Quality Assessment of Published Health Economic Analyses from South America

Márcio Machado, Michael Iskedjian, and Thomas R Einarson

BACKGROUND: Health economic analyses have become important to healthcare systems worldwide. No studies have previously examined South America’s contribution in this area.

OBJECTIVE: To survey the literature with the purpose of reviewing, quantifying, and assessing the quality of published South American health economic analyses.

METHODS: A search of MEDLINE (1990–December 2004), EMBASE (1990–December 2004), International Pharmaceutical Abstracts (1990–December 2004), Literatura Latino-Americana e do Caribe em Ciências da Saúde (1982–December 2004), and Sistema de Informacion Esencial en Terapéutica y Salud (1980–December 2004) was completed using the key words cost-effectiveness analysis (CEA), cost-utility analysis (CUA), cost-minimization analysis (CMA), and cost-benefit analysis (CBA); abbreviations CEA, CUA, CMA, and CBA; and all South American country names. Papers were categorized by type and country by 2 independent reviewers. Quality was assessed using a 12 item checklist, characterizing scores as 4 (good), 3 (acceptable), 2 (poor), 1 (unable to judge), and 0 (unacceptable). To be included in our investigation, studies needed to have simultaneously examined costs and outcomes.

RESULTS: We retrieved 25 articles; one duplicate article was rejected, leaving 24 (CEA = 15, CBA = 6, CMA = 3; Brazil = 9, Argentina = 5, Colombia = 3, Chile = 2, Ecuador = 2, 1 each from Peru, Uruguay, Venezuela). Variability between raters was less than 0.5 point on overall scores (OS) and less than 1 point on all individual items. Mean OS was 2.6 (SD 1.0, range 1.4–3.8). CBAs scored highest (OS 2.8, SD 0.8), CEs next (OS 2.7, SD 0.7), and CMAs lowest (OS 2.0, SD 0.5). When scored by type of question, definition of study aim scored highest (OS 3.0, SD 0.8), while ethical issues scored lowest (OS 1.5, SD 0.9). By country, Peru scored highest (mean OS 3.8) and Uruguay had the lowest scores (mean OS 2.2). A nonsignificant time trend was noted for OS ($R^2 = 0.12$; $p = 0.104$).

CONCLUSIONS: Quality scores of health economic analyses articles published in South America were rated poor to acceptable and lower than previous research from other countries. Thus, efforts are needed to improve the reporting quality of these analyses in South America. Future research should examine the region’s level of expertise and educational opportunities for those in the field of health economics.

KEY WORDS: health economics, literature analysis, quality assessment, South America.
ed that a lack of expertise could have been a contributing factor to lack of adherence to guidelines.

Health economics has been largely developed in Australia, Britain, Canada, Europe, and the US. Although the population of South America is currently approximately 370 million people, no guidelines have appeared from that region. We therefore undertook this research to determine what South American economic analyses have been published and to assess their quality.

**Methods**

The countries of interest for this research included all those within continental South America including Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, and Venezuela. We examined all articles that presented health economic analyses. To qualify, studies must have examined a drug, treatment, procedure, program, or medical device against a relevant comparator with respect to both costs and outcomes. Papers must have been published in a healthcare journal; however, no restriction was placed on date or language. Presentations at meetings, colloquia, and studies only presented in abstract form were not accepted.

To locate these articles, we performed a search from inception to December 2004 of pertinent databases including MEDLINE (1990–December 2004), EMBASE (1990–December 2004), International Pharmaceutical Abstracts (1990–December 2004), Literatura Latino-Americana e do Caribe em Ciências da Saúde (1982–December 2004),* and Sistema de Información Esencial en Terapéutica y Salud (1980–December 2004).* The latter 2 databases focus on health-related literature in the Spanish and Portuguese languages, particularly from Latin America. Key words employed in the search included cost-effectiveness analysis (CEA), cost-utility analysis (CUA), cost-minimization analysis (CMA), and cost-benefit analysis (CBA), as well as all South American country names. References from retrieved papers and reviews of the topic were searched for further possible studies. Two reviewers independently identified potential articles, with discrepancies resolved through consensus discussion. In the case of failure to achieve consensus, a third reviewer was appointed to adjudicate.

Accepted articles were evaluated using a modified version of the 13 item checklist previously described and used by our group. The only modification consisted of deleting the ninth question, “Is the evaluation suitable if made within a clinical trial?” This question was considered unimportant since no clinical trial was found in any of the reviewed articles. The final number of items available in our checklist was 12.

The total possible score for each item was 4 (4 = good, 3 = acceptable, 2 = poor, 1 = unable to judge, 0 = completely unacceptable). If an item was considered not applicable, it was labeled as such and not included in the evaluation. Thus, in the initial calculation, articles could earn a score ranging from 0 to 48 points. To arrive at the final score for each article, the total number of points awarded was divided by the number of pertinent questions. Final scores for articles therefore ranged from 0 to 4, and these article scores could be interpreted in a manner similar to that used with the individual items. When interpreting means for articles or across all studies, the midpoints between scores were taken into consideration (ie, “good” included all scores from 3.5 to 4.0, “acceptable” included 2.5 to <3.5, “poor” included 1.5 to <2.5).

We correlated the studies’ mean overall score (OS) with the raters’ Global Impression (GI). GI was defined as the rater’s personal overall global estimate of the quality score for each study. Correlation between OS and GI showed how well the rater’s personal views of the study aligned with the values obtained by analyzing and averaging the scores for all items.

Inter-rater reliability was tested using a sample of identical articles between the raters. Each rater assessed a set of 3 papers, and scores were compared. Adequate agreement was achieved when raters obtained equal scores or differences of less than 1 point for individual items and 0.5 for OSs.

Authors’ language proficiency was as follows: native English (TRE and MI), advanced English (MM), native Portuguese (MM), advanced Spanish (MM), and intermediary Spanish (TRE and MI).

Descriptive statistics were used to describe the data. Included were the mean, range, standard deviation, mode, and minimum and maximum values. Differences in mean OS for Latin American publications and previous quality assessment from other countries were assessed using Student’s t-test. Categorizations of quality scores between South American publications and those from other countries were compared using a Mann–Whitney U test. Global impression scores and OSs were correlated using Pearson’s r. Group differences by type of analysis, country, languages, and funding source were analyzed using a Kruskal–Wallis test.

**Results**

A total of 25 studies were retrieved; 1 duplicate was rejected, leaving 24 articles for the quality assessment.* Seventeen articles were in English (OS 2.7, SD 0.8), 4 in Spanish (OS 2.5, SD 0.6), and 3 in Portuguese (OS 2.6, SD 0.5). There were no statistically significant differences among the scores across languages (Kruskal–Wallis test; $\chi^2 = 0.32; p = 0.85$).

The cumulative number of publications has increased exponentially over time, from 1 in 1984 to 24 in 2004. A description of included studies and their main characteristics (ie, authors, alternatives compared, perspective, types of outcomes and costs) can be found in Table 1. Only 8 studies described their funding source, which included 3 pharmaceutical companies (OS 3.1, SD 1.3), 3 government agencies (OS 2.8, SD 0.5), and 2 nongovernmental
Table 1. Description of Selected Articles

<table>
<thead>
<tr>
<th>Reference</th>
<th>Drug/Alternatives Compared</th>
<th>Benefits/Outcomes</th>
<th>Costs</th>
<th>Perspective</th>
<th>Type of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigal (2003)</td>
<td>proposed treatment (acetylsalicylic acid, triptan, acetylsalicylic acid + metoclopramide) vs acetylsalicylic acid only</td>
<td>response rate to treatment</td>
<td>direct: consultations, preventive drugs, acute therapy drugs</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Botto (2003)</td>
<td>ramipril vs placebo</td>
<td>life years saved</td>
<td>direct: ramipril, major cardiovascular events, hospitalization by angina and cardiac insufficiency, diabetes diagnostic tests</td>
<td>societal</td>
<td>CEA</td>
</tr>
<tr>
<td>Burckel (1999)</td>
<td>vaccine vs no vaccine</td>
<td>avoided costs</td>
<td>direct: physician visits, prescription and nonprescription drugs, hospitalization, diagnostic tests</td>
<td>employer</td>
<td>CBA</td>
</tr>
<tr>
<td>Calabró (2003)</td>
<td>voriconazole vs amphotericin B</td>
<td>response rate to treatment</td>
<td>direct: drug acquisition, adverse events, hospital resources</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Creese (1984)</td>
<td>alternative strategy vs routine practice vs immunization campaign</td>
<td>number of avoided deaths</td>
<td>direct: input of staff time, transportation, vaccine, other materials, supervising, supplying, training, publicizing</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Creese (1987)</td>
<td>campaign vs routine immunization</td>
<td>fully immunized infant</td>
<td>direct: human and physical resources</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Dayan (2001)</td>
<td>vaccine vs no vaccine</td>
<td>reductions in illness rate</td>
<td>direct: vaccine; adverse effects; outpatient visits; X-rays; antipyretic, antiviral, and antibiotic treatment; hospitalization</td>
<td>societal</td>
<td>CEA</td>
</tr>
<tr>
<td>Doyle (2001)</td>
<td>venlafaxine vs SSRIs vs TCAs</td>
<td>treatment success and symptom-free days</td>
<td>direct: physician services, laboratory services, facility services, electroconvulsive therapy, pharmacotherapy</td>
<td>government</td>
<td>CEA</td>
</tr>
<tr>
<td>Farina (2002)</td>
<td>palivizumab vs no intervention</td>
<td>reductions in hospitalization rate</td>
<td>direct: drugs, drug prescriptions, hospitalization</td>
<td>societal</td>
<td>CEA</td>
</tr>
<tr>
<td>Ferraz (1995)</td>
<td>screening vs no screening hepatitis B vaccine program</td>
<td>equal benefits</td>
<td>direct: screening, vaccination</td>
<td>NR</td>
<td>CMA</td>
</tr>
<tr>
<td>Gehrke (2002)</td>
<td>antihypertensive drug class</td>
<td>% patients with controlled hypertension</td>
<td>direct: purchase of drugs or supplies, payment for visits to doctor, laboratory tests, health insurance, expenses with meals and transportation</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Morales (2004)</td>
<td>vaccine (IMOVAX) vs no vaccine</td>
<td>% influenza-like symptoms</td>
<td>direct: vaccines, materials, administration indirect: time to get the vaccine, ADE in terms of productivity, influenza-like symptoms, employees working sick</td>
<td>employer</td>
<td>CBA</td>
</tr>
<tr>
<td>Murad (1997)</td>
<td>UFT vs fluorouracil</td>
<td>equal benefits</td>
<td>direct: prechemotherapy, physician visits, premedication, chemotherapy, laboratoy procedures, ADE, hospitalization</td>
<td>NR</td>
<td>CMA</td>
</tr>
<tr>
<td>Murad (1997)</td>
<td>UFT vs fluorouracil</td>
<td>equal benefits</td>
<td>direct: prechemotherapy, physician visits, premedication, chemotherapy, laboratory procedures, ADE, hospitalization</td>
<td>government</td>
<td>CMA</td>
</tr>
<tr>
<td>Larrieu (2000)</td>
<td>conventional surgical and unconventional treatment (albendazole or puncture-aspiration–injection–respiration) vs hydatidosis control program</td>
<td>avoided costs</td>
<td>direct: consultation, derivation, hospitalization, diagnosis, treatment</td>
<td>NR</td>
<td>CBA</td>
</tr>
<tr>
<td>Levine (1993)</td>
<td>vaccine vs no vaccine</td>
<td>avoided costs</td>
<td>direct: treatment, exams, procedures, hospitalization, sequelae</td>
<td>government</td>
<td>CBA</td>
</tr>
<tr>
<td>Quintero (2001)</td>
<td>sevoflurane vs remifentanil + propofol</td>
<td>time between surgery end and hospital discharge, ADE</td>
<td>direct: drug acquisition</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Rodrigo (2003)</td>
<td>ipratropium bromide + salbutamol vs salbutamol only</td>
<td>reduction in hospitalization frequency</td>
<td>direct: treatment, exams, procedures, hospitalization, sequelae</td>
<td>NR</td>
<td>CEA</td>
</tr>
</tbody>
</table>

ADE = adverse drug event; CBA = cost-benefit analysis; CEA = cost-effectiveness analysis; CMA = cost-minimization analysis; NR = not reported; SSRIs = selective serotonin-reuptake inhibitors; TCAs = tricyclic antidepressants; UFT = tegafur–uracil.
agencies (OS 3.3, SD 0.6). These scores were not significantly different (Kruskal–Wallis test; $\chi^2 = 0.86; p = 0.65$).

There were 15 (62.5%) CEAs, 6 (25.0%) CBAs, and 3 (12.5%) CMAs; no CUAs were found. CBAs scored highest (OS 2.8, SD 0.8), CEAs next (OS 2.7, SD 0.7), and CMAs lowest (OS 2.0, SD 0.5). These scores were not significantly different (Kruskal–Wallis test; $\chi^2 = 2.24; p = 0.32$). The average OS was 2.6 (SD 1.0). By country, the OS score was 2.7 (SD 0.6, n = 5) for Argentina, 2.4 (SD 0.6, n = 9) for Brazil, 2.4 (SD 1.2, n = 3) for Colombia, 2.3 (SD 0.5, n = 2) for Chile, 3.0 (n = 1) for Ecuador, 3.8 (n = 1) for Peru, 2.2 (n = 1) for Uruguay, and 3.4 (SD 0.6, n = 2) for Venezuela. These scores were not statistically different (Kruskal–Wallis test; $\chi^2 = 4.56; p = 0.71$).

The mean OS for individual articles ranged from a minimum of 1.4 (35%) to a maximum of 3.8 (95%) out of a possible score of 4. The mean OS for all articles combined was 2.6 (SD 1.0), and the median was 2.4. Expressed as a percentage, the average was 65% and the median was 60%. These scores could be considered poor to acceptable since half scored above 60% and half scored below that value. There was no trend in scores over time ($R^2 = 0.12; p = 0.104$).

Table 2 shows that mean GI scores and OSs were almost identical (2.7 and 2.6, respectively) and highly correlated ($r = 0.91, df = 22, p = 0.005$). The item with the highest mean score was “definition of study aim,” and the question with the lowest mean OS was “ethical issues.” The correlation between GI score and OS was 0.91 (df = 22, $p < 0.001$). Variability between the raters was less than 0.5 point on OSs and less than 1 point on all items.

There was a statistically significant difference between mean OSs for South American publications (OS 2.6) compared with previous studies from other parts of the world (OS 2.9, $p < 0.001$).

Table 3 summarizes quality category differences between South American publications and those published previously.

### Discussion

The mean OS for South American articles related to health economic analysis was 2.6 out of 4, considered to be acceptable. Expressed as a percentage (65%), it would be on the lower end of that classification. The scores ranged from 1.4 (very poor) to 3.8 (good). Only 4 (17%) articles were categorized as good. The majority (54%; n = 13) would be described as poor, indicating that improvements are needed.

We accessed databases from their inception to obtain the maximum number of articles available in the literature. This strategy was considered efficient since the years of publication ranged from 1984 to 2004.

Our findings are comparable to those from 2 previous studies that analyzed economic evaluations worldwide from 1989 to 1993 and from 1992 to 1995. However, the mean OS score of 2.6 for our study in South American countries was somewhat lower than that shown in the previous quality assessments, which were 2.9 and 3.0, respectively. In all of these studies, the mean OS was very similar to the mean GI score (OS 2.6, GI 2.7). Differences between our findings and those from studies from other countries in total mean OSs and among scores for individual items could be related to a lack of health economics expertise in South America.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Drug/Alternatives Compared</th>
<th>Benefits/Outcomes</th>
<th>Costs</th>
<th>Perspective</th>
<th>Type of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rollán (2000)</td>
<td>FAM vs OAT vs LAC1 vs LAC2</td>
<td>% of disease eradication</td>
<td>direct: consults, endoscopy, diagnostic tests</td>
<td>NR</td>
<td>CBA</td>
</tr>
<tr>
<td>San Sebastian (2001)</td>
<td>hospital vs community health workers vaccination program</td>
<td>fully vaccinated children</td>
<td>direct: fuel, vaccine supplies, salaries, per diem allowances</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Shepard (1989)</td>
<td>campaign vs routine immunization</td>
<td>number of deaths averted</td>
<td>indirect: time spent at vaccination</td>
<td>NR</td>
<td>CEA</td>
</tr>
<tr>
<td>Suarez (2002)</td>
<td>second-line drug treatment for MDR tuberculosis vs 2 other second-line treatments for MDR tuberculosis</td>
<td>DALYs</td>
<td>direct: drug treatment, food parcels, physician consultations, DOT visits to MDR tuberculosis unit, committee evaluation, exams, adverse effects</td>
<td>government</td>
<td>CEA</td>
</tr>
<tr>
<td>Temporada Cookson (1997)</td>
<td>CVD 103-HgR (vaccine) vs cholera treatment</td>
<td>savings from vaccination program</td>
<td>direct: cholera treatment costs, CVD 103-HgR costs</td>
<td>NR</td>
<td>CBA</td>
</tr>
<tr>
<td>Ward (1986)</td>
<td>drugs vs radiotherapy vs surgery</td>
<td>% remission</td>
<td>direct: diagnostics, treatment, follow-up, complications</td>
<td>NR</td>
<td>CEA</td>
</tr>
</tbody>
</table>

CBA = cost-benefit analysis; CEA = cost-effectiveness analysis; DALYs = disability-adjusted life years; DOT = directly observed therapy; FAM = famotidine + amoxicillin + metronidazole; LAC1 = lansoprazole + amoxicillin + clarithromycin; LAC2 = lansoprazole + amoxicillin + clarithromycin; MDR = multidrug-resistant; NR = not reported; OAT = omeprazole + amoxicillin + tinidazole.
Other factors could include the inability to produce such analyses due to a lack of funding.

The types of questions with highest (definition of the study aim) and lowest (ethical issues) mean OS scores were the same as in previous quality assessments. The mode for this latter item was 1, which represents “not reported.” Others have suggested the appropriateness and importance of discussing ethical implications of pharmacoeconomic analysis in policy-making and medical decision-making.36,37

We found only one item (definition of study aim) from the 12 item checklist that had a mean OS higher than 3 (acceptable); the others ranged from 1.5 to 2.9 (not reported to poor to acceptable). We also noted that some studies did not correctly present or report their statistical, cost, or benefit analyses, which could directly compromise the overall quality and reliability of the results presented by the South American articles.

It is known that this distribution (CBA > CEA > CMA) also reflects the complexity of the analyses, meaning that CMAs are easier to perform because they allow authors to focus only on the costing aspects. There are 2 explanations that can account for these inverse score results: (1) the expertise in health economic analysis in South America is focused more on complex studies, thereby resulting in the highest scores, or (2) the scoring scale could be biased against noncomplex analyses. Noncomplex studies presented answers for individual scores that were categorized as not applicable (ie, CMA, issues such as “Measurement of outcomes/benefits”) and thus were not included in the scores.

We also found that several studies from South America did not report funding sources and other information, such as the perspective of the study and methods of statistical analysis. This information is considered very important. In the case of economic evaluations, it is crucial to identify whether the methodology was appropriately applied and the results can be extrapolated.

There are 2 major limitations to our study: (1) the definition of quality was applied more to the reporting than the actual execution of economic evaluations (which is impossible to observe) and (2) although the inter-rater reliability was found to be adequate, intra-rater reliability was not tested on the final analysis since raters analyzed studies independently. We assumed that the raters’ experiences and backgrounds were adequate for this quality assessment.

There is always the potential of publication bias as well (ie, studies with negative or null findings may not have been published). Therefore, our findings may underestimate the true number of South American health economic analyses in the literature. Unfortunately, unlike meta-analysis, which can use funnel plots, statistical estimates, or file drawer calculations, there is no technique that can be applied to estimate such bias. Therefore, the extent and impact of this limitation remain unknown. However, the potential for publication bias is also present in the studies we compared (ie, quality assessment from other parts of the world), and OS from South America remain significantly lower.

Conclusions

We conclude that quality scores for South American health economic analyses were, on average, acceptable but lower than those shown in research from other industrialized countries. We recommend that South American countries develop and implement guidelines for economic evaluations to assist in providing higher-quality reports. Future research should examine the level of expertise, educational opportunities, and availability of training programs in South America’s government, pharmaceutical industry,
and universities to assess the level of skills of those performing health economic analyses.

Márcio Machado PharmD, PhD Candidate, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile; Research Fellow, PharmIdeas Research and Consulting Inc., Oakville, ON, Canada

Michael Iskedjian MSc, President, PharmIdeas Research and Consulting Inc.

Thomas R Einarson PhD, Associate Professor, Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada; Vice-President, PharmIdeas Research and Consulting Inc.

Reprints: Dr. Einarson, Leslie Dan Faculty of Pharmacy, University of Toronto, 19 Russell St., Toronto, ON M5S 2S2, Canada, fax 416/978-8511, teinarson@utoronto.ca

We acknowledge the assistance of Monica L Zilberman MD PhD in obtaining articles from Brazil for this project.

References

28. Quintero JM. Costo efectividad en anestesia pediátrica: anestesia total intravenosa versus inhalada con sevofluorane en los tiempos de recupera-
MEDLINE, EMBASE, LILACS, y SIETES usando las siguientes palabras claves: costo-efectividad, costo-utilidad, reducción de costo, costo beneficio, las abreviaturas (por sus siglas in inglés) CEA, CUA, CMA, CBA, y todos los nombres de los países de América del Sur. Los artículos fueron clasificados por tipo y país por 2 evaluadores independientes. La calidad de los artículos fue determinada usando una hoja de 12 criterios con las siguientes puntuaciones: 4 (bueno), 3 (aceptable), 2 (pobre), 1 (no puedo juzgar), y 0 (inaceptable).

RESULTADOS: Se seleccionaron 25 artículos y se rechazó uno que estaba duplicado, para un total de 24 (CEA = 15, CBA = 6, CMA = 3; Brazil = 9, Argentina = 5, Colombia = 3, Chile = 2, Ecuador = 2, 1 para Perú, Uruguay, y Venezuela). La variabilidad entre evaluadores fue menor de 0.5 puntos de las puntuaciones totales (OS, por sus siglas en inglés) y menor de 1 punto en todos los criterios individuales. El promedio de las puntuaciones totales fue de 2.6 (desviación estándar = 1.0, rango 1.4–3.8). Los análisis de costo beneficio tuvieron las puntuaciones más altas (OS = 2.8, DE = 0.8), seguido por los de costo efectividad (OS = 2.7, DE = 0.7), siendo los de reducción de costos los de puntuación más baja (OS = 2.0, DE = 0.5). El criterio denominado “definición del objetivo del estudio” obtuvo la puntuación mayor (OS = 3.0, DE = 0.8) y el de “aspectos éticos” la menor (OS = 1.5, DE = 0.9). Por país, Perú obtuvo la puntuación más alta (puntuación total promedio = 3.8) y Uruguay la más baja (puntuación total promedio = 2.2). No se observó una tendencia significativa en las puntuaciones totales con respecto al tiempo de publicación (R² = 0.12; p = 0.104).

CONCLUSIONES: Las puntuaciones de calidad fueron de “pobres” a “aceptables” y más bajas que en investigaciones previas en otros países. Es necesario establecer estrategias para mejorar la calidad de los análisis económicos en salud en América del Sur. En el futuro, las investigaciones deben examinar el nivel de peritaje y las oportunidades educativas de los que tienen responsabilidad de realizar estudios económicos en salud en América del Sur.