  - Less bias in determination
  - Easy to interpret and to communicate (“absolute risk reduction”, NNT)
- Mortality differences often required by regulatory and licensing bodies (FDA, Health Canada)
- Dichotomous outcome
Limitations of using 28-day mortality as a primary endpoint:

- Does not capture mortality that commonly occurs after this arbitrary time point
- Does not reflect the burden of morbidity in survivors
  - If baseline mortality is decreasing in critical illness, treating and measuring morbidity becomes increasingly important
The effect of drotrecogin alfa (activated) on long-term survival after severe sepsis*

(Crit Care Med 2004; 32:2199–2206)

Derek C. Angus, MB, ChB, MPH; Pierre-Francois Laterre, MD; Jeff Helterbrand, PhD; E. Wesley Ely, MD, MPH; Daniel E. Ball, BS, MBA; Rekha Garg, MD; Lisa A. Weissfeld, PhD; Gordon R. Bernard, MD; for the PROWESS Investigators

96.5% Consent Rate
Among contacted pts

Unable to contact
11% of patients

Number at Risk

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>840</td>
<td>850</td>
</tr>
<tr>
<td>3</td>
<td>487</td>
<td>509</td>
</tr>
<tr>
<td>6</td>
<td>467</td>
<td>475</td>
</tr>
<tr>
<td>12</td>
<td>441</td>
<td>447</td>
</tr>
<tr>
<td>18</td>
<td>391</td>
<td>401</td>
</tr>
<tr>
<td>24</td>
<td>260</td>
<td>256</td>
</tr>
<tr>
<td>30</td>
<td>147</td>
<td>154</td>
</tr>
</tbody>
</table>
Quality-adjusted Survival in the First Year after the Acute Respiratory Distress Syndrome Syndrome

DEREK C. ANGUS, AMJAD A. MUSTHAFA, GILLES CLERMONT, MARTIN F. GRIFFIN, WALTER T. LINDE-ZWIRBLE, TONY T. DREMSIZOV, and MICHAEL R. PINSKY

Critical Care Medicine Division, Department of Anesthesiology and Critical Care Medicine and the Center for Research on Health Care, University of Pittsburgh, Pittsburgh, Pennsylvania; Redding Critical Care Medical Group, Redding, California; and Health Process Management, Inc., Doylestown, Pa.

Internet address: www.atsjournals.org
Disadvantages of using a dichotomous outcome

- Increases sample size requirements considerably
  - Endpoints may not occur as frequently, and this can further affect sample size requirements
- Makes quantification of therapeutic benefit or harm more difficult
  - No ‘gradient of response’
Other Possible Outcomes

- **Physiological endpoints**
  - Many examples of therapies that improve physiological endpoints but are not associated with improved clinical outcomes

- **Surrogate clinical endpoints**
  - *ie.* ICU LOS, duration of mechanical ventilation
  - Not always associated with mortality
  - Mortality interferes with their interpretation
  - Not as robust; often more subject to practice variability
Other Possible Outcomes

- Composite endpoints:
  - ie. death or persistence of MV at 30 days; death or need for dialysis
  - Must have biological plausibility for grouping endpoints
  - Composite endpoint must not be driven by one component
    - ie. need for revascularization versus mortality, non-fatal MI
VFDs were improved by lower tidal volume ventilation, but the improvement was mostly caused by the improved mortality rate...
... Additional instruments allow researchers to transform patients’ values (utilities) for different health states on a single linear scale that ranges from zero to one. Patient utilities can be used to calculate quality-adjusted life years (QALYs) which incorporate the quality and the quantity of life into a single measure ...
Quality-Adjusted Life Years

• Most widely utilised and validated method for combining mortality AND morbidity (HRQL)
• Enables comparisons of outcomes across disciplines / disease states / populations
• Utilities for different health states:
  • global ratings of a given health state
    • ranging from 0 (death) to 1 (perfect health)
  • QALY = global rating multiplied by duration of time spent in that state
Welcome to...

Health Utilities Inc

Health-related quality-of-life

Thanks to all who dropped by the HUI booth for a chat at ISPOR in Washington, DC recently.

Health Utilities Inc (HUIInc) specializes in preference-based (utility) measures of health-related quality of life for use in:
  * describing treatment processes and outcomes in clinical studies;
  * economic evaluations of health care programs;
  * the measurement and monitoring of population health.

HUIInc provides support services for the Health Utilities Index (HUI®) measurement systems.

The HUI® measurement system is a generic, preference-scored, comprehensive system for measuring health status, health-related quality of life, and producing utility scores.

For assistance in navigating the HUIInc web pages see the 'Site Map' for brief descriptions.

To visit our associated academic web site, click the HUG/HUI logo...

(R) HUI Registration: USA 2660116, 2716082; UK 2228611, 2228610; CAN TMA54408, TMA550246.

HUI pages last updated - March 31, 2005
HUI Webpages designed and maintained by John R. Horsman
© Copyright HUInc 1999-2005. All rights reserved.
QALYs in Critical Care

- Advantages:
  - Well-validated in other specialties / disease states
    - Enables comparisons between interventions
    - Applicable to most critical care illnesses
  - Good face validity: provides meaningful measure of patient outcomes
  - Should provide more useful information than mortality alone (mortality is still captured)
  - Continuous measurement: *may* reduce sample size requirements
  - Accepted minimal clinically-important differences have already been established
QALYs in Critical Care

• Disadvantages:
  • Requires longer follow-up of patients
  • Measuring outcomes occurring further downstream from study intervention can actually increase sample size requirements
    • more ‘noise’
    • potential for confounding factors during hospitalization and convalescence that can decrease effect size
  • HUI has not been validated in critical care population - but is being used…
Cost-effectiveness Analysis

More Expensive and More Effective

More Expensive and Less Effective

Cheaper & More Effective

Cheaper and Less Effective
QALYs in Critical Care?

Less Feasible and More Important

Less Feasible and Less Important

More Feasible and More Important

More Feasible and Less Important
INTEGRATING MORTALITY & MORBIDITY OUTCOMES: USING QALYS IN CRITICAL CARE TRIALS

Niall D. Ferguson, Damon C. Scales, Ruxandra Pinto, M. Elizabeth Wilcox, Deborah J. Cook, Gordon H. Guyatt, Holger J. Schunemann, John C. Marshall, Margaret S. Herridge, Maureen O. Meade

American Journal of Respiratory & Critical Care Medicine - In Press
Methods

• Use SF-36 data from ARDS cohort of Herridge et al. to generate utilities
  • Calculate QALYs by integrating utility over time (area under utility-time curve)

• Use this information to:
  • Examine distribution and estimate sample size requirements for RCTs
Calculation of Utilities

• Assigned utility = 0 while in ICU
  • Patients were ventilated for 90% of their ICU days

• Assigned utility = 0 for patients who died from that point forward
  • ICU deaths – total QALY = 0

• Converted raw SF-36 data to utility by calculating the SF-6D (Brazier et al.)
  • ‘Continuous’ measure from 0.29 – 1.0
  • Utilities calculated from f/u data taken at 3, 6, and 12 months
Calculation of QALYs: Area under the Utility-Time Curve

QALY = \[\frac{(a/2 + a)(0.25-x)/2} + \frac{(a+b)(0.25)/2} + \frac{(b+c)(0.5)/2}\]
QALYs to 6 Months

Overall group
Mean: 0.10
SD: 0.11
Median: 0.09
IQR: 0.0 – 0.20

Survivors Only
Mean: 0.19
SD: 0.07
Median: 0.19
IQR: 0.15 – 0.23
QALYs to 12 Months

**Overall group**
- Mean: 0.25
- SD: 0.27
- Median: 0.10
- IQR: 0.0 – 0.49

**Survivors Only**
- Mean: 0.50
- SD: 0.15
- Median: 0.49
- IQR: 0.42 – 0.60
Sample Size Estimations

• **Base-case**
  • Hypothetical intervention
  • 4% ARR in mortality from baseline of 48%
    • Effect size point estimate seen in 2 large ventilation trials in ARDS (LOVS, ExPress)
    • Observed 6-month mortality in ARDS cohort
  • Increase in 0.05 in QALYs per year of follow-up among survivors
    • Minimally important difference

• 5 additional permutations
Sample Size Simulations

1. Generate a binomial distribution \((n_1,p_1)\)
   - \(n_1\) = total sample size in the control group
   - \(p_1\) = probability of dying in the control group (48%)

2. Assign a QALY = 0 for those with an outcome of 1 (dead)

3. For those with an outcome of 0 (alive) generate a QALY distribution that is normal around mean \(\mu_1\) with standard deviation \(\sigma_1\)
   - \(\mu_1\) = observed mean in QALYs in survivors
   - \(\sigma_1\) = observed standard deviation of QALYs in survivors
QALYs to 6 Months

Overall group
Mean: 0.10
SD: 0.11
Median: 0.09
IQR: 0.0 – 0.20

Survivors Only
Mean: 0.19
SD: 0.07
Median: 0.19
IQR: 0.15 – 0.23
Sample Size Simulations

4. Generate a binomial distribution \((n_2, p_2)\)
   - \(n_2 = n_1\) = total sample size in the treatment group
   - \(p_2\) = probability of dying in the treatment group (44%)

5. Assign a QALY = 0 for those with an outcome of 1 (dead)

6. For those with an outcome of 0 (alive) generate a QALY distribution that is normal around mean \(\mu_2\) with standard deviation \(\sigma_2\)
   - \(\mu_2 = \mu_1 + \text{delta QALY}\)
   - \(\sigma_2 = \sigma_1 = \text{observed QALY standard deviation in survivors}\)
Sample Size Simulations

7. Generate a p-value for a 2-sided Mann-Whitney U test comparing generated QALY values in the control (steps 1-3) vs. treatment (steps 4-6) group

- Repeat steps 1 - 7 x 10,000 samples
- Power calculated as percentage of p-values < 0.05 in the 10,000 samples
- Repeat whole process across a spectrum of sample sizes (n1) to find that which achieves 80% power
- Adjust total estimates to account for observed loss to follow-up
- Repeat all of above for each permutation
Cohort Characteristics

• 195 ARDS patients enrolled
  • QALYs at 6 months in 168 (86%)
  • QALYs at 12 months in 159 (82%)

• Medians (IQR):
  • Age: 48 (37-61) years
  • APACHE II: 25 (20-30)
  • ICU LOS: 21 (11-37) days
Results

• Simple sample size estimation for mortality differences in hypothetical groups
  • Control: mortality = 48%
  • Treatment: mortality = 44%

• Sample size required to detect difference (chi square) with beta=0.2, alpha=0.05:
  • 2436 patients per group
## Results: Base-case

<table>
<thead>
<tr>
<th>Intervention Effects</th>
<th>Follow-up duration (months)</th>
<th>QALY $\Delta_{\text{survivors}}$ ($\sigma_{\text{survivors}}$)</th>
<th>Mortality $p_1$-$p_2$</th>
<th>QALY $\Delta_{\text{overall}}$ ($\sigma_{\text{overall}}$)</th>
<th>N per group</th>
<th>N per group adjusted for loss to F/U**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved mortality; Improved survivor QoL</td>
<td>6</td>
<td>0.025 (0.07)</td>
<td>0.04</td>
<td>0.022 (0.11)</td>
<td>500</td>
<td>571</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.05 (0.15)</td>
<td>0.04</td>
<td>0.048 (0.28)</td>
<td>525</td>
<td>630</td>
</tr>
</tbody>
</table>

For all scenarios: alpha (2-sided) = 0.05

*Baseline mortality ($p_1$)=0.30 (for all other scenarios baseline mortality ($p_1$)=0.48)
### Results: Permutations

<table>
<thead>
<tr>
<th>Intervention Effects</th>
<th>Follow-up duration (months)</th>
<th>QALY $\Delta_{\text{survivors}}$ ($\sigma_{\text{survivors}}$)</th>
<th>Mortality $p_1-p_2$</th>
<th>QALY $\Delta_{\text{overall}}$ ($\sigma_{\text{overall}}$)</th>
<th>N per group</th>
<th>N per group adjusted for loss to F/U**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved mortality; Improved survivor QoL; Lower baseline mortality*</td>
<td>6</td>
<td>0.025 (0.07)</td>
<td>0.04</td>
<td>0.026 (0.11)</td>
<td>250</td>
<td><strong>300</strong></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.05 (0.15)</td>
<td>0.04</td>
<td>0.057 (0.27)</td>
<td>275</td>
<td>353</td>
</tr>
<tr>
<td>Improved mortality; Unchanged survivor QoL</td>
<td>6</td>
<td>0 (0.07)</td>
<td>0.04</td>
<td>0.008 (0.11)</td>
<td>3000</td>
<td>3426</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0 (0.13)</td>
<td>0.04</td>
<td>0.020 (0.27)</td>
<td>3000</td>
<td>3603</td>
</tr>
<tr>
<td>Unchanged mortality; Improved survivor QoL</td>
<td>6</td>
<td>0.025 (0.07)</td>
<td>0</td>
<td>0.013 (0.11)</td>
<td>1600</td>
<td>1827</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.05 (0.13)</td>
<td>0</td>
<td>0.026 (0.28)</td>
<td>1800</td>
<td>2162</td>
</tr>
</tbody>
</table>

For all scenarios: alpha (2-sided) = 0.05
*Baseline mortality ($p_1$)=0.30 (for all other scenarios baseline mortality ($p_1$)=0.48)
Results: Permutations

<table>
<thead>
<tr>
<th>Intervention Effects</th>
<th>Follow-up duration (months)</th>
<th>QALY $\Delta_{\text{survivors}}$ ($\sigma_{\text{survivors}}$)</th>
<th>Mortality $p_1-p_2$</th>
<th>QALY $\Delta_{\text{overall}}$ ($\sigma_{\text{overall}}$)</th>
<th>N per group</th>
<th>N per group adjusted for loss to F/U**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worsened mortality; Improved survivor QoL</td>
<td>6</td>
<td>0.025 (0.07)</td>
<td>-0.04</td>
<td>0.0043 (0.11)</td>
<td>48,000</td>
<td>54,816</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.05 (0.13)</td>
<td>-0.04</td>
<td>0.003 (0.27)</td>
<td>50,000</td>
<td>60,053</td>
</tr>
<tr>
<td>Improved mortality; Worsened survivor QoL</td>
<td>6</td>
<td>-0.025 (0.07)</td>
<td>0.04</td>
<td>-0.006 (0.10)</td>
<td>15,000†</td>
<td>17,130†</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-0.05 (0.13)</td>
<td>0.04</td>
<td>-0.008 (0.26)</td>
<td>23,000†</td>
<td>27,624†</td>
</tr>
</tbody>
</table>

For all scenarios: alpha (2-sided) = 0.05

*Baseline mortality ($p_1$)=0.30 (for all other scenarios baseline mortality ($p_1$)=0.48)
Sample Size Summary

• Improved mortality & QOL
  • WIN!!
  • Even better with lower baseline mortality
• Improved mortality or QOL; no change in other
  • No great loss or gain in power; patient-centred endpoint = win!
• Mortality & QOL move in opposite directions
  • Problem!
Conclusions

• QALYs may be a feasible outcome in ICU trials
• Need to examine both components of this ‘composite’ endpoint independently
• Major potential increase in power when both mortality and QOL in survivors are improved by an intervention
• Minimizing loss to follow-up is KEY
• Next steps:
  • Future trials should collect utility data and treat QALYs as a secondary outcome
leading science, leading practice

October 28-31, 2012
Sheraton Centre Hotel, Toronto

www.criticalcarecanada.com