Clinical-Trial Based Cost-effectiveness Analyses of Antipsychotics

Daniel Polsky, PhD[1,2], Jalpa Doshi, PhD[1,2],
Mark Bauer, MD[3], Henry Glick, PhD[1,2]

1. University of Pennsylvania School of Medicine
2. Leonard Davis Institute of Health Economics
3. Providence VA Medical Center & Brown University

Funding Support: NIAAA #R01 AA12664-01

45th NCDEU Meeting
Boca Raton, Florida
June 6, 2005

Presentation Material

Presentation handouts are posted on our website:
www.uphs.upenn.edu/dgimhsr

E-mail:
jdoshi@mail.med.upenn.edu

Forthcoming article:

Polsky D, Doshi JA, Bauer M, Glick H. "Clinical Trial Based Cost-effectiveness Analyses of Antipsychotics." (In Press *American Journal of Psychiatry*)

Purpose

- To familiarize participants with methodological issues in the design, measurement, and analysis of economic assessments in randomized clinical trials (RCT) of antipsychotics
  - Systematic review of all such economic evaluations of
    - second-generation antipsychotic medications (SGAs) relative to
    - first-generation antipsychotic medications (FGAs)
- To summarize the evidence on cost-effectiveness of SGAs relative to FGAs in light of the identified methodological issues

Why Antipsychotics?

- Antipsychotic medications make up one of the fastest growing segments of the rapidly growing pharmaceutical sector
  - Primarily a result of the shift from FGAs to newer SGAs
- Since the newer SGAs are relatively more expensive than the FGAs antipsychotic expenditures have increased several fold over the last decade
  - Medicaid expenditures for antipsychotic medications increased from $484 million in 1995 to $3.73 billion in 2002

Why Cost-effectiveness?

- Given limited health care resources, assessment of the value for cost of SGAs relative to FGAs is critical for resource allocation decisions
- Cost-effectiveness analysis (CEA) is a common method used in the medical literature to address such questions
  - Cost-effectiveness ratio (CER) explicitly expresses tradeoffs among treatment interventions by comparing differences in costs and health benefits achieved
    CER=$\Delta C/\Delta E$ where $\Delta C = Ct_1 - Ct_2$ and $\Delta E = Et_1 - Et_2$
- CEA are frequently based on data from RCTs where prospective economic information is collected alongside clinical endpoints

Literature Review Methods

- Medline search for all studies which included terms related to
  - costs
  - clinical trials
  - schizophrenia treatment
- Search was limited to publications in English, involving human subjects, and published during 1985 to 2003
- Study exclusion criteria
  - Not a randomized trial
  - Did not compare SGAs vs. FGAs
  - Did not collect or analyze patient specific costs
  - Applied clinical trial data in a decision analytic model

Literature Search Results

- Search identified 8 RCT-based cost-effectiveness studies of SGAs vs. FGAs from a total of 6 clinical trials
  - Clozapine vs. Haloperidol
    - Essock, 2000
    - Rosenheck, 1998
    - Rosenheck, 1999
  - Olanzapine vs. Haloperidol
    - Hamilton, 1999
    - Tunis, 1999
    - Rosenheck, 2003
  - Risperidone vs. Haloperidol
    - Chouinard, 1997
  - Olanzapine vs. Risperidone vs. FGAs
    - Jerrell, 2003

Study Conclusions on SGAs vs. FGAs

- 6 of the 8 studies explicitly reported or suggested that SGAs were cost-effective compared to FGAs
- Only 2 studies suggested that SGAs were unlikely to be cost-effective
- Thus, the weight of evidence appears to suggest that compared to FGAs, SGAs are good value for the cost

Are RCT Results the Gold Standard?

- While RCTs are considered to be the gold standard for comparing alternative medical therapies, randomization does not in and of itself guarantee valid or reliable results
- Without appropriate methods of measurement, analysis, and design, economic evaluation using RCTs may be biased, imprecise, or have limited applicability

Seven Issues in RCT-CEA Studies

- Methods of measurement, analysis, and design of cost-effectiveness studies in RCTs
  1) Measurement of costs
  2) Measurement of effectiveness
  3) Analysis of costs
  4) Analyzing uncertainty
  5) Analysis of incomplete cost data
  6) Minimizing loss to follow-up
  7) Threats to external validity

Issue # 1: Measurement of Costs

Measurement of Costs: Recommendations

- Costs that are most appropriate for inclusion in an analysis depend on its "perspective"
  - e.g. insurer's perspective, societal perspective
- Direct health care costs
  - e.g. costs of drugs, hospitalization, and health care personnel
- Direct nonhealthcare costs
  - e.g. time family members spend providing home care, costs of transportation to and from clinic
- Indirect costs
  - e.g. productivity costs, patient time costs

Measurement of Costs: Recommendations

- All studies should assess direct health care costs
- To better serve the broadest set of decision-makers analysts should evaluate a therapy's effect from a societal perspective
  - Separately report differences in health care costs as well as differences in total costs

Measurement of Costs: Findings

- 3 studies included direct healthcare and nonhealthcare costs and indirect costs
- 4 studies evaluated direct healthcare costs alone
- 1 study quantified drug acquisition costs only

Issue # 2: Measurement of Effectiveness

Measurement of Effectiveness: Recommendations

- Ideally, analysts should adopt a single outcome for use in economic evaluations that
  - Allows direct comparison of therapies across number of domains
  - Allows comparison across therapeutic areas and illnesses
  - For which one generally understands how much one is willing to pay to buy a unit of this outcome
- In most areas of health technology assessment, this outcome is the quality-adjusted life year (QALY)
  - Incorporate length of survival and its quality into one measure
  - Values generally range between 0 (death) and 1 (perfect health)
    e.g., a health state with a utility value of 0.8 indicates that a year in that state is worth 0.8 of a year with perfect health

Measurement of Effectiveness: Findings

- Three studies used QALYs as an effectiveness measure
- The remainder used a wide variety of alternative outcome measures
  - Brief Rating Psychiatric Score
  - Positive and Negative Syndrome Scale (PANSS) total scores
  - SF-36 Physical and Mental Health component scores
  - 12 mental health and satisfaction measures
  - Extrapyramidal symptom-free months
  - Measures of disruptiveness and psychiatric symptomatology
  - Composite Health Index for schizophrenia

Measurement of Effectiveness: Findings

- A wide variety of measures of effectiveness have been used in clinical trials evaluating new antipsychotic drugs
  - They do not permit comparability across studies within schizophrenia nor between therapeutic areas
  - There is a limited understanding of willingness to pay for observed changes in these measures
- Current QALY assessment instruments have been validated and may be appropriate for schizophrenia
- To the extent that they may not be appropriate
  - Development of QALY instruments that are more specific to psychosis should be a priority
  - If non-QALY outcomes are preferred to QALYs, then studies of willingness to pay for changes in these non-QALY outcomes are necessary

Issue # 3: Analysis of Costs

Analysis of Costs: Recommendations

- Cost-effectiveness should be based on differences in the arithmetic mean of costs between treatment groups
  - Report means, and not medians, since means are the statistic of interest for policy decisions
  - Indicate whether the observed difference occurred by chance
- Costs of health care are typically characterized by a highly skewed distribution with a long and sometimes awkwardly heavy right tail
- As a result means and common parametric statistical methods for analyzing means such as t-tests of means and ANOVA have been rejected by many researchers

Analysis of Costs: Recommendations

- Violation of normality assumption has led to use of alternate approaches (which are more problematic!)
  - T-test of log costs
  - Nonparametric tests of median costs (Mann Whitney U or Wilcoxon rank sum)
- Statistical inferences from these tests may not be representative of inferences about the arithmetic mean
  - T-test of log costs indicates one can infer that the mean of the logs are different, and thus the geometric means of cost are different
  - Mann Whitney U test indicates one can infer that the medians are different

Analysis of Costs: Recommendations

- If arithmetic means are the most meaningful summary statistic of costs, one should test for significant differences in arithmetic mean costs
- T-test shown to be robust to violations of the normality assumption in large samples
- Non-parametric bootstrap is the preferred statistical test in small to moderately sized samples
  - Test of arithmetic mean that avoids parametric assumptions
  - Estimates the distribution of the observed difference in mean costs
  - Yields a test of how likely it is that 0 is included in this distribution

Analysis of Costs: Findings

- All 8 studies reported measures of arithmetic mean costs in the comparison groups
- However, they demonstrated little agreement in their approach to the statistical analysis of differences in costs
- Only 3 studies performed statistical tests of arithmetic mean costs
  - 2 of the studies reported t-tests of untransformed costs
  - 1 study performed a non-parametric bootstrap to derive 95% confidence intervals around the cost difference in addition to performing t-tests on untransformed and transformed costs

Analysis of Costs: Findings

- The remaining 5 studies analyzed something other than the arithmetic mean
  - 3 studies conducted t-tests on the log transformation of costs
  - 1 performed a non-parametric rank test or Mann-Whitney test
  - 1 study did not conduct a statistical test

Analysis of Costs: Findings

- Illustration of problems with log of cost evaluation (Hamilton, 1999)

	N	Mean	SD
Olanzapine	543	6114	9,672
Haloperidol	260	6502	10,239

  - Using log of costs, Hamilton et al. concluded that costs were significantly different in the two groups (p=0.03)
  - However, calculations on the untransformed scale: Difference in costs: $388 (p=0.6)

Issue #4 Analyzing Uncertainty

Sampling Uncertainty: Recommendations

- Economic outcomes observed in RCTs are the result of samples drawn from the population
- Thus, all studies should report the uncertainty in these outcomes that results from such sampling
- Common measures of this uncertainty are confidence intervals for the CER

Sampling Uncertainty: Findings

- Only 3 studies reported confidence intervals for CERs
- In all three, the point estimates indicated that compared to FGAs, SGAs have acceptable CERs, but the 95% CI contained CERs that -- depending upon one's maximum willingness to pay -- may be unacceptably high (i.e., not cost-effective)
- Thus, in none of the 3 studies could we be 95% confident that SGAs represent good value for the cost compared with haloperidol or other FGAs

Sampling Uncertainty: Findings

- The remaining 5 studies failed to report on the stochastic uncertainty surrounding this comparison
- One study found significant clinical improvements without any significant difference in costs (i.e., they presumed the result represented identification of a dominant strategy)
  - But the 3 studies that reported uncertainty also had similar results, yet their 95% CI suggested they did not have sufficient statistical power to allow one to confidently conclude that one had observed such dominance
- Two other studies found no differences for either costs or clinical outcomes
  - However, estimation of a confidence interval for the CER would still have been useful

Power and Sample Size: Recommendations and Findings

- When designing RCT-based economic studies, investigators should perform power calculations for the cost-effectiveness end point, so that one can understand whether a statistically insignificant result may be due to a lack of power
- Sample sizes across the studies varied from 65 to 817
- Hence, although many studies reported large differences in costs or effects, the lack of statistical significance of these results might be due to these studies being underpowered

Issue #5: Analysis of Incomplete Cost Data

Analysis of Incomplete Cost Data: Recommendations

- Trials in general, and psychiatric trials in particular, are prone to high levels of drop out that leads to right censoring of cost data
- Drop-out may be due to adverse events, patient decision, lack of efficacy, or noncompliance with the clinical protocol
- In each of these cases costs incurred after the time patients are lost to follow-up may be different from costs incurred during the period they were observed

Analysis of Incomplete Cost Data: Recommendations

- Studies should explicitly report the amount of missingness and whether or not there was an indication that the data appear to be missing-at-random
- They should also adopt appropriate statistical methods to address the problems posed by missing data
- If these problems are not appropriately addressed, estimates of the treatment effect will be biased
- Recent interest in the issue of censored cost data has led to the proposal of several analytic methods for addressing censored cost data
  - It is recommended that these methods be used because naïve estimation can lead to serious bias

Analysis of Incomplete Cost Data: Findings

- 5 studies explicitly reported the amounts of missing data
- Large amounts of missing data in a number of studies
  - For example, one reported 37% while another reported 47% of all follow-up data to be missing
  - Termination of follow-up / withdrawal over a 1-year time frame was 83% in two studies using the same clinical trial data
    - Data were unavailable after withdrawal from the RCT

Analysis of Incomplete Cost Data: Findings

- 4 studies used a random-effects repeated measures model for their analyses
  - Allows inclusion of available data from individuals who eventually drop out of the RCT
- 1 study imputed missing values using a linear mixed model
- Underlying assumption: Costs for subjects during the period when they are not observed = adjusted means of the subjects who are observed during the period
  - Results unbiased only if the reasons for missingness are not correlated with the outcome (i.e. missing completely at random)

Analysis of Incomplete Cost Data: Findings

- 1 study with 83% termination of follow-up/withdrawal imputed missing values by use of a per day cost during the period observed for each censored subject
- Underlying assumption:
  (1) costs are missing completely at random and that
  (2) costs are homogenous over time
- Neither assumption may be warranted given that one of the primary reasons for missing data in this study was lack of response in clinical trial
- This violation suggests that the study may have adopted biased methods for addressing missing data

Issue #6: Minimizing Loss to Follow-up

Minimizing Loss to Follow-up: Recommendations

- Ideally one should design studies in such a way that they minimize the occurrence of such data
- Study designs should include plans to aggressively pursue subjects and data throughout the trial
- Psychiatric RCTs should also ensure that follow-up continues until the end of the study period, and that data collection not be discontinued simply because a subject reaches a clinical or treatment stage such as failure to respond
  - Economic impact of these outcomes or events can only be measured if patients are followed beyond the time of their occurrence

Minimizing Loss to Follow-up: Findings

- 5 studies were based on trials that reported explicit strategies for minimizing missing cost data
- 2 studies were based on a clinical trial that used a design that consciously created missing data
  - 40% of participants dropped out of the study during the trial's 6-week acute phase
  - Data collection for the 46-week maintenance phase was intentionally discontinued for those patients who did not meet a pre-defined level of treatment response
- 1 study did not indicate what they did or did not do to prevent missing data

Issue #7: Threats to External Validity

Threats to External Validity: Recommendations

- Frequently, the priority of the design of RCTs of psychopharmacologic agents is to generate an internally valid measure of the efficacy of treatment
- Given that the primary purpose of CEA is to inform real-world decision-makers about how to respond to real-world health care needs, design of CEA studies should in addition consider issues of external validity

Threats to External Validity: Recommendations

- In this section, we address selected methods of RCT design that specifically relate to maximizing the usefulness of CEA from clinical trials
  - Intention-to-treat analysis
  - Time frame
  - Sample inclusion criteria

Intention-to-treat Analysis: Recommendations and Findings

- In real world settings, economic questions relate to treatment decisions (e.g., whether to prescribe an SGA), not whether the patient received the drug prescribed nor whether, once they started the prescribed drug, they were switched to other drugs
- Hence, costs and benefits associated with these later decisions should be attributed to the initial treatment decision
  - i.e. cost-effectiveness analyses in RCTs should adopt an intention-to-treat (ITT) design
- Only 4 of the eight studies used an ITT approach

Time Frame: Recommendations and Findings

- RCTs of chronic treatments are time-limited while treatment for chronic conditions is not
  - Long term use may yield outcomes that generally cannot be observed by use of shorter time frames
  - Cost-effectiveness ratio may be heterogeneous with time of follow-up
- Making decisions based solely on results observed within short-term RCTs may be inappropriate.
  - One approach to addressing this is via decision analysis
- The longest follow-up among the 8 studies was 2-years, while most had follow-ups of 1-year or less

Sample Inclusion Criteria: Recommendations and Findings

- Many RCTs may employ study samples that do not resemble the more heterogeneous population found in general practice that decision-makers must consider when making resource allocation decisions
- 3 studies were limited to schizophrenic patients with both treatment resistance and high inpatient use
  - The efficacy of FGAs and SGAs may be different in such patients as compared to the general schizophrenic population
- Antipsychotics may be used for indications for which no economic evidence of value for the cost exists
- In the VA system 43% of patients prescribed SGAs received them for the treatment of psychiatric illnesses other than schizophrenia or schizoaffective disorder

Validity of the Conclusions of Reviewed Economic Evaluations of Antipsychotics

Clozapine vs. FGAs (3 studies)

- All 3 studies concluded that for treatment resistant patients with high hospital use clozapine had lower costs and better outcomes over some effectiveness domains
  - Generally these studies were well designed, but none reported a statistically significant difference in costs between treatment groups (issue #3)
  - One study reported 95% confidence about the cost-effectiveness result only among the small subset of patients with the highest levels of inpatient service use (issue #4)
  - However, given that samples in these studies were limited to schizophrenic patients with both treatment resistance and at least 30 days of hospitalization, their results are unlikely to be applicable to the majority of non-treatment resistant schizophrenics (issue #7)

Olanzapine vs. FGAs (3 studies)

- 2 studies found olanzapine reduced costs and improved quality of life compared to haloperidol
  - One year cost-effectiveness results of these two studies may be severely biased
  - Both studies were based on the same RCT which discontinued patients who either failed to respond by week 6 or changed their treatment regimen (issues #5 and #6)
  - Less than 17% of those randomized were still in the study by the end of the first year
- The third study found no significant advantages in effectiveness of olanzapine and higher costs when compared with haloperidol
  - The study was based on a trial that followed both treatment non-responders and those who changed their treatment regimen
  - In this study, 59% of the subjects remained in the trial by the end of the first year

Risperidone vs. FGAs (2 studies)

- 1 study reported a favorable cost-effectiveness ratio for risperidone
  - Study included drug acquisition costs alone (issue #1)
  - Study did not report a statistical test for differences in outcomes (issue #3)
  - However, results based on a small sample size (22 risperidone patients and 21 haloperidol patients followed for 8 weeks) should lead to a lack of statistical significance for their comparative findings (issue #4)
- The second study comparing risperidone and olanzapine with FGAs found both SGAs to have no advantages in outcomes but higher mental health treatment costs than FGAs
  - However, this conclusion was limited by small sample sizes and large loss-to-follow up (issues #4 and #5)

Conclusions

Conclusion

- Economic claims made by the authors of a number of RCT-based economic evaluations have generally - but not unanimously - been favorable for SGAs over FGAs
- However, the methodological problems we have identified raise questions as to the quality of the evidence behind those claims
- Our review suggests that currently there is no clear evidence that atypical antipsychotics generate cost savings or are cost-effective in general use among all schizophrenic patients

Conclusion

- Identified methodological problems are not limited to the antipsychotic literature alone, but found to persist in most areas of medicine
- However, in light of the exponential growth in atypical antipsychotic expenditures over the last few years, the poor quality of the economic evidence supporting the value of SGAs over FGAs takes on added significance

Conclusion

- Clinicians, administrators, insurers, and other stakeholders should recognize that there is a need for comprehensive Phase IV studies that compare SGAs to FGAs
- The goal of these studies should be to determine cost-effective treatment strategies and thus provide optimal input into future decision-making processes

Conclusion

- Improved economic evidence may lead to cost-effectiveness data playing a greater role in insurance coverage and formulary decisions for antipsychotics
- Future RCT-based economic evaluations of current SGAs versus FGAs, of newer antipsychotic agents, and of other psychotropic therapies should attempt to address the priority issues identified here to enhance the validity of their findings and ensure their usefulness to decision-makers