

Economic analyses of human genetics services: A systematic review

Josh J. Carlson, MPH², Nora Beidler Henrikson, MPH², David L. Veenstra, PharmD, PhD³, and Scott D. Ramsey, MD, PhD¹

Purpose: The study's purpose was to conduct a structured review of economic analyses of genetic services. These will be increasingly valuable tools for assessing the clinical and economic outcomes of new medical technologies.

Methods: We searched for economic studies published between January 1990 and August 2004 from a variety of publicly available databases. Articles were first reviewed to determine whether they were original studies, and second to determine whether they were formal cost-effectiveness analyses by established criteria. Articles meeting these criteria were graded using a validated rating scale. **Results:** Of 149 articles, 63 met established criteria for cost-effectiveness analyses. The majority (87%) were published since 1996. The majority of studies considered adult (31) or prenatal (25) conditions with the remainder considering preconception or pediatric conditions. More than half used life years gained or an ad hoc measure of outcome (e.g., cases detected). Twenty-five percent measured outcome as quality-adjusted life years. The disease areas most considered were cancer (21%) and aneuploidies (18%). The average quality ranking was 87 of 100 possible (range 48–100). Common shortcomings included lack of statement of perspective, lack of discussion of potential bias, and lack of disclosure of funding sources. **Conclusions:** Relatively few economic evaluations are available for genetic services, and most are clustered in specific disease areas. Overall quality was high, but varied widely. Most shortcomings that would improve study quality are easy to address. To improve the relevance of these studies, researchers need to incorporate measures of outcome that are familiar to decision makers, including quality-adjusted life years. **Genet Med 2005;7(8):519–523.**

Key Words: genetic services, genetic test, genetic technologies, economic evaluation, cost-effectiveness

In an era of increasing health care expenditures and budget constraints, the rapid proliferation of genetic tests and services has been both a source of hope and concern. Genetic tests offer hope for early diagnosis and identification of persons at risk for serious diseases, with a goal of prevention or improved outcome through early treatment. However, interventions have both benefits and risks, and can be costly to both individuals and societies. In addition, under fixed budgets, spending more on genetic services will mean spending less on other health services. In this context, cost-effectiveness studies of genetic tests and services can be an important tool for decision makers seeking to maximize health benefits for available medical expenditures. A cursory review of the literature reveals that like genetic test technologies themselves, the number of economic evaluation studies

in this area has increased rapidly over time. Recently, a review of economic studies of pharmacogenomic technologies was published, although study quality was not evaluated.¹ To date no study has collected and evaluated all cost-effectiveness studies of human genetic tests and related services.

For cost-effectiveness studies to be useful, they must be timely, address technologies of interest to decision makers, and have high-quality methods and transparent reporting. Standard methods for conducting cost-effectiveness have been widely available for many years, as have efforts to increase the uniformity of reporting of these studies in the literature.^{2–4} To determine the extent to which these goals have been met, and to provide an overview of the current literature, we conducted a systematic search for and assessment of economic evaluations of genetic technologies.

METHODS

Literature search

During August 2004, we performed literature searches using PubMed, Proquest, LexisNexis, Expanded Academic Index, The Harvard Review of Economic Analyses (<http://www.hsph.harvard.edu/cearegistry/>), PsycINFO, National Institute for Clinical Excellence (<http://www.nice.org.uk>), and The Canadian Council on Technology Assessment in Health Care (http://www.ccohta.ca/entry_e.html). For databases that use

From the ¹Cancer Prevention Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; ²Institute for Public Health Genetics, University of Washington, Seattle, Washington; and ³Department of Pharmacy, University of Washington, Seattle, Washington.

Scott D. Ramsey, MD, PHD, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B230, PO Box 19024, Seattle, WA 98109-1024.

Received: July 12, 2005.

Accepted: August 4, 2005.

DOI: 10.1097/01.GIM.0000182467.79495.e2

MeSH search terms, we used the search terms “economic(s)” and/or “cost(s)” combined with “gene,” “genetic,” or “genotype.” The searches were limited to studies with abstracts, in English language, and with publication dates equal to or after 1990. For the remaining databases, we used manual searches to locate potential articles.

We reviewed the abstracts and culled the set to include only original publications of economic evaluations of genetic services. In situations in which we could not make a judgment solely from the abstract, we reviewed the full article before deciding.

Identifying economic analyses

In selecting potential articles for the study from the initial search results, we used the following definition of economic evaluation from Drummond et al.:³ “the comparative analysis of alternative courses of action in terms of both their costs and consequences.” To be considered a genetic service, the disease or condition had to be either primarily genetic or involve a genetic test. We broadly defined a genetic test as “the analysis of human DNA, RNA, chromosomes, proteins, and certain metabolites in order to detect heritable disease-related genotypes, mutations, phenotypes or karyotypes for clinical purposes.”⁵ We included articles based on consensus and discussion of whether each article met the criteria of being an economic evaluation and involving a genetic service.

Assessing the content and quality of each article

Two reviewers (N.B.H., J.J.C.) examined the articles meeting initial selection criteria and independently completed an abstract form for each article. The abstract form included the disease or condition of interest, the type of economic evaluation (i.e., cost-minimization, cost-effectiveness, cost utility, cost-benefit, or a combination thereof); the clinical emphasis (preconception, prenatal, pediatric/newborn, or adult); and the study population, intervention, comparator, results, and brief summary of sensitivity analysis.

Both reviewers also assigned a quality score to each article using a grading system for cost-effectiveness studies developed by Chiou et al.⁶ This grading system has been shown to be internally consistent and valid for assessing the quality of economic evaluation studies. According to the rating system, possible quality scores range from zero (worst quality) to 100 (best quality). Before grading began, we agreed on interpretations of the criteria that contained ambiguity for the purposes of our analysis. For example, we agreed that studies received credit for the quality of their data sources if the authors disclosed and discussed it as such, because we could not confidently judge the quality of all data sources in all the diseases represented in this analysis. A third reviewer (D.L.V.) resolved any disagreements between the raters’ scores through a tie-breaker process.

RESULTS

The initial database search using the MeSH terms and manual searches yielded 1252 articles. After an initial review of the titles and abstracts, 149 articles met the initial inclusion criteria

as original articles. Of these, 63 articles were determined to be comparative economic evaluations with costs and consequences identified.

Table 1 provides summary statistics for the articles included in this analysis. Articles are summarized by economic evaluation type, clinical category, and specific disease group. The most common type of economic evaluations were cost-effectiveness analyses (59%). The least common was cost-minimization (6%). The cost-utility subgroup received the highest mean quality score (mean 94.3).

Cost-effectiveness studies by definition relate costs to a single unit of effect that may differ in magnitude between alternative programs.³ In the genetics literature, the most common measure of effect was life years gained or life years saved. Other outcomes used were cases prevented/averted, cases detected, mutations detected, events prevented, births averted, fetuses detected, or carriers detected. Cost-utility studies by definition use quality-adjusted life years (QALYs) as the outcome, whereas cost-benefit studies report costs and outcomes in monetary terms. Table 2 shows cost-utility study results, with U.S. dollar values stated in the year of publication. In terms of economic value of the genetic services evaluated, the results varied widely from study to study.

The most common disease category was cancer (21%). The majority of studies in cancer focus on detecting susceptibility in at-risk relatives of probands or in high-risk populations. One study focused on tumor testing to plan treatment, and one study focused on genetic testing to diagnose leukemia. Studies on diseases diagnosed in infancy or in utero (aneuploidies, fetal anomalies, and inborn errors of metabolism) collectively made up approximately 21% of the studies in the sample. The

Table 1
Summary statistics and quality scores

	n	%	Mean quality score	Min-max
Cost-effectiveness	37	58.73%	87.51	59-100
Cost utility	16	25.40%	94.25	80-100
Cost-benefit	12	19.05%	86.00	60-100
Cost-minimization	4	6.35%	73.75	48-91
Adult	31	49.21%	87.52	59-100
Prenatal	25	39.68%	84.04	48-100
Pediatric/newborn	8	12.70%	94.80	85-100
Preconception	5	7.94%	77.40	59-93
Cancer:	13	20.63%	88.77	70-100
Breast cancer	5	7.94%	87.00	70-100
Colorectal cancer	5	7.94%	92.20	87-100
Other cancers	3	4.76%	86.67	79-94
Aneuploidies (eg. trisomy 21, 18)	11	17.46%	81.55	64-100
Cystic fibrosis	8	12.70%	89.38	77-100
Thrombophilia	6	9.52%	81.00	59-100
Fetal anomalies (eg. spina bifida)	6	9.52%	82.83	60-96
Hemochromatosis	5	7.94%	89.00	73-100
Hypercholesterolemia (familial)	4	6.35%	90.75	81-100
Other	3	4.76%	96.33	89-100
Hemoglobinopathy	3	4.76%	83.33	78-97
Inborn errors of metabolism	2	3.17%	92.50	85-100
Rheumatoid arthritis, lupus	2	3.17%	94.00	94-94
Total	63		87.05	48-100

Table 2.
Results of 14 cost utility (CU) studies^a

CU study	Intervention	Result
Elkin, E. B., M. C. Weinstein, et al. (2004). ¹⁰	6 HER-2 testing and trastuzumab treatment strategies for metastatic breast cancer	HercepTest with FISH confirmation of positive results: \$125,000/QALY
Higashi, M. K., D. L. Veenstra, et al. (2002). ¹¹	Genetic testing for susceptibility to periodontitis	\$32,633/QALY
Eckman MH, Singh SK, Erban JK, Kao G: (2002). ¹²	Factor V Leiden testing followed by prolonged anticoagulation therapy (3 years or lifelong) for positives	\$16,823/QALY
Marchetti, M., S. Quaglini, et al. (2001). ¹³	Screening for Factor V Leiden and prothrombin G20210A, anticoagulation only for double heterozygotes	\$13,624/QALY
Marchetti M, Pistorio A, Barosi G: (2000). ¹⁴	Screen for Factor V Leiden plus 2 yrs warfarin for mutation positives	\$12,833/QALY
Calvert, N. W., A. B. Morgan, et al. (2003). ¹⁵	Radical prostatectomy and experimental selection-based management using DNA-ploidy	DNA alternative: £12,068/QALY. RP is dominated.
Harris, R. A., A. E. Washington, et al. (2004). ¹⁶	Amniocentesis and chorionic villus sampling	CVS dominated at all ages by amniocentesis; amniocentesis \$11,300/QALY-\$14,200/QALY
Auerbach AD, Sanders GD, Hambleton J: (2004). ¹⁷	Multiple testing/treatment strategies for DVT	Test, treat positives for 24 months: \$11,100/QALY
Rowley, P. T., S. Loader, et al. (1998). ¹⁸	Carrier screening for CF	\$8,290/QALY
King, C. H., D. F. Fischler, et al. (2002). ¹⁹	Newborn screen for genetic susceptibility to rheumatic fever	\$7,900/QALY
Schoen, E. J., J. C. Baker, et al. (2002). ²⁰	Routine MS/MS testing for inborn errors of metabolism	\$5827/QALY
Venditti, L. N., C. P. Venditti, et al. (2003). ²¹	Newborn screening by MS/MS for MCAD	\$100/QALY-\$5,600/QALY
Asberg, A., S. Tretli, et al. (2002). ²²	Hemochromatosis screening (phenotypic)	\$250/QALY
Adams, P. C. and L. S. Valberg (1999). ²³	Hemochromatosis screening (genotypic)	Intervention is dominated

^aQuality score range 80-100

“other” disease category includes infectious diseases and periodontal disease. For this category, most studies involved genotyping to gauge genetic susceptibility.

Article quality

Overall, the mean quality score for the sample was 87.1; quality scores ranged from 48 to 100. Table 3 shows the percentage of articles missing each quality criteria, as agreed on by all raters. The most commonly missed criteria were failing to state the perspective of the analysis (36%); not explicitly discussing direction of bias (35%); and not disclosing the funding source for the study (35%).

The overall intraclass correlation between the raters was 0.816 (95% confidence interval: 0.696–0.889). In general, the criteria that required the most objective response, such as disclosing the funding source and conduct of sensitivity analyses, received the highest proportion agreeing. The criteria with the lowest percent agreement between raters were discussion of magnitude and direction of potential bias and inclusion of short- and long-term and negative outcomes.

Year of publication

We hypothesized that there may be a trend toward higher quality ratings as the year of publication increased, so we conducted a linear regression analysis and characterized trends by

groups of years. The regression showed a modest trend ($P = .19$, not shown) toward improving quality over time. When we analyzed the trend by 5-year categories, studies published after 2000 received consistently higher quality scores than those published between 1996 and 2000, as seen in Table 4. Because there were relatively few studies (eight) published between 1990 and 1995, the earlier trend is more difficult to characterize.

DISCUSSION

In a structured review of economic evaluations of genetic services published between 1990 and 2004, we found a modest but rapidly increasing number of studies. The review identified strengths and weakness in the current literature. The most common single type of study was cost-effectiveness analyses; cost-utility studies (the second most common study type) are recommended by the U.S. Panel on Cost-Effectiveness in Medicine. The proportion of high-quality studies in genetics was much higher than a recent review of another clinical area (digestive diseases) using the same grading system (64% vs. 29%),⁷ and there was a modest (but nonsignificant) trend toward increasing quality over time. Nevertheless, there were several areas of concern regarding this literature. Quality scores varied widely among studies, with many falling below 75 points, a threshold others have suggested may indicate modest

Table 3
Missing quality criteria for economic evaluations in genetic services^a

Criteria (weight)	Number of studies missing criteria	Percent (%)	Percent of agreement between raters ^b
Were the perspective of the analysis (societal, third-party payer, etc.) and reasons for its selection stated? (4)	23	36.5	84.1
Did the author(s) explicitly discuss direction and magnitude of potential biases? (6)	22	34.9	65.1
Was there a statement disclosing the source of funding for the study? (3)	22	34.9	98.4
Did the analytic horizon allow time for all relevant and important outcomes? Were benefits and costs that went beyond 1 year discounted (3–5%) and justification given for the discount rate? (7)	16	25.4	85.7
Were the choice of economic model, main assumptions and limitations of the study stated and justified? (7)	12	19.1	77.8
Was uncertainty handled by: 1) statistical analysis to address random events; 2) sensitivity analysis to cover a range of assumptions? (9)	9	14.3	95.2
Were the primary outcome measure(s) for the economic evaluation clearly stated and were the major short term, long term and negative outcomes included? (6)	8	12.7	60.3
Were the economic model (including structure), study methods and analysis, and the components of the numerator and denominator displayed in a clear transparent manner? (8)	8	12.7	84.1
Was the study objective presented in a clear, specific, and measurable manner? (7)	5	7.9	71.4
Was the measurement of costs appropriate and the methodology for the estimation of quantities and unit costs clearly described? (8)	5	7.9	87.3
Was incremental analysis performed between alternatives for resources and costs? (6)	3	4.8	76.2
Was the methodology for data abstraction (including value health states and other benefits) stated? (5)	3	4.8	82.5
Were variable estimates used in the analysis from the best available source (i.e. Randomized Control Trial—Best, Expert Opinion—Worst)? (8)	2	3.2	77.8
Were the health outcomes measures/scales valid and reliable? If previously tested valid and reliable measures were not available, was justification given for the measures/scales used? (7)	1	1.6	88.9
Were the conclusions/recommendations of the study justified and based on the study results? (8)	1	1.6	90.5
If estimates came from a subgroup analysis, were the groups prespecified at the beginning of the study? (1)	0	0.00	98.4

^a Rating system developed by Chiou et al.⁶

^b Overall intraclass correlation (alpha): 0.816 (95% confidence interval: 0.696–0.889).

Table 4
Year of publication

Year of publication ^a	N	Mean score (SD)
1990-1995	8	87.1
1996-2000	27	82.1
2001-2004	28	91.5

^a Analysis of variance $P < .05$. SD, standard deviation.

or poor quality.⁸ The majority of studies used measures of outcome other than life years or QALYs gained. Ad hoc outcome measures limit comparison with other evaluations in health care, and thus are less useful to decision makers. That said, we acknowledge the difficulties of evaluating outcomes for prenatal genetic services in terms of years or life lost or gained. It may be desirable to identify a common outcomes measure that is familiar to the genetics community for inclusion in future cost-effectiveness studies. For example, symptom-free days was determined by asthma researchers to be a valuable and desirable metric when new outcomes studies were planned for this disease.⁹ Finally, the scope of studies is relatively small in rela-

tion to the number of available genetic services. Cancer was most common (13 studies), followed by aneuploidies (11 studies). Very few studies focused on pediatric populations and preconception genetic services.

We note several important limitations in this analysis. First, in assessing quality of studies, we tended to focus on identifying criteria rather than quality *within* criteria. Thus, our bias may be toward higher scores. Bias introduced here would be internally consistent and not affect the hierarchy of scores. Second, the quality assessment system we applied does involve some subjectivity. We mitigated this by arbitrating discrepancies with a senior reviewer and having two independent reviewers; the ultimate agreement between reviewers was relatively high. The grading system we used is somewhat new, and thus few comparisons are available with other interventions or health conditions. Furthermore, the quality assessment tool is limited in its ability to capture articles that lacked transparency or were poorly presented. Finally, the methods and reporting of results varied widely, making it difficult to construct league tables that would facilitate comparison of studies.

On the basis of our review of the literature, we have several suggestions for future economic analyses of genetic services. First, authors should attend to simple disclosure issues. These issues reduced the quality ratings of many studies and are easily addressed. The most commonly missed criteria were statement of a funding source and statement of the study perspective (e.g., societal, third-party payer). A second common issue was lack of discounting of costs or effects. Studies can show discounted and undiscounted results without changing the study design. Third, whenever possible, future analyses should include uniform measures of outcome that are familiar to decision makers who use this literature. Specifically, using QALYs or life years gained will facilitate comparison of these interventions with others in medicine. In cases in which such outcomes are problematic (e.g., preconception or prenatal testing), international guidelines developed from within the genetics community for common measures would be helpful, as noted above. In the interim, researchers should justify their choices as carefully as possible. Fourth, many areas of genetic service are not addressed in the economics literature, despite their potential impact on population health and the costs of medical care. We particularly encourage studies of well-known genetic variants in newborn and pediatric populations. On the other hand, certain topics have been well studied (e.g., cystic fibrosis, hemochromatosis), and future efforts that would only offer further replication may not be the best use of resources unless new genetic variants, testing, or treatment strategies emerge for these conditions.

Funding

Projects No. U35MC02601 and No. U35MC02602 from the Maternal and Child Health Bureau (Title V, Social Security Act), No. 11223, Health Resources and Services Administration, Department of Health and Human Services.

References

- Phillips KA, Van Bebber SL. A systematic review of cost-effectiveness analyses of pharmacogenomic interventions. *Pharmacogenomics* 2004;5(8):1139–1149.
- Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;313(7052):275–283.
- Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 2nd ed. Oxford: Oxford University Press; 1997.
- Gold MR, Siegal JE, Russel LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
- Holtzman NA, Watson MS. Promoting safe and effective genetic testing in the United States. Final report of the Task Force on Genetic Testing. *J Child Fam Nurs* 1999;2(5):388–390.
- Chiou CF, Hay JW, Wallace JF, Bloom BS, et al Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care* 2003;41(1):32–44.
- Spiegel BM, Targownik LE, Kanwal F, Derosa V, et al The quality of published health economic analyses in digestive diseases: a systematic review and quantitative appraisal. *Gastroenterology* 2004;127(2):403–411.
- Ofman JJ, Sullivan SD, Neumann PJ, Chiou CF, et al Examining the value and quality of health economic analyses: implications of utilizing the QHES. *J Manag Care Pharm* 2003;9(1):53–61.
- The Global Initiative For Asthma Dissemination and Implementation of Asthma Guidelines. Available at: <http://www.ginasthma.com/dis.pdf>. Accessed January 10, 2005.
- Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC. HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. *J Clin Oncol* 2004;22(5):854–863.
- Higashi MK, Veenstra DL, del Aguila M, Hujuel P. The cost-effectiveness of interleukin-1 genetic testing for periodontal disease. *J Periodontol* 2002;73(12):1474–1484.
- Eckman MH, Singh SK, Erban JK, Kao G. Testing for factor V Leiden in patients with pulmonary or venous thromboembolism: a cost-effectiveness analysis. *Med Decis Making* 2002;22(2):108–124.
- Marchetti M, Quaglini S, Barosi G. Cost-effectiveness of screening and extended anticoagulation for carriers of both factor V Leiden and prothrombin G20210A. *QJM* 2001;94(7):365–372.
- Marchetti M, Pistorio A, Barosi G. Extended anticoagulation for prevention of recurrent venous thromboembolism in carriers of factor V Leiden—cost-effectiveness analysis. *Thromb Haemost* 2000;84(5):752–757.
- Calvert NW, Morgan AB, Catto JW, Hamdy FC, et al Effectiveness and cost-effectiveness of prognostic markers in prostate cancer. *Br J Cancer* 2003;88(1):31–35.
- Harris RA, Washington AE, Nease RF, Jr., Kuppermann M. Cost utility of prenatal diagnosis and the risk-based threshold. *Lancet* 2004;363(9405):276–282.
- Auerbach AD, Sanders GD, Hambleton J. Cost-effectiveness of testing for hypercoagulability and effects on treatment strategies in patients with deep vein thrombosis. *Am J Med* 2004;116(12):816–828.
- Rowley PT, Loader S, Kaplan RM. Prenatal screening for cystic fibrosis carriers: an economic evaluation. *Am J Hum Genet* 1998;63(4):1160–1174.
- King CH, Fischler DF, Gerkin RD. Will genetic testing alter the management of disease caused by infectious agents? A cost-effectiveness analysis of gene-testing strategies for prevention of rheumatic fever. *Clin Infect Dis* 2002;34(11):1491–1499.
- Schoen EJ, Baker JC, Colby CJ, To TT. Cost-benefit analysis of universal tandem mass spectrometry for newborn screening. *Pediatrics* 2002;110(4):781–786.
- Venditti LN, Venditti CP, Berry GT, Kaplan PB, et al Newborn screening by tandem mass spectrometry for medium-chain Acyl-CoA dehydrogenase deficiency: a cost-effectiveness analysis. *Pediatrics* 2003;112(5):1005–1015.
- Asberg A, Tretli S, Hveem K, Bjerve KS. Benefit of population-based screening for phenotypic hemochromatosis in young men. *Scand J Gastroenterol* 2002;37(10):1212–1219.
- Adams PC, Valberg LS. Screening blood donors for hereditary hemochromatosis: decision analysis model comparing genotyping to phenotyping. *Am J Gastroenterol* 1999;94(6):1593–1600.