

# UCSF

## UC San Francisco Previously Published Works

### Title

Understanding Cost-Effectiveness Analyses: An Explanation Using Three Different Analyses of Lung Cancer Screening.

### Permalink

<https://escholarship.org/uc/item/2d94k9bf>

### Journal

AJR. American journal of roentgenology, 205(2)

### ISSN

0361-803X

### Authors

Curl, Patti K  
Kahn, James G  
Ordovas, Karen G  
et al.

### Publication Date

2015-08-01

### DOI

10.2214/ajr.14.14038

Peer reviewed

# Understanding Cost-Effectiveness Analyses: An Explanation Using Three Different Analyses of Lung Cancer Screening

Patti K. Curl<sup>1</sup>  
 James G. Kahn<sup>2</sup>  
 Karen G. Ordovas<sup>1</sup>  
 Brett M. Elicker<sup>1</sup>  
 David M. Naeger<sup>1</sup>

**OBJECTIVE.** Cost-effectiveness analyses (CEAs) contribute to informed decision making, at both the practitioner and societal levels; therefore, understanding CEAs is valuable for radiologists. In light of the recently published National Lung Cancer Screening Trial (NLST) CEA, we aim to explain the terminology, methods, and heterogeneity of CEAs.

**CONCLUSION.** We compared the NLST results to two example lung cancer screening CEAs (which do not rely on NLST data). Both examples assessed screening but reached substantially different conclusions.

**C**ost effectiveness is playing an increasingly important role in clinical decision making and the development of standards of care. The cost-effectiveness of screening and preventive care interventions in particular is scrutinized because the benefit is limited to a small proportion of people receiving the intervention whereas the costs can be substantial [1]. Evaluation of the cost-effectiveness of CT screening for lung cancer in high-risk patients appears particularly relevant as policy makers and practitioners grapple with the appropriate utilization of this technology. As radiologists, we must understand cost effectiveness, including how to interpret published CEAs and the methodologic decisions made by the CEA researchers. In light of the recent publication of the National Lung Cancer Screening Trial (NLST) [2, 3] group's official CEA, we aim to explain the terminology, methods, and heterogeneity of CEAs. We will discuss these issues via an analysis of the NLST CEA and two additional analyses of CT screening that have been published by groups other than the NLST; they report substantially different results, allowing a comparison and discussion of the assumptions and methods used in performing a CEA.

## Terminology

### Cost

The definition and assessment of intervention costs are generally straightforward, usually restricted to measurements of resources and their monetary cost in dollars or other

currency. Measurements of net cost can become complicated because all costs influenced by the intervention should also be included, including costs of follow-up. If an intervention averts the need for future expensive clinical interventions, this will reduce the net cost estimate for the intervention. Conversely, if an intervention results in more clinical interventions, the net cost estimate will be higher. Nonmedical costs (e.g., time costs) are often included when they are likely to be significant and are measurable. Potential harm to the patient is generally not factored into the cost part of the equation, except where it affects medical care utilization; instead, it is accounted for by adjusting the estimated effectiveness [4].

### Effectiveness

The effectiveness of a CEA can be any measure of clinical impact (e.g., number of cancers detected or deaths averted); however, the most commonly reported effectiveness outcomes in the United States are life years and quality-adjusted life years (QALYs) [4]. Life years gained represent an estimate of the expected time gained by averting a death—that is, the difference between how long people lived on average after receiving the intervention versus without it. This is valuable information, but it estimates only the duration of life, without accounting for the quality of life. QALYs include effects on morbidity as well and thus incorporate all health consequences of disease and interventions. In QALYs, the duration of life is mul-

**Keywords:** cost-effectiveness analyses (CEAs), low-dose chest CT, lung cancer screening

DOI:10.2214/AJR.14.14038

Received October 31, 2014; accepted after revision January 20, 2015.

<sup>1</sup>Department of Radiology and Biomedical Imaging, University of California, San Francisco, 505 Parnassus Ave, M-391, San Francisco, CA 94143-0628. Address correspondence to D. M. Naeger (david.naeger@ucsf.edu).

<sup>2</sup>Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco, San Francisco, CA.

AJR 2015; 205:344–347

0361–803X/15/2052–344

© American Roentgen Ray Society

multiplied by a coefficient estimating the average health state utility, which varies from 0 (death) to 1 (perfect health) [4]. A CEA that uses QALYs as the effectiveness unit is often referred to as a “cost-utility analysis.”

**Comparing Cost and Effectiveness**

The outcome of a CEA for an intervention is the ratio of added cost to added effectiveness (e.g., net cost per cancer identified per QALY gained). A CEA must compare two scenarios: costs and total QALYs are estimated with and without the intervention or test, and the differences are then divided. The result is the incremental cost-effectiveness ratio (ICER) [4]. For example, if the net cost of screening was \$10,000 per person higher than no screening and the health benefit was 0.1 QALYs, then the ICER for screening versus no screening would be \$10,000 / 0.1, or \$100,000 per QALY.

**Lung Cancer Screening Analyses**

The NLST was a randomized controlled trial comparing low-dose CT screening to chest radiography in patients at high risk for lung cancer. The study followed 53,000 people for 5–7 years and showed a 20% reduction in lung cancer mortality in the low-dose CT group. The recently published NLST CEA was performed using the NLST trial data for inputs including screening efficacy and resource utilization [2, 3].

In addition to the NLST analysis, two other recent publications have reported CEAs of lung cancer screening. Although both were published after the NLST results were available (but not the NLST CEA), neither explicitly used the results as inputs. One analysis, by Villanti et al. [5], is based on a model created by the private consulting group Milliman [6]. The other analysis, by McMahon et al. [7, 8], is based on the National Cancer Institute’s lung cancer policy model. All three analyses are cost-utility analyses because they use QALYs to calculate ICERs for screening versus usual care as their primary outcome. Whereas the NLST reported an ICER of \$81,000 per QALY [3], Villanti et al. [5] reported an ICER of \$28,000 per QALY gained, and McMahon et al. [7] reported ICERs between \$110,000 and \$169,000 per QALY gained; this four- to sixfold difference is potentially large enough that it would affect policy recommendations.

**Cost-Effectiveness Analysis Inputs**

Intervention cost inputs are simple at face value, but there are several decisions in-

volved. In the United States, the billed charge by the provider to the payer (e.g., Medicare or private insurance), the amount actually reimbursed (“allowed charges”), and the out-of-pocket cost to the patient are all different values. The most common practice is to use the average amount reimbursed by payers as a reasonable approximation of true cost. If an analysis were to use billed charges rather than the paid amount, the estimate of the ICER could be more than twice as large. If an analysis used actual costs to clinical providers, excluding profits, then the ICER estimate could be lower. The accuracy and relevance of cost inputs are highly specific to the context of a CEA; consequently, understanding the population and economic setting of the analysis is important to appropriately interpret the results [9].

Estimating effectiveness can be complicated. Inputs for models are typically numerous and diverse. They include probabilities (e.g., the chance that a new cancer will be detected in a screening interval), duration of life, and health state utilities [4]. For a screening test, estimates of effectiveness should include the consequences of false-positive and -negative tests, such as the morbidity of further diagnostic steps and treatment options [1]. Anxiety resulting from a false-positive can be accounted for as a utility decrease. (In the case of lung cancer screening, a recent study found no significant utility difference in patients with initially false-positive results versus those with true-negatives results [10]). For a population with a high overall mortality risk, such as older individuals screened for lung cancer, the analysis must accurately account for the effect of comorbidities on life expectancy without cancer when estimating the potential benefit of curative and life-extending treatments.

Often inputs that are most directly useful for a particular CEA are unavailable in the published literature. For example, for cancer interventions, the duration input most useful for calculating an ICER is mean overall survival in years, but the value reported in clinical trials is often the percent surviving after a specific time interval. Having access to the actual trial data to derive desired inputs can reduce uncertainty in a CEA and is a significant advantage of the analysis performed with the NLST data [3].

**Inputs in Lung Cancer Screening Analyses**

All three CEAs used a payer reimbursement estimate to quantify the cost of the in-

tervention, but the estimates differ considerably. The average annual cost of CT screening (including follow-ups) estimated in the study by Villanti et al. [5] was \$210. The value estimated for the cost of a single CT screening in the study by McMahon et al. [7] was \$283, and the value used in the NLST analysis was \$285 [3]. The lower estimate in Villanti et al.’s study [5] is interesting: they used an estimated cost of CT that is lower than the current actual cost on the basis of the assumption that large-scale screening would result in lower fees for screening; they estimated the reduced cost in their model by applying the ratio of fees charged for screening versus diagnostic mammography [5, 6].

Treatment costs are not reported in the same manner between studies but also appear to differ considerably: both McMahon et al. [7] and Villanti et al. [5] used sources for treatment costs that exceed \$100,000 per lung cancer treatment, whereas the NLST analysis reported an average cost of treatment that equates to \$27,000 per diagnosed cancer [3]. Finally, the number of follow-ups for an initially positive test could have a significant effect on cost: McMahon et al.’s model [7] averaged approximately four follow-ups whereas the NLST averaged closer to one [3]. The difference in costs between the CEAs—most significantly, the difference in screening CT cost estimates—contributes to the overall difference in ICER estimates, with a lower cost of screening in the study by Villanti et al. [5] contributing to a lower ICER.

The analyses also used different sources to estimate the health state utility (which largely determines the effectiveness). For example, a patient with metastatic non-small cell lung cancer after treatment and after recurrence was assigned a utility modifier of 0.57 in the Villanti et al. study [5] and 0.62 in the McMahon et al. study [7, 11, 12]. This difference, though relatively small, leads to an increase in the estimated effectiveness of preventing cancer in the analysis of Villanti et al. [5], which leads to an increase in the estimated QALYs gained and consequently a lower ICER.

**Cost-Effectiveness Analysis Methods**

There are many models that can be used for CEA, but they all involve some method of simulating what will happen with or without an intervention, often relying on uncertain data about disease acquisition and progression [13]. Modeling intervention effects adds further uncertainty to intervention evaluations. As has been shown thoroughly

in clinical research, observational data may bias effectiveness estimates because of important differences between groups that are not randomized [1]. The analysis done by the NLST group included precise estimates for the period of close follow-up; however, for portrayal of the continued costs and benefits after the duration studied in the trial, simulation was required, and their analysis modeled costs based on Medicare reimbursements rather than the trials' actual costs [3]. This decision was likely appropriate given that research costs tend to differ from clinical reimbursements. The length of time that should be simulated—the time horizon—is one of the most important factors in evaluating a CEA [4]. In cases where an intervention is expected to provide only a temporary benefit, a short time horizon may be appropriate; by contrast, in the evaluation of interventions that have the potential to extend life or avert long-term sequelae, a lifetime horizon is the most appropriate. The McMahon et al. [7] and NLST [3] analyses used a full lifetime horizon, and the Villanti et al. analysis [5] used a 15-year horizon. In the population being studied, using a 15-year versus lifetime time horizon does not dramatically alter results and therefore does not substantially contribute to the difference in ICER estimates between the two studies (though the difference slightly favors a higher ICER estimate in Villanti et al.'s study).

When models include costs and clinical outcomes that occur years after the initial intervention, experts recommend discounting values that are remote to that intervention at a rate of 3% annually [4]. This practice—based on the economic principle that costs or benefits in the immediate present have more value than costs or benefits that will be experienced in the distant future—decreases the impact of late consequences of an intervention. The NLST [3] and McMahon et al. [7] analyses discounted both costs and QALYs at the customary rate of 3%, whereas the Villanti et al. analysis [5] did not discount either costs or QALYs. Not discounting costs increases the total cost estimate (and thus increases the ICER), whereas not discounting QALYs increases the total benefit estimate (and thus decreases the ICER). For a cancer screening CEA with substantial up-front costs and delayed benefit, the resulting increase in benefit will likely outweigh the increase in cost and consequently have a net effect of decreasing the ICER. Thus, the difference in discounting further contributes to the lower ICER reported

**TABLE 1: Summary of Differences Between Three Recent Lung Cancer Screening Cost-Effectiveness Analyses**

Component of Analysis	Black et al. (NLST) [3]	McMahon et al. [7]	Villanti et al. [5]	Likely Effect on ICER
Demographics				Unclear
Age (y)	55–74	50–74	50–64	
Pack-y	≥ 30	≥ 20	≥ 30	
Cost of screening CT (\$)	285 <sup>a</sup>	283 <sup>a</sup>	210 <sup>b</sup>	↓ Villanti
6-Year mortality reduction (%)	20	28	> 20	↑ Black
Discount rate (%)	3	3	0	↓ Villanti
Time horizon	Lifetime	Lifetime	15 y	Small
ICER (\$)	81,000	110,000–169,000	28,240	

Note—NLST = National Lung Cancer Screening Trial, ICER = incremental cost-effectiveness ratio.

<sup>a</sup>Per CT examination.

<sup>b</sup>Per year.

ed by Villanti et al. Table 1 summarizes many of the aforementioned differences between the two CEAs, as well as the effects that such differences have on the ICERs.

For complicated models such as those for simulating lung cancer screening, a valuable tool for improving precision is calibration [14]. Calibration (also known as “benchmarking”) is the practice of taking an established clinical outcome (e.g., the mortality difference at 6 years follow-up seen in a well-regarded randomized trial like the NLST) and ensuring that the model predicts a similar outcome [14]. Calibration is used in the process of building a model, and knowing what data the authors used for calibration is important for validation: checking how closely the model's results correspond to outcomes from well-regarded clinical trials is a quick and effective way to assess the accuracy of a model. For lung-cancer screening, the NLST reported a lung cancer-specific mortality reduction at 6 years of 20%, the McMahon et al. model reported a 28% reduction in a similar scenario, and the Villanti et al. model did not report an exact number although the value was described as “more optimistic” than the NLST results [2, 3, 5–7]. Thus, the mortality benefit in the NLST model was lower than in the other two models.

### Cost-Effectiveness Analysis Reporting

Most CEAs report the ICER produced by a base-case analysis. A base-case analysis uses the best estimated value for each input [4]. This single ICER can give a false impression of precision. The NLST [3] and Villanti et al. [5] reported base-case results with ICERs of \$81,000 and \$28,000, respectively, per QALY gained. Alternatively, an analysis may report a range of ICERs; McMahon et al.

[7] reported ICERs ranging from \$110,000 to \$169,000 per QALY gained; they reported a range of ICERs because they modeled multiple scenarios with different population characteristics (age, pack-years, gender) without highlighting one as a base case.

Whether the primary outcome is reported as a single base-case result or a range, reporting sensitivity analyses is necessary in any CEA given the large number of assumptions required. Sensitivity analyses vary inputs from the lowest to the highest reasonable values and report the effect of this variation on the ICER and its components (costs, effects). Sensitivity analyses allow the reader to assess the degree of uncertainty around the ICER estimate and help to identify the input values most influential in determining the cost effectiveness [4]. All three CEAs discussed performed sensitivity analyses [3, 5, 7].

### Conclusion

CEA can seem mysterious, and there is inherent uncertainty in the inputs and the precision of the results. However, understanding the methods used, as well as the nature of the decisions made, can significantly reduce this mystery. Although the analyses discussed here reported vastly different results, the two estimates that preceded the NLST did surround the NLST's more robust estimate and did so without the benefit of randomized trial data in their inputs. Further, looking at the inputs and methods identifies important reasons for the differences and allows an assessment of which estimate is most appropriate. In a setting with screening efficacy equal to that observed in the NLST, current CT screening costs, and relatively low average treatment costs, the ICER reported by

## Understanding Cost-Effectiveness Analyses

the NLST group (Black et al. [3]) is likely the most accurate; in a setting with higher efficacy but increased number of follow-up screens and higher average treatment costs, the ICER reported by McMahon et al. [7] is likely to be accurate; and in a setting with increased efficacy and considerably decreased screening costs, the ICER reported by Villanti et al. [5] will be more accurate.

Awareness of the cost and the value of imaging is crucial to the field of radiology. CEAs may guide decisions at the individual radiologist level and often substantially influence decisions made at the society and payer level. Medical imaging continues to be a costly, yet valuable, component of modern medical care, and imaging-based screening tests contribute to this cost and value [15]. Given the tremendous sway that CEAs hold over the field, it is of benefit to radiologists to understand the inputs, methods, and reporting of CEAs. This article has reviewed the basics of this type of analysis by using two published CEAs as examples; for interested readers, more detailed reading is available in the literature [1, 13, 14, 16–20].

### References

1. Newman TB, Kohn MA. Screening tests. In: Newman TB, Kohn MA. *Evidence-based diagnosis*. New York, NY: Cambridge University Press, 2009:116–138
2. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365:395–409
3. Black WC, Gareen IF, Soneji SS, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. *N Engl J Med* 2014; 371:1793–1802
4. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996; 276:1339–1341
5. Villanti AC, Jiang Y, Abrams DB, Pyenson BS. A cost-utility analysis of lung cancer screening and the additional benefits of incorporating smoking cessation interventions. *PLoS ONE* 2013; 8:e71379
6. Pyenson BS, Sander MS, Jiang Y, Kahn H, Mulshine JL. An actuarial analysis shows that offering lung cancer screening as an insurance benefit would save lives at relatively low cost. *Health Aff (Millwood)* 2012; 31:770–779
7. McMahon PM, Kong CY, Bouzan C, et al. Cost-effectiveness of computed tomography screening for lung cancer in the United States. *J Thorac Oncol* 2011; 6:1841–1848
8. McMahon PM, Kong CY, Johnson BE, et al. Estimating long-term effectiveness of lung cancer screening in the Mayo CT screening study. *Radiology* 2008; 248:278–287
9. Guyatt GH, Oxman AD, Kunz R, et al. Incorporating considerations of resources use into grading recommendations. *BMJ* 2008; 336:1170–1173
10. Gareen IF, Duan F, Greco EM, et al. Impact of lung cancer screening results on participant health-related quality of life and state anxiety in the National Lung Screening Trial. *Cancer* 2014; 120:3401–3409
11. Sturza J. A review and meta-analysis of utility values for lung cancer. *Med Decis Making* 2010; 30:685–693
12. Earle CC, Chapman RH, Baker CS, et al. Systematic overview of cost-utility assessments in oncology. *J Clin Oncol* 2000; 18:3302–3317
13. Krahn MD, Naglie G, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis. Part 4. Analyzing the model and interpreting the results. *Med Decis Making* 1997; 17:142–151
14. Naimark D, Krahn MD, Naglie G, Redelmeier DA, Detsky AS. Primer on medical decision analysis. Part 5. Working with Markov processes. *Med Decis Making* 1997; 17:152–159
15. Health Forum. *AHA hospital statistics: 2012 edition*. Chicago, IL: American Hospital Association, 2012
16. Detsky AS, Naglie G, Krahn MD, Naimark D, Redelmeier DA. Primer on medical decision analysis. Part 1. Getting started. *Med Dec Making* 1997; 17:123–125
17. Detsky AS, Naglie G, Krahn MD, Redelmeier DA, Naimark D. Primer on medical decision analysis. Part 2. Building a tree. *Med Dec Making* 1997; 17:126–135
18. Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis. Part 3. Estimating probabilities and utilities. *Med Dec Making* 1997; 17:136–141
19. Sox HC Jr, Blatt MA, Higgins MC, Marton KI. *Medical decision making*. Boston, MA: Butterworth-Heinemann, 1988
20. Shlipak MG. Decision analysis. In: Friedland DJ, ed. *Evidence-based medicine: a framework for clinical practice*. Stamford, CT: Appleton & Lange, 1998:35–57