IVI INNOVATION AND VALUE INITIATIVE

11100 Santa Monica Boulevard, Suite 500
Los Angeles, CA 90025

2 Bethesda Metro Center, Suite 850
Bethesda, MD 20814

Phone: 310 984 7793
Fax: 310 982 6311

Technical Report

Expanding Cost-Effectiveness Analysis to All of Health Care: Comparisons between CEAs on Pharmaceuticals and Medical/Surgical Procedures

Version: 2.0

Authors:

James Baumgardner, PhD
Senior Research Economist, Precision Health Economics

Michelle Brauer, BS
Associate Research Scientist, Precision Health Economics

Michelle Skornicki, MPH
Senior Research Scientist, Precision Health Economics

Peter Neumann, ScD
Director, Center for the Evaluation of Value and Risk in Health at the Institute for Clinical Research and Health Policy Studies at Tufts Medical Center; Professor of Medicine, Tufts University School of Medicine

February 21, 2018

Funding

This study was funded by the Innovation and Value Initiative, a collaboration among thought leaders in academia, patient advocacy organizations, payers, pharmaceutical companies, providers, delivery systems and other organizations dedicated to preserving innovation, value, and choice in the healthcare system.

Disclosures

James Baumgardner, Michelle Brauer, and Michelle Skornicki are employees of Precision Health Economics, a health economics consultancy providing services to the life sciences industry. Peter Neumann is Director of the Center for the Evaluation of Value and Risk in Health at Tufts Medical Center and a principal scientific advisor to Precision Health Economics.

About the Innovation and Value Initiative

The Innovation and Value Initiative (IVI) is a multi-stakeholder initiative that seeks to improve the way value is measured and rewarded in the healthcare system to promote the development and use of high value interventions that advance human health. To achieve this, IVI pursues the following goals:

- Establish best practices for measuring the real-world value of healthcare technologies using both existing and innovative scientific methods;
- Provide a range of marketplace stakeholders – including patients, consumers, providers, healthcare systems, and payers – with salient, accurate, and actionable information about value in healthcare;
- Develop and test innovative approaches to link healthcare spending to value.

The IVI is hosted by Precision Health Economics, a health economics consultancy. IVI's direction and research agenda are determined in collaboration with its Strategic Advisory Panel, which includes representatives from patient advocacy organizations, pharmaceutical firms, academia, insurers, and health systems. All funding supports IVI's overall activities, with no funding or funder tied to specific activities or research projects.

Acknowledgements

Anirban Basu of the University of Washington served as an external reviewer and provided helpful comments on an earlier version of this report. Mark Linthicum, Ahva Shahabi and Katie Everson of Precision Health Economics provided advice and assistance in earlier phases of the project.

Table of Contents

List of Tables

List of Figures

List of Appendix Tables

List of Abbreviations

Abbreviation	Description
CEA	Cost-effectiveness analysis
FDA	Food and Drug Administration
Med/Surg	Relating to a CEA evaluating only medical and surgical procedures
Mixed	Relating to a CEA evaluating both pharmaceuticals and medical or surgical procedures
Pharm	Relating to a CEA evaluating only pharmaceuticals
QALY	Quality-adjusted life year
RCT	Randomized controlled trial
Second Panel	Second Panel on Cost-Effectiveness in Health and Medicine

Executive Summary

The number of cost-effectiveness analyses (CEAs) on healthcare interventions has increased dramatically – from 34 per year in the 1990 to 1999 period to over 500 per year in the 2010 to 2014 period.[1, 2] There is, however, an imbalance in the types of healthcare interventions on which these CEAs have been conducted. Between 2010 and 2012, 46 percent of CEAs were conducted on pharmaceuticals even though pharmaceuticals account for less than 15 percent of total personal healthcare expenditures in the US. In contrast, only 22 percent of CEAs evaluated medical or surgical procedures, yet these interventions comprise more of total health care spending than pharmaceuticals.[1, 3, 4] Although there may be legitimate reasons that require disproportionately more information about the cost-effectiveness of pharmaceuticals, at first blush it would appear that the production of information on the value of goods and services would roughly reflect their proportion in overall spending – for example, it makes sense for a consumer to take much more time and resources to assess the value of a potential house purchase than a box of salt.

A more representative balance of CEA activity might better serve the information needs of decision makers and patients. Two types of inefficient decisions can occur from the relative lack of CEAs on medical/surgical procedures: 1) the use of procedures that are not cost-effective is more likely; and 2) opportunities for efficient substitution between pharmaceutical and medical/surgical interventions (in either direction) may be missed.

In this study we collected and analyzed data on the characteristics of CEAs that were published in 2015 and the effectiveness studies that provided measures of health benefit used in those CEAs. Our primary interest was to look for differences between analyses and studies on pharmaceuticals versus those on medical/surgical procedures for clues about the causes of the imbalance toward disproportionately more pharmaceutical CEAs.

Key Questions

One of our initial hypotheses was that the availability of suitable data was a reason for the imbalance – that there were simply better data available on pharmaceuticals. Six times as many clinical trials are done on pharmaceuticals than medical/surgical procedures and that greater availability of clinical trial data may allow and lead to more CEA activity on pharmaceuticals.[5] That 'type of data' reason is not fully satisfying, however, because advanced statistical techniques exist for dealing with potential biases in observational data. Thus, we were interested in examining whether trial data, in particular, randomized controlled trial (RCT) data were more often used for pharmaceutical analyses, if observational data were more often used for medical/surgical analyses, and whether advanced statistical techniques for dealing with endogeneity (or confounding) bias were being applied when observational data were used.

We were also interested in exploring the roles of property rights and sponsorship. A lesson from economics is that poorly enforced or poorly defined property rights (or a lack of property rights) can lead to under-provision. In our context, CEAs are a source of information that may be under-provided if property rights are lacking. In such cases, government subsidization can encourage production of that information. We were therefore interested in examining whether both property rights and industry sponsorship were more common for pharmaceutical CEAs and effectiveness studies. Also, because better, or unbiased, effectiveness measures come from either RCT data or observational data when coupled with advanced statistical techniques, we were interested in seeing how property rights and sponsorship were related to the type of data and statistical technique used.

In addition, we explored the comprehensiveness of CEAs, again contrasting pharmaceutical with medical/surgical CEAs. In particular, we looked at how well analyses conformed to more recent guidelines from the Second Panel on Cost-Effectiveness in Health and Medicine [6] ('Second Panel') and also examined other items, such as time horizon and inclusion of patient or caregiver time burden or non-health costs in the analysis.

Data and Results

Our data analysis used a sample of 201 CEAs and 424 effectiveness studies that provided measures of effectiveness for those CEAs. The CEA sample came from the Tufts Medical Center Cost-Effectiveness Analysis Registry (cearegistry.org), a database of published, English-language CEAs presented in the form of costs per quality-adjusted life years gained (QALYs). Details on the database are provided elsewhere.[7] We randomly selected analyses from CEAs published in 2015 and sought a balance of 100 CEAs from those classified as dealing with pharmaceuticals and those analyzing medical or surgical procedures. For each CEA, we traced the sources of effectiveness measures used and randomly selected up to 3 effectiveness studies used for each CEA. For those studies we collected data on characteristics, such as the type of data used, the type of statistical technique used, the type of sponsor, and whether there were property rights for the intervention studied.

The final sample included 201 CEAs and 424 effectiveness studies. The CEAs and their related effectiveness studies fall into 3 general categories: pharmaceutical only; medical/surgical procedure only; or mixed. The latter category captures analyses that looked at both a pharmaceutical and medical/surgical interventions as part of the analysis.

Figure E1. Likelihood of Each Characteristic, by Type of CEA

		Sponsorship		Type of Data		Property Rights	Statistical Technique	
		Industry	*Government*	*RCT*	*Observational*		*Advanced Statistical Technique*	*Multivariable Regression*
Cost Effectiveness Analyses	Pharmaceutical Only	51.6%	21.1%					
	Medical/Surgical Only	17.1%	29.3%					
	Mixed	20.8%	37.5%					
Effectiveness Studies	Pharmaceutical Only	71.6%	13.2%	82.2%	17.3%	86.8%	5.9%	60.9%
	Medical/Surgical Only	23.6%	30.9%	27.5%	71.4%	50.0%	3.9%	47.2%
	Mixed	44.9%	51.0%	51.0%	49.0%	65.3%	8.3%	61.2%

Note: CEA = cost effectiveness analysis; RCT = randomized controlled trial; Medical/surgical includes CEA conducted on medical procedures or surgical procedures; Advanced statistical techniques include: Instrumental variables, difference-in differences, and regression discontinuity, similar results were obtained when propensity score matching was included; Advanced statistical technique was examined only among effectiveness studies that used observational data, while multivariable regression results were obtained from an analysis of all 424 effectiveness studies regardless of the type of data used.

Our main findings for the CEAs are as follows:

- Pharmaceutical industry sponsorship was more common for pharmaceuticals (52%) than for medical/surgical (17%) or mixed (21%) CEAs.
- Government sponsorship was more common for mixed (38%) and medical/surgical (29%) analyses than for pharmaceuticals (21%) (those differences were not statistically significant).
- On issues of technique, there appeared to be a common standard across CEAs with no significant differences between pharmaceutical, medical/surgical, or mixed with respect to characteristics such as compliance with recommendations of the Second Panel, time horizon, inclusion of time costs, inclusion of non-medical costs, and use of probabilistic sensitivity analysis.

Our main findings for the effectiveness studies are the following:

- Industry sponsorship was much more common with pharmaceuticals (72%) than medical/surgical (24%) and mixed (44%).
- Government sponsorship was most common for mixed (51%) followed by medical/surgical (31%) and pharmaceutical (13%).
- RCT data were much more commonly used for pharmaceuticals (82%), while observational data were typically used for medical/surgical (71%) studies.
- A logistic regression showed that both the existence of a property right and industry sponsorship significantly increased the likelihood that RCT data were used.
- Even in an analysis limited to medical/surgical studies, property rights were associated with a greater probability of the use of RCT data (39% compared with only 16% of cases without property rights).
- Use of advanced statistical techniques for observational data was rare, with no statistically significant differences between pharmaceutical (6%), medical/surgical (4%), or mixed studies (8%) that used observational data.
- Multivariable regression, which can suffer from biases when applied to observational data but can provide insights beyond simple comparisons of frequencies, was more likely to be used in pharmaceutical studies (61%) and mixed studies (61%) than in medical/surgical (47%). Relatedly, multivariable regression was also more likely to be used if there was an industry sponsor and if RCT data were used.

Discussion

Though there are plausible reasons that could justify a disproportionate number of CEAs toward particular types of health care interventions, the results suggest potential policies that could lead to a more balanced mix of CEAs across pharmaceutical and medical/surgical interventions if that were considered desirable. A lesson from textbook economics is that government subsidization can be especially important in areas where property rights are lacking because of the breakdown between social and private valuations, and costs, when property rights cannot be well defined or enforced. Our results illustrated the importance of property rights as being associated with use of RCT data (and that result was even

observed in a sub-analysis on only medical/surgical studies), although that may reflect the supply of such data owing to regulatory requirements. It is plausible that both regulatory requirements and property rights were necessary to bring about the level of RCT evidence that we observe for pharmaceuticals.

The evidence suggests that government already appears to be working toward establishing a balance, with greater government sponsorship seen in medical/surgical and mixed applications. We also saw that medical/surgical studies use mainly observational data, while more plentiful pharmaceutical work tends to use RCT data. Policies that encourage more RCTs on the medical/surgical side would make more of such 'gold standard' data available and could lead to a greater number of medical/surgical CEAs and greater overall balance.

Finally, increasing the awareness and acceptance of advanced statistical techniques for dealing with the potential biases in observational data could also lead to more medical/surgical CEAs. We found that use of techniques like instrumental variables was rare, even though the vast majority of studies that are done on medical/surgical procedures make use of observational data.

1. Introduction

For several decades, cost effectiveness analysis (CEA) has been used by policymakers and payers to determine the amount of value delivered by a good or service at a given price. Interest in and utilization of CEAs by health care policymakers and payers has peaked in recent years as prices have risen and the industry has experienced a movement toward value based pricing. The number of CEAs on healthcare interventions has increased markedly from an average of 34 per year in the 1990-1999 period to more than 500 per year in the 2010-2014 timeframe.[1, 2]

Research has shown that CEAs are disproportionately conducted on pharmaceutical interventions as opposed to other types of healthcare, such as medical and surgical procedures. Between 2010 and 2012, about 46% of published healthcare CEAs were conducted on pharmaceuticals even though pharmaceutical spending is less than 15% of total personal healthcare expenditures.[3, 4] In contrast, only 22% of CEAs evaluated medical/surgical procedures[1], although calculations performed on Medicare data indicate that spending on such procedures exceeds spending on pharmaceuticals (See Appendix 9.1) The imbalance in CEAs toward pharmaceuticals and away from medical/surgical procedures raises a concern as to whether a proper balance of information is available to guide efficient decisions in healthcare delivery. Two types of problems may occur because of the imbalance: 1) procedures may be used when it is not cost-effective to do so; 2) opportunities for efficient substitution between a pharmaceutical or procedural mode of treatment may be missed. Although there may be legitimate reasons that the amount of information on the net value of goods and services may not be strictly proportional to the amount spent on them, the apparent imbalance warrants further exploration.

In a recent post on the Health Affairs blog, Baumgardner and Neumann (2017) brought attention to the imbalance in CEAs between pharmaceutical versus medical/surgical procedure applications.[8] They hypothesized that type of data was a key driver of the imbalance, because pharmaceuticals have 6 times as many clinical trials as compared to medical/surgical procedures.[5] But other hypothesized factors included a limited use or familiarity with advanced statistical techniques for use with observational data and the far more likely occurrence of clear property rights for pharmaceuticals.

In this study we compare the attributes of CEAs, and the related effectiveness studies to support them, on pharmaceuticals versus CEAs conducted on medical/surgical procedures. The analysis seeks to shed light on the underlying reasons for the disproportionate differences in the types of healthcare intervention undergoing CEAs and explores policy changes that could move toward greater balance.

2. Objectives

Our main goal is to find potentially important differences in characteristics of pharmaceutical CEAs versus CEAs on medical/surgical procedures as well as the effectiveness studies used in those CEAs. A third category of CEA, which we describe as 'mixed,' evaluated at least one pharmaceutical and one medical/surgical procedure evaluated in the analysis. We often simplify our exposition by referring to 'pharmaceutical' versus 'medical/surgical procedure,' but we will also draw comparisons with the 'mixed' category in our analyses.

Our study is exploratory in nature. We have collected data in order to allow us to measure the characteristics of CEAs and the effectiveness studies that underlie them and to see whether the data are consistent with several general hypotheses. In particular, we are interested in exploring the following:

1) **Type of Data and Statistical Technique**
   The type of data and statistical technique used are related issues. There are two ways to get unbiased measures of effectiveness: 1) use data from randomized controlled trials (RCT), or 2) if observational data is used, then use an appropriate statistical technique for dealing with endogeneity in the choice of intervention (often characterized as confounding).[9] We collected data by reviewing effectiveness studies used as inputs into CEAs. We looked for differences between pharmaceutical and medical/surgical procedure studies in terms of data and statistical techniques used. One hypothesis is that more CEAs are conducted on pharmaceuticals because RCT data are more prevalent for drugs. We are also interested in seeing whether medical/surgical procedures are less likely to use RCT data given less regulation and the lack of FDA-approval process for procedures. However, even if RCT data are less available for medical/surgical procedures, observational data could be used to produce unbiased estimates of effectiveness if advanced statistical techniques were employed. We looked at the relative use of RCT data between studies that feed pharmaceutical versus medical/surgical CEAs and, in cases where observational data were used, we looked for the use of an appropriate statistical technique such as instrumental variables.

   Although it is not a technique that deals with endogeneity or confounding bias, we additionally examined the use of multivariable regression to determine if there were any differences in its use between pharmaceutical versus medical/surgical applications or other characteristics of interest.

2) **Existence of Property Rights and Sponsorship**
   Standard economic reasoning points to a lack of property rights as the fundamental cause of problems related to public goods and externalities – if there are no property rights, society may need government subsidization (or taxation in the case of negative externalities) in order to obtain efficient resource allocation. In the health care

context, the reasoning would be that if there are not property rights, society may obtain too few studies of an intervention unless there is government subsidization of those studies. One hypothesis is that property rights are less likely to exist for procedures and that may at least partially explain the disproportionate number of CEAs conducted on pharmaceuticals. This cannot be tested directly because we can only observe situations in which a CEA was actually performed. Nonetheless, the data we collected can provide suggestive evidence. We also looked at whether the existence of property rights impacted the use of an advanced statistical technique or affected the type of data used, as there may be less incentive to obtain RCT data or employ more difficult statistical techniques where property rights do not exist. Related arguments can be made about sponsorship. Industry sponsorship, which implies a private entity is putting 'skin in the game,' may occur more often when property rights exist and may also be associated with greater use of RCT data or use of advanced techniques when observational data is used. Government sponsorship may be needed (or may be relatively helpful) in ensuring that the RCT is conducted when there are not property rights to the intervention(s) being analyzed.

Additionally, we are interested in examining characteristics of studies where the entity has property rights and sponsors both the CEA and effectiveness study. Specifically, we are interested in identifying the proportion of those 'multi-stage projects' sponsored by industry versus government versus another type of entity to determine if one sponsorship type is more likely.

3) **Consistency with Recently Revised CEA Guidelines**

We also examined whether study type affects the degree to which a CEA would already have been compliant with the new guidelines put forth by the Second Panel on Cost-Effectiveness in Health and Medicine (Second Panel).[6] The Second Panel's guidelines were published in 2016, while the CEAs in our data analysis were all published in 2015. Thus, this is not a test of adherence to these guidelines, per se, but does indicate whether analyses will have to raise their standards in order to meet the new guidelines. We also examined whether pharmaceutical or medical/surgical CEAs were more likely to meet the new guidelines.

4) **Expansiveness of CEAs**

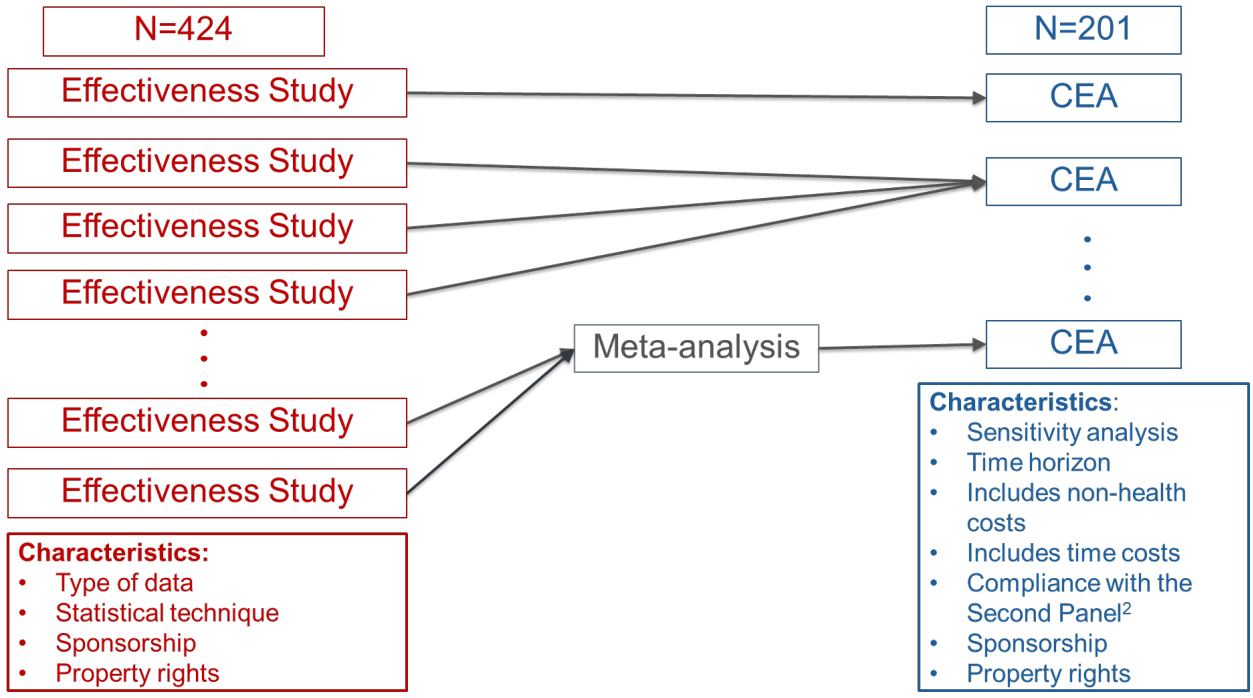
We also measured how expansive or comprehensive pharmaceutical CEAs are in comparison to medical/surgical procedure CEAs; we considered analyses to be more expansive to the extent they included a longer time horizon, patient or caregiver time burden, non-health costs or reported a cost/QALY threshold. We also identified the proportion of pharmaceutical versus medical/surgical CEAs that conducted a probabilistic sensitivity analysis.

3. Methods

As depicted in Figure 1, the typical CEA makes use of measures of effectiveness that are conducted in separate effectiveness studies. Sometimes a CEA will take its effectiveness parameters from a meta-analysis study, which in turn has synthesized information from a number of effectiveness studies.

Our analysis focuses on the characteristics of two elements in Figure 1: 1) the CEAs, and 2) the effectiveness studies that are used by the CEAs, whether or not there is an intermediate meta-analysis. Figure 1 also lists the types of characteristics that we look at for the effectiveness studies and CEAs, respectively.

Figure 1. Study Schematic



Notes: Second Panel = the Second Panel of Cost-Effectiveness in Health and Medicine

We randomly selected approximately 100 pharmaceutical and 100 medical/surgical procedure CEAs published in 2015 and extracted relevant data already captured in the Tufts database. To randomly select 100 pharmaceutical CEAs, a random number generator (Microsoft Excel 2013) was used to identify 100 of the 213 pharmaceutical CEAs and 100 of the 120 medical/surgical CEAs. For medical/surgical CEAs, we aimed to identify 50 medical procedure CEAs and 50 surgical procedure CEAs. However, fewer than 50 surgical CEAs published in 2015 exist in the Tufts CEA database. Therefore, we included all 46 surgical CEAs and supplemented with 54 medical procedure CEAs, selected using a similar simple randomization technique. Of our final data set of 201 CEAs, 95 were pharmaceutical only, 82

studied medical or surgical procedures only, and 24 were mixed (i.e. both pharmaceutical and procedures were considered in the study).

The database includes variables related to CEA study sponsorship, time horizon, inclusion of costs related to patient or caregiver time, as well as other direct and indirect health outcomes components (see Table 1), inclusion of a cost per quality-adjusted life-year (QALY) threshold, and the perspective for evaluation of costs and outcomes (societal, health care payer, etc.). To supplement the data in the Tufts registry, we identified CEAs that conducted probabilistic sensitivity analyses. We flagged each CEA with a study type identifier based on intervention (pharmaceutical, medical or surgical procedure, or 'mixed' pharmaceutical and medical/surgical procedure).

For each CEA, we randomly selected up to three studies that were sources for that CEA's measures of the effectiveness of the intervention and collected relevant data. To select studies, the random number generator was used to generate 3 random numbers from a range corresponding to the number of effectiveness studies used in the CEA. The effectiveness studies corresponding to the numbers output by the random number generator were selected for inclusion in our analysis. We categorized the effectiveness studies based on data source (RCT, quasi RCT, claims, registry, or other observational data source) and sponsorship type (government, industry, other). Quasi-RCTs were defined as those with a control group, but lacking randomization (see Appendix 9.2 for examples). We also flagged studies that had the same entity sponsoring both the effectiveness study and CEA. Lastly, we created variables that identified whether or not property rights exist for the main intervention being analyzed (see Appendix 9.3 for examples), and determined the statistical techniques used to determine effectiveness (instrumental variables, difference in differences, regression discontinuity, propensity score matching, multivariable regression, or simple comparisons).

3.1. CEA-level analysis

First, we compared characteristics of studies at the CEA level, examining aspects of the CEAs themselves, before honing in on attributes of the underlying effectiveness studies.

At the CEA level, we calculated the proportion of the 201 studies that were industry sponsored, the proportion that were government sponsored, and the likelihood that the study used a probabilistic sensitivity analysis. We also examined the relative distributions of analytic time horizons for pharmaceutical, medical/surgical and mixed analyses.

Next, we evaluated the level of adherence to cost-effectiveness recommendations published by the Second Panel. Specifically, we identified variables in the Tufts database that align with the lists of 7 and 21 components the Second Panel recommends for CEAs that take a payer perspective and societal perspective, respectively (Table 1). We calculated percent

adherence as a simple proportion (number of components included divided by number of components recommended and available in the database x 100%).[a]

Lastly, we evaluated CEA expansiveness by separately calculating the proportion that included a time horizon, non-health costs, time costs (defined as patient or caregiver time), and a cost/QALY threshold.

Table 1. Identification of Variables in Tufts Registry that Align with Second Panel Recommendations

Second Panel	Tufts
1. Longevity effects	No Tufts variables identified; assumed all CEAs include either longevity effects, HrQoL or other health effects and considered recommendations #1-3 as one component.
2. Health-related quality-of-life effects (HrQoL)	
3. Other health effects (e.g. adverse events and secondary transmissions of infectious diseases)	
4. Medical costs paid by 3rd party payers	Costs_DirectMedical
5. Medical costs paid for by patients	Costs_Outofpocket
6. Future related medical costs	FutureCosts_HealthCareRelated
7. Future unrelated medical costs	FutureCosts_HealthCareUnRelated (Future costs related to other conditions or if the economic evaluation states that all medical care was considered)
8. Patient time costs	Costs_PatientTime
9. Unpaid caregiver time costs	Costs_CaregiverTime
10. Transportation costs	Costs_Transportation
11. Labor market earnings lost	Costs_IncomeLoss
12. Cost of unpaid lost productivity due to illness	Costs_ProductivityGains
13. Cost of uncompensated household production	N/A
14. Future consumption unrelated to health	FutureCosts_NonHealthCare (Future costs outside of health care, other costs for added life year)
15. Cost of social services as part of intervention	N/A
16. Number of crimes related to intervention	N/A
17. Cost of crimes related to intervention	Costs_Legal
18. Impact of intervention on educational achievement of population	Costs_Education
19. Cost of intervention on home improvements	Costs_Housing
20. Production of toxic waste pollution by intervention	Costs_Environment
21. Other impacts	Costs_SectorsOther

Notes: Second Panel = the Second Panel on Cost-Effectiveness in Health and Medicine; The Second Panel recommends including components 1-7 for a payer-perspective CEA and all 21 components for a societal perspective CEA.

[a] We calculated a simple measure of adherence – a ratio of items included over total items recommended by the Second Panel. A more sophisticated, and complicated, assessment would consider that some recommended components are not relevant depending on the intervention and context being evaluated.

3.2. Effectiveness-study level analysis

For the 424 effectiveness studies, we examined the sponsorship type, whether or not property rights exist for the intervention, the type of data employed, and the statistical technique used. To categorize effectiveness studies as pharmaceutical, medical/surgical or mixed, we drew from the categorization of the CEA that it supported. For this reason, an effectiveness study that evaluated a pharmaceutical may fall under the mixed category if the CEA that it supported evaluated both pharmaceutical and medical/surgical interventions.

To observe how sponsorship differs by study type, we calculated the proportion of each study type sponsored by industry (a pharmaceutical or device company) and the proportion sponsored by government (a government institution or government grant). We separately examined sponsorship type among studies sponsored by the same entity as that sponsoring the associated CEA, and identified the proportion of studies by study type where the entity has property rights and sponsors both the CEA and supporting effectiveness study.
To explore the type of data underlying each analysis, we calculated the proportion of pharmaceutical, medical/surgical and mixed effectiveness studies with the following data sources (evaluated as yes/no):

1. RCT data
2. RCT or quasi RCT data
3. Observational data

Next, we estimated the proportion of each effectiveness study type that had property rights for the intervention analyzed, and among that subset, the proportion that used RCT data. We also used multiple logistic regression to examine how factors (property rights, sponsorship) may affect a study's likelihood of using RCT data. We repeated this regression (regressing property rights and industry sponsorship on use of RCT) on the subset of medical/surgical effectiveness studies only, to separate any association between pharmaceutical studies and RCT data. As we expected most pharmaceutical studies to have property rights, we also calculated the percentage of studies using RCT data for those with and without property rights in the subset of medical/surgical studies.

Next, we calculated the proportion of pharmaceutical, medical/surgical, and mixed effectiveness studies that reported using any of the following types of statistical techniques:

1. Instrumental variables, difference-in differences, regression discontinuity, propensity score matching, or multivariable regression[b]
2. Instrumental variables, difference-in differences, regression discontinuity, or propensity score matching

[b] For 'multivariate regression' we included use of either least squares regression, the Cox proportional hazards model, or logistic regression with multiple independent variables.

3. Propensity score matching
4. Simple multivariable regression techniques (excluding other methods listed above)

Then we conducted a series of regressions to examine factors that may affect the likelihood of using a special statistical technique. First, we examined the use of techniques designed for use with observational data – instrumental variables, difference-in-differences, regression discontinuity, and propensity score matching (#2 above). For that set of analyses, we limited the data to the subset of relevant effectiveness studies – namely, those that use observational data (N=185), and used property rights and industry sponsorship as explanatory variables. We conducted a second regression to examine likelihood of using multivariable regression techniques, which unlike the other special statistical techniques, may be useful in both RCT- and observational data-based studies. Here we included covariates for property rights, industry sponsorship, and RCT data, and analyzed the entire set of effectiveness studies.

All calculations of proportions of studies with categorical dichotomous variables equal to 0 or 1 were calculated using the proc freq SAS procedure with the Pearson chi-square test option (for overall significance, and pairwise comparisons of interest). In cases where the number of observations was not large enough, we employed Fisher's exact test for pairwise comparisons. For the continuous time horizon variable, we conducted a Kruskal-Wallis test to compare medians across study type groups. All regressions, because the dependent variables were 1-0 indicators for the use of a type of data or the use of a set of statistical techniques, were conducted using the proc logistic SAS procedure. P values were obtained from the Wald chi-square statistic for each predictor. We estimated the correlation between predictor variables using the proc corr SAS procedure. We evaluated significance at the $p<0.05$ level.

4. Results

4.1. Analysis at the CEA Level

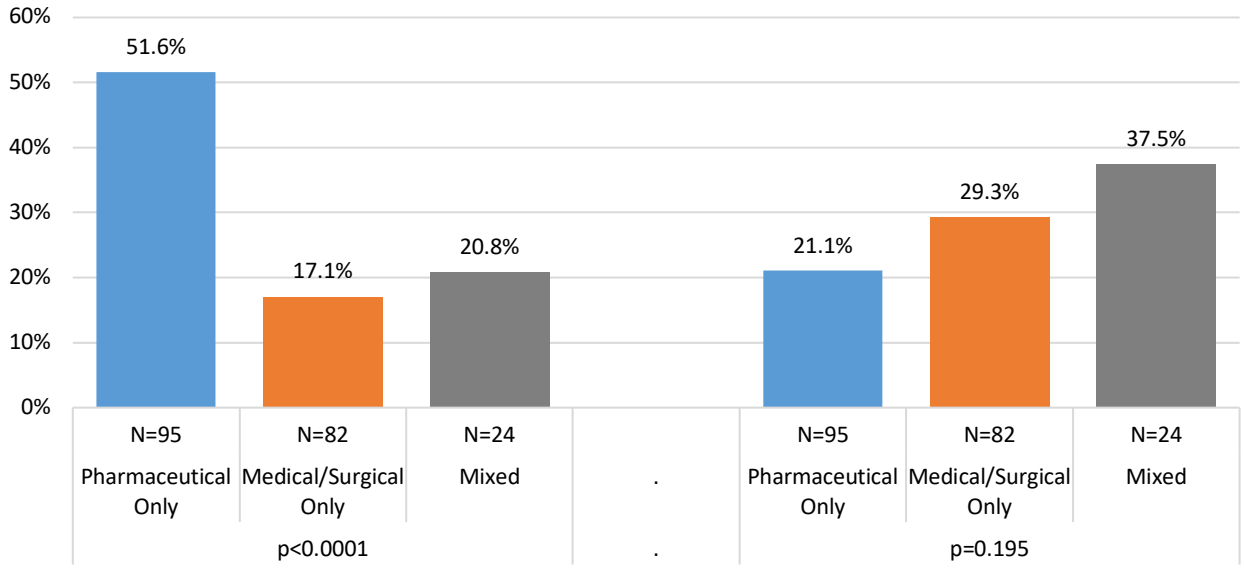
In our analysis of characteristics at the CEA-level, we were primarily interested in assessing analysts' motivations for conducting CEAs, the expansiveness of CEAs, and the degree to which studies complied with recently revised guidelines on the conduct of CEAs. In general, we were exploring for clues explaining the imbalance of CEAs toward pharmaceutical applications. Hence, we were interested in looking for differences between pharmaceutical and medical/surgical CEAs in several dimensions. Because industry sponsorship suggests that a financial stake and government sponsorship may be indicative of government provision of information as a public good, we sought to determine the likelihood of pharmaceutical CEAs' and medical/surgical CEAs' being sponsored by industry versus government. Second, we evaluated how comprehensive pharmaceutical or medical/surgical analyses were in including time burden and non-health costs and reporting of a time horizon and cost/QALY threshold. Third, we examined each type of CEA's likelihood of including components in their analysis as recommended by the Second Panel on Cost-Effectiveness in Health and Medicine. [6] Through this analysis, we hoped to provide insight into the degree

to which analyses of each type would have to change their approach to cost and value assessments to comply with recent guidelines.

Industry sponsorship was significantly more common for pharmaceutical CEAs than medical/surgical CEAs (Figure 2). Specifically, we found that 51.6% of pharmaceutical CEAs were sponsored by industry while 17.1% of medical/surgical analyses were sponsored by industry ($p<0.0001$). Among mixed CEAs, 20.8% were sponsored by industry.

Although government sponsorship was less frequent among pharmaceutical CEAs (21.1%) and successively more common among medical/surgical (29.3%) and mixed analyses (37.5%) ($p=0.195$), the differences were not statistically significant. The lack of statistically significant results could be a function of a lack of power due to the small sample size. To further examine this issue, we re-visit this question in an analysis of effectiveness studies for which we have a larger sample (See Figure 6).

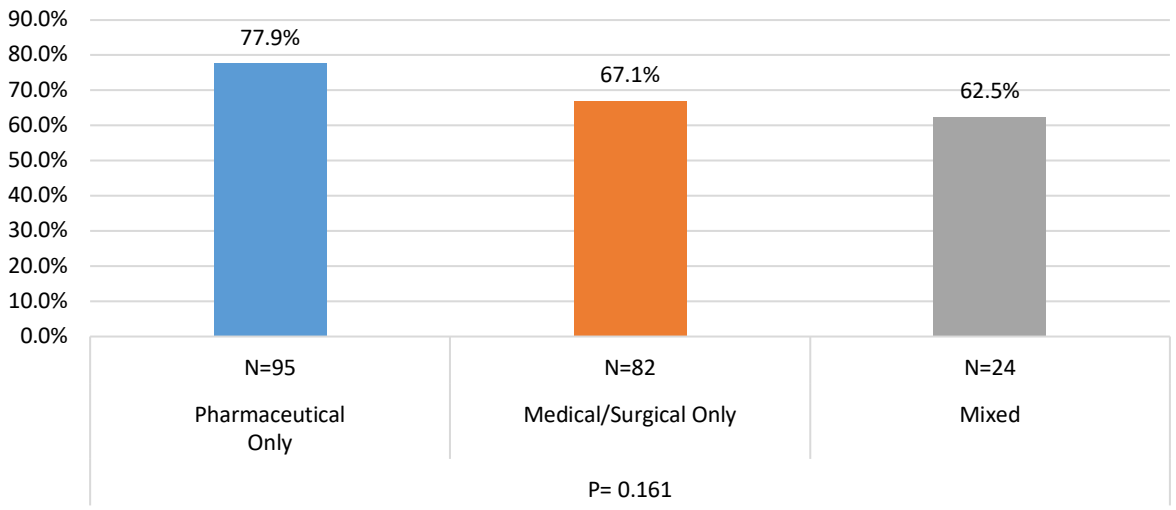
Figure 2. CEA Industry and Government Sponsorship, by Type of CEA



Notes: CEA = cost effectiveness analysis

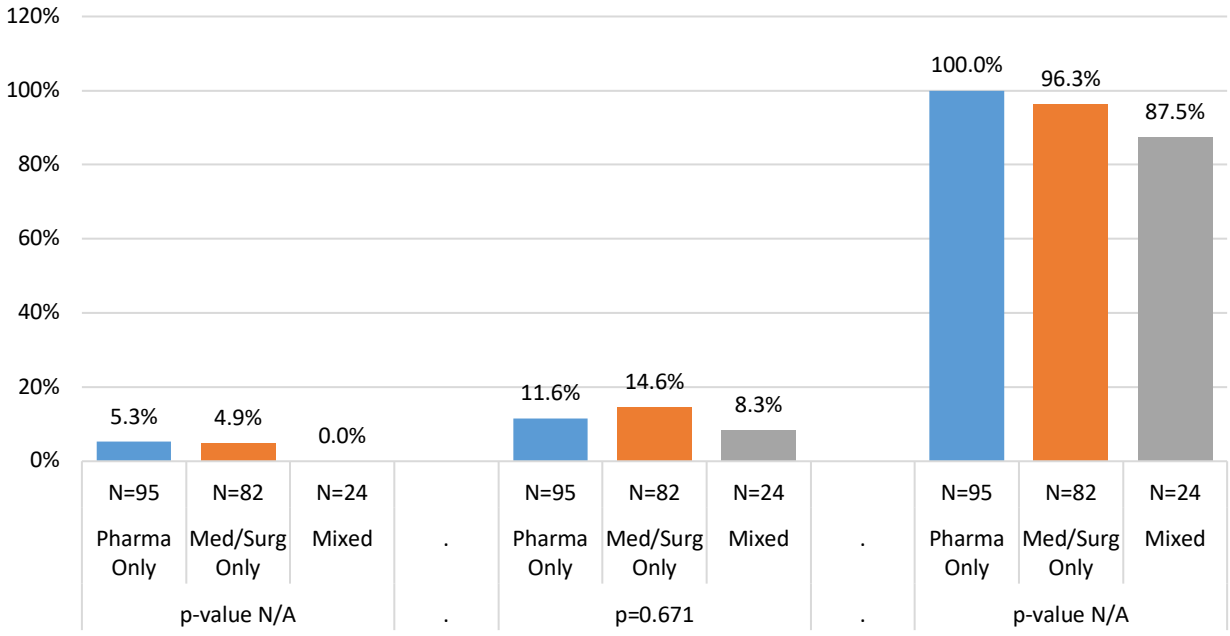
There was no statistically significant relationship between use of probabilistic sensitivity analysis and type of CEA (Figure 3). Among our sample, the largest proportion of CEAs to employ a probabilistic sensitivity analysis were pharmaceutical 77.9%) followed by medical/surgical (67.1%) and mixed analyses (62.5%) ($p=0.106$).

Figure 3. Percent CEAs with Probabilistic Sensitivity Analysis, by Type of CEA



Notes: CEA = cost effectiveness analysis

Figure 4. Percent CEAs Reporting Time Burden (Patient or Caregiver Time), Non-Health Costs, and Cost/QALY Threshold, by Type of CEA



Notes: CEA = cost effectiveness analysis; Med/Surg = medical and surgical procedures; N/A = not available; Pharm = pharmaceutical; QALY = quality adjusted life year

Inclusion of time burden, non-health costs and cost/QALY threshold was not significantly related to the type of CEA (Figure 4). Time burden was included infrequently across all types of CEAs (pharmaceutical CEAs: 5.3%, medical/surgical CEAs: 4.9%, mixed: 0.0%; p=1.000). Slightly more medical/surgical CEAs included non-health costs in their analysis (14.6%) than pharmaceutical (11.6%) and mixed (8.3%) CEAs, however the difference was not significant (p=0.547). A cost/QALY threshold was mentioned in all pharmaceutical CEAs, most medical/surgical CEAs (96.3%) and slightly less often in mixed analyses (87.5%) (p=0.060).

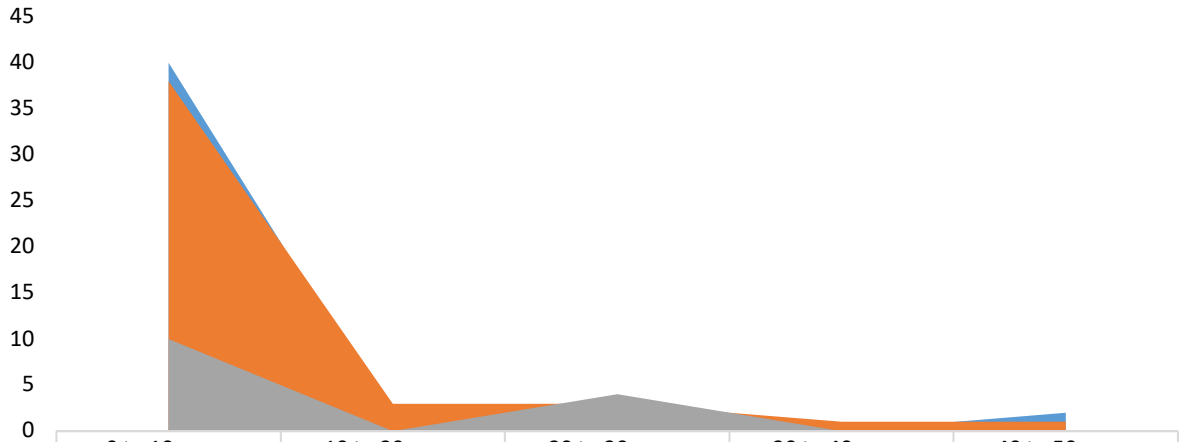
Generally, pharmaceutical CEAs tended to use a shorter time horizon than medical/surgical and mixed analyses (Table 2). For studies identified as having a "lifetime" horizon, we set the variable equal to 50 years for simplicity. Though the time horizon for both pharmaceutical and medical/surgical analyses ranged from 3 months to 50 years, or a lifetime, the median among pharmaceutical analyses was 2.6 years compared to 5 years for medical/surgical analyses. Comparing the third quartile values, this difference is more pronounced: 75% of pharmaceutical analyses used a time horizon of 5.5 years or shorter whereas 75% of medical/surgical analyses used a time horizon of 10 years or less. Note, however, that the differences in time horizons for the three types of CEAs are not significant (Kruskal-Wallis test, p=0.1432). The frequency distributions of time horizons by type of CEA are displayed in Figure 5. It is worth noting that a prior analysis over a longer time frame (2005 to 2014) found that 71% of 782 studies had a time horizon of >5 years. [10]

Table 2. Time Horizon Distribution

	Pharmaceutical (N=44)	Medical/Surgical (N=46)	Mixed (N=14)
Maximum	50	50	30
75% Q3	5.5	10	30
50% Median*	2.6	5	4.7
25% Q1	1	1	2
Minimum	0.25 (3 months)	0.25 (3 months)	0.25 (3 months)
Missing	N=51	N=36	N=10

Notes: Q1 = first quartile; Q3 = third quartile; *The differences in time horizons for the three types of CEAs are not significant (Kruskal-Wallis test, p=0.1432); Assigned a 50 year time horizon to analyses whose horizon was described as "lifetime"

Figure 5. Frequency Distribution of Time Horizon (Years)



	0 to 10 years	10 to 20 years	20 to 30 years	30 to 40 years	40 to 50 years
Pharma	40	1	1	0	2
Med/Surg	38	3	3	1	1
Mixed	10	0	4	0	0

Notes: Med/Surg = medical and surgical procedures; Pharma = pharmaceutical

Compliance with the recommendations of the Second Panel was similar across all types of CEAs. We assumed that in order to calculate cost effectiveness ratios, all CEAs included either longevity effects, health-related quality of life or other health effects, even though they were not explicitly captured in the Tufts database. We grouped these outcomes as one component, "health outcomes". Of the 201 CEAs, Tufts reviewers categorized 173 studies as evaluating costs and outcomes using a payer perspective and 25 CEAs as taking a societal perspective, leaving three CEAs un-categorized. Among the 173 payer perspective analyses that were examined, 65 out of 68 pharmaceutical CEAs and 81 out of 83 medical/surgical CEAs and all 22 mixed analyses adhered to 2 out of 5 components recommended by the Second Panel: health outcomes and direct medical costs. Two pharmaceutical CEAs and 3 medical/surgical CEAs adhered to 3 out of 5 components (addition of out of pocket costs). However, none of them included future related and future unrelated medical costs.

Only 25 of our sample of 201 CEAs conducted analyses from the societal perspective. The 11 pharmaceutical CEAs adhered to 22% of the 16 recommended components on average, while the 12 medical/surgical and mixed CEAs adhered to an average of 24% and 22% of the 16 components on average.

Our analysis highlights the degree to which all types of CEAs need to evolve over time to meet the new recommendations for the conduct of CEAs. Compliance with the Second Panel's Recommendations was particularly low among analyses conducted from the societal perspective. This discordance may reflect the fact that there is a much longer list of recommended components for societal perspective studies. Components incorporated by some CEAs from this perspective included: direct medical, productivity gains, income loss, transportation, caregiver time, patient time, and other sectors costs. None of the CEAs

incorporated future costs, or costs related to education, housing or the justice system, or the environment in their analyses.

Though there is clear need for improved conduct of CEAs, caution must be exercised when interpreting the results. Not all components may be relevant to each intervention; therefore, the context of each intervention must be considered. Additionally, even though these analyses were conducted prior to the publication of the Second Panel's Recommendations, many of the recommendations were also included in prior published guidelines.[12] Thus, our results could be interpreted as the degree to which CEAs violated contemporaneous best practice guidelines as well as the degree to which they must evolve to meet the new guidelines.

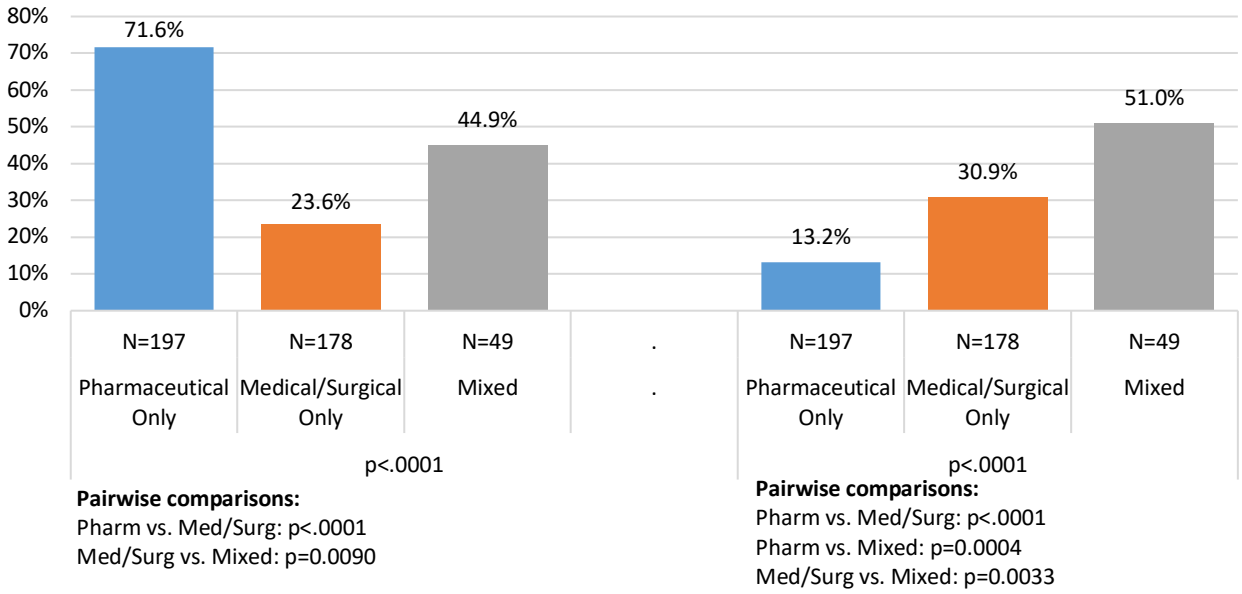
In summary, there are no significant differences between pharmaceutical and medical/surgical CEAs on matters of inclusion of components and conduct; however, significant differences exist in sponsorship, with industry sponsorship much more common among pharmaceutical CEAs. (Issues related to type of data and statistical technique are considered with respect to effectiveness studies in subsequent sections of the report.)

4.2. Analysis at the Effectiveness Study Level

In almost all cases CEAs determined efficacy from the results of other studies of the effectiveness of an intervention. We sought to determine if the likelihood of key characteristics of these studies underlying the CEAs significantly differed between pharmaceutical and medical/surgical studies. We were particularly interested in investigating sponsorship and data type, as we hypothesized that the wide availability of RCT data for pharmaceuticals may be a reason for the imbalance in the number of CEAs towards pharmaceutical applications. As an additional analysis, we sought to examine the likelihood of a property right's being associated with each study type and whether a medical/surgical CEA with a property right was significantly more likely to use RCT data. If it turned out that medical/surgical studies more often were using observational data, the inherent problems could potentially be overcome with certain advanced statistical techniques. Thus, we were also interested in the use of particular statistical techniques that can deal with potential biases that exist in observational data.

At the effectiveness study level, industry sponsorship is significantly more common for pharmaceutical effectiveness studies than medical/surgical studies, while government sponsorship is significantly more common for medical/surgical effectiveness studies than pharmaceutical studies (Figure 6). Specifically, we found that 71.6% of pharmaceutical effectiveness studies were sponsored by industry while 23.6% of medical/surgical effectiveness studies had industry sponsorship ($p<0.0001$). In contrast, 13.2% of pharmaceutical effectiveness studies had government sponsorship while 30.9% of medical/surgical and 51% of mixed studies were sponsored by government ($p<0.0001$). While this pattern mirrors the results of our CEA-level sponsorship analysis, in this case the differences in percent of government sponsorship were statistically significant.

Figure 6. Percent of Effectiveness Studies with Industry and Government Sponsorship, by Type of CEA

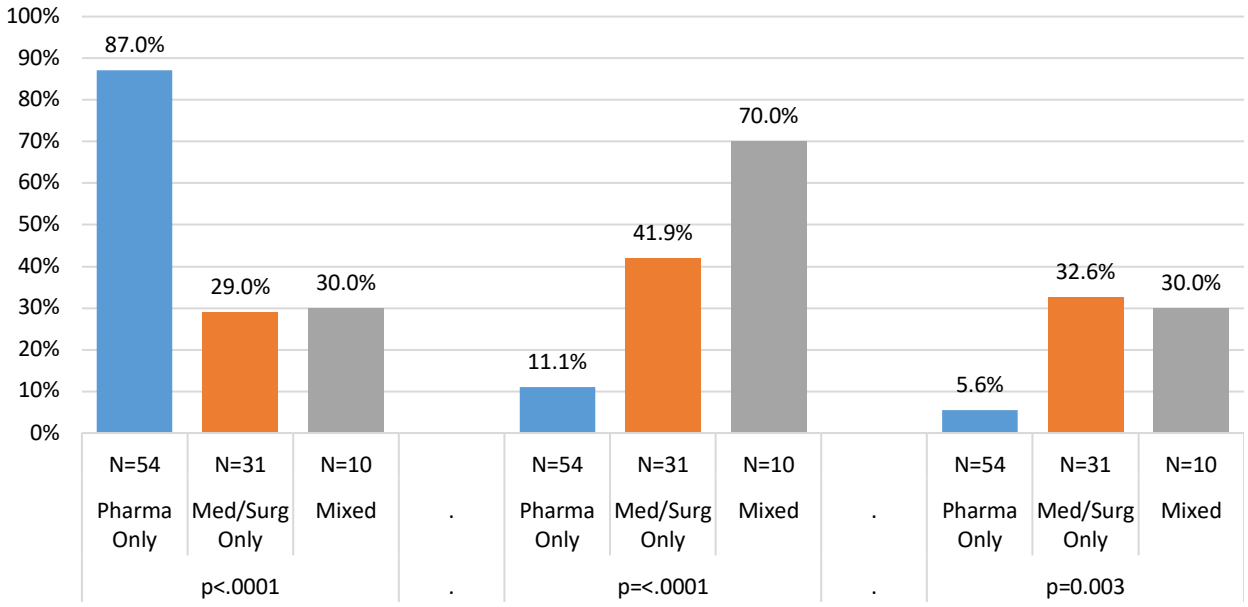


Notes: CEA = cost effectiveness analysis; Med/Surg = medical and surgical procedures; Pharm = pharmaceutical

We also evaluated cases that appeared to be part of a complete research program in which the same entity that sponsored an effectiveness study was also the entity sponsoring the CEA. Clearly, if an entity had a strong interest in an intervention, it might get involved in sponsorship at both levels.

Compared to medical/surgical and mixed studies, pharmaceutical studies that were sponsored by the same entity as the CEA were significantly more likely to have industry sponsorship ($p<0.001$)(Figure 7). Government sponsorship was significantly more likely among mixed studies ($p<0.001$) that had same sponsorship of effectiveness study and CEA. Medical/surgical and mixed studies were significantly more likely than pharmaceutical cases to have another entity, such as a non-profit sponsor the research underlying the CEA ($p=0.003$) as well as the CEA. Results of this analysis suggest that industry is more highly motivated to sponsor these multi-level programs that produce both effectiveness studies and CEAs for pharmaceuticals. The greater likelihood of government sponsorship for mixed cases may be indicative of governmental interests in seeking efficient use of resources among multiple types of healthcare interventions. Multi-level research programs on medical/surgical interventions were more likely than pharmaceuticals to be government sponsored. That may be indicative of governments filling the void by sponsoring areas not covered by private entities.

Figure 7. Percentage of Studies with Industry, Government, or Other Sponsorship in Cases Where the Same Entity Sponsors the Effectiveness Study and CEA



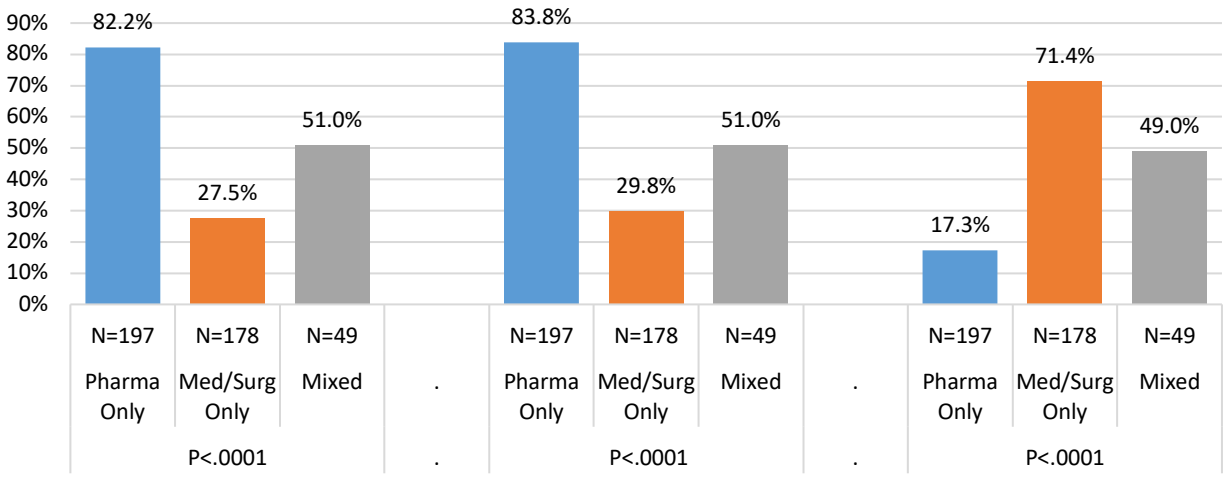
Note: Cases where the Tufts registry coded a study with two sponsorship types (N=10) is displayed in both applicable figures (6 cases where study was flagged as government or other, 2 where it was flagged as industry or other, and 2 where it was flagged as government and industry).

4.3. Type of Data

We also examined whether medical/surgical procedures were less likely to use RCT data, perhaps due to a less rigorous regulatory approval pathway. We also looked at the underlying roles of property rights and industry sponsorship as motivators for making use of RCT data.

In exploring the type of data underlying each analysis, we found that use of RCT data was significantly more common among pharmaceutical studies, while use of observational data was significantly more common among medical/surgical studies (Figure 8). Specifically, 82.2% of the data supporting pharmaceutical CEAs came from RCTs compared to 27.5% for medical/surgical CEAs ($p<0.0001$) (Figure 8). Similar results were obtained when quasi-RCTs (controlled trials that were not randomized) were included in the analysis. This trend is reversed for observational data, however, with 71.4% of medical/surgical CEAs obtaining effectiveness data from observational studies compared to 17.3% for pharmaceutical studies ($p<0.0001$).

Figure 8. Effectiveness Study Data Type by Type of CEA



Notes: CEA = cost effectiveness analysis, Med/Surg = medical and surgical procedures; Pharma = pharmaceutical; RCT = randomized controlled trial

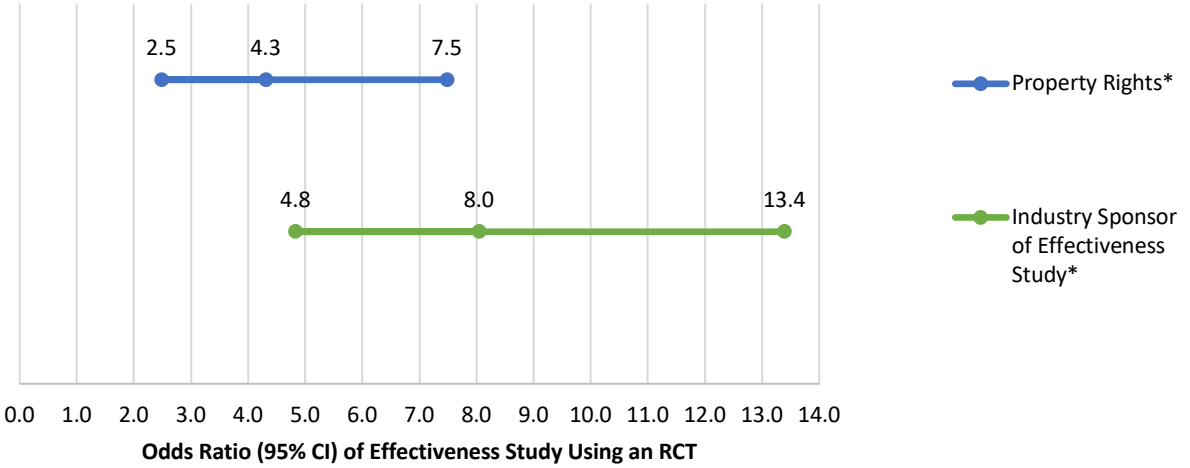
Regressing property rights and industry sponsorship on use of RCT data showed that having property rights and industry sponsorship significantly increases the likelihood that an effectiveness study uses RCT data. The log odds of using RCT data is estimated to be 1.46 times higher for studies with property rights versus studies with no property rights, given industry sponsorship stays constant ($p<0.0001$). When the study is industry sponsored, the change in log odds of using RCT is expected to be 2.08, holding property rights constant (Table 3) ($p<0.001$). Expressed as odds ratios, the likelihood of using RCT is 4.3 times higher if there are property rights for the intervention, holding sponsorship constant. Holding property rights constant, having industry sponsorship increases odds of using RCT data eightfold (Figure 9).

Table 3. Logistic Regression: Likelihood of Using an RCT

Parameter	Estimate	p-value
Property Rights	1.46	p<0.0001
Industry Sponsor of Effectiveness Study	2.08	p<0.0001

Note: Interaction terms were not significant

Figure 9. Logistic Regression: Likelihood of using an RCT



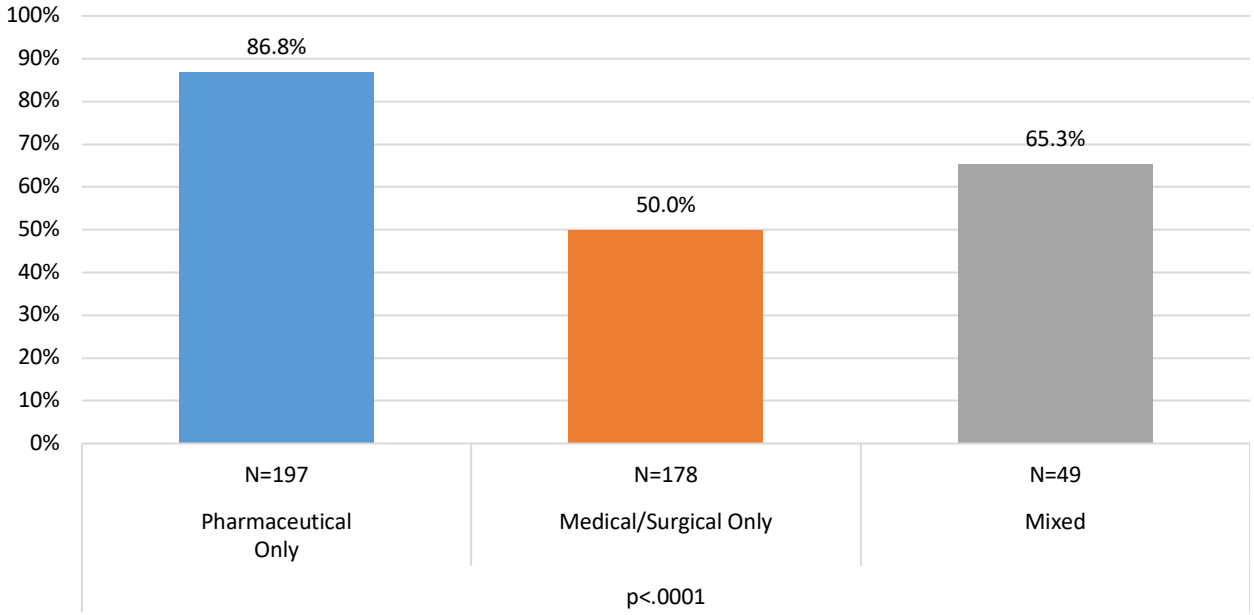
Note: *while simultaneously controlling for other included variable; CI = confidence interval; RCT = randomized controlled trial

Property rights were significantly more common for pharmaceutical studies than for medical/surgical studies with almost all pharmaceutical studies (86.8%). Further, 50% of medical/surgical studies had a property right associated with the effectiveness study intervention ($p<0.0001$) (Figure 10). Pharmaceutical products typically have a property right associated with them.

We also replicated this analysis on the subsample of only medical/surgical studies. The goal of this analysis was to address the possibility that property rights were essentially a proxy for pharmaceuticals with their strict regulatory requirements for producing RCT data.

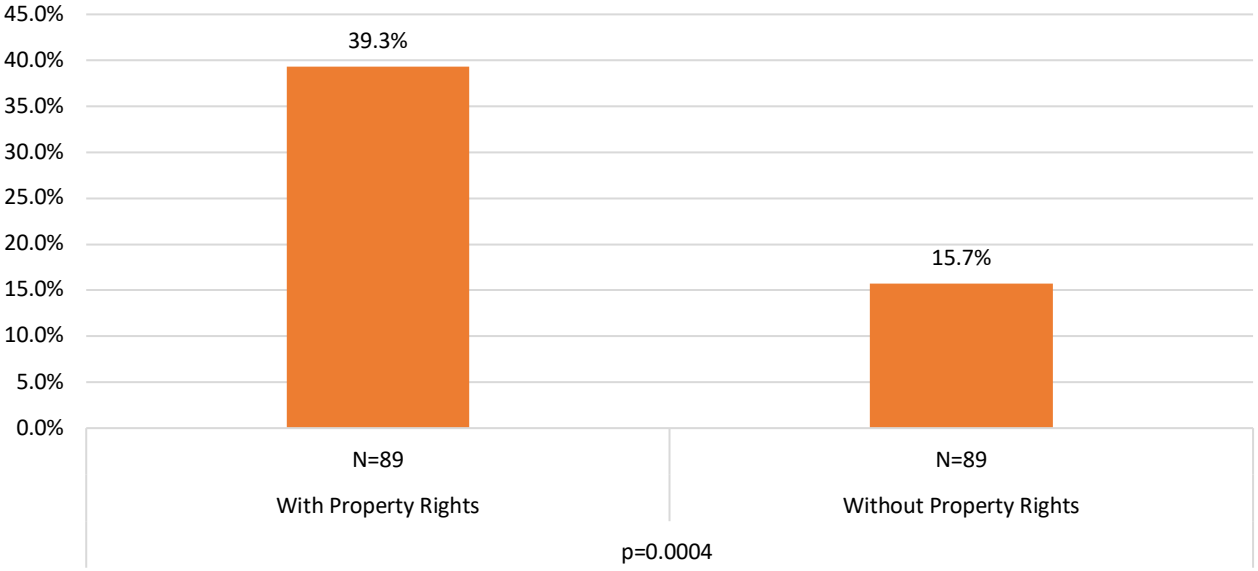
Among medical/surgical effectiveness studies that had property rights, significantly more used RCT data than other types of data. We found that 39.3% of medical/surgical studies with property rights used RCT data compared to 15.7% without property rights ($p=0.0004$) (Figure 11).

Figure 10. Percent of Effectiveness Studies with Property Rights Associated with the Intervention, by Type of CEA



Notes: CEA = cost effectiveness analysis; Med/Surg = medical and surgical procedures; Pharma = pharmaceutical

Figure 11. Percent of Medical/Surgical Effectiveness Studies using RCT data, by Property Right



Notes: RCT = randomized controlled trial

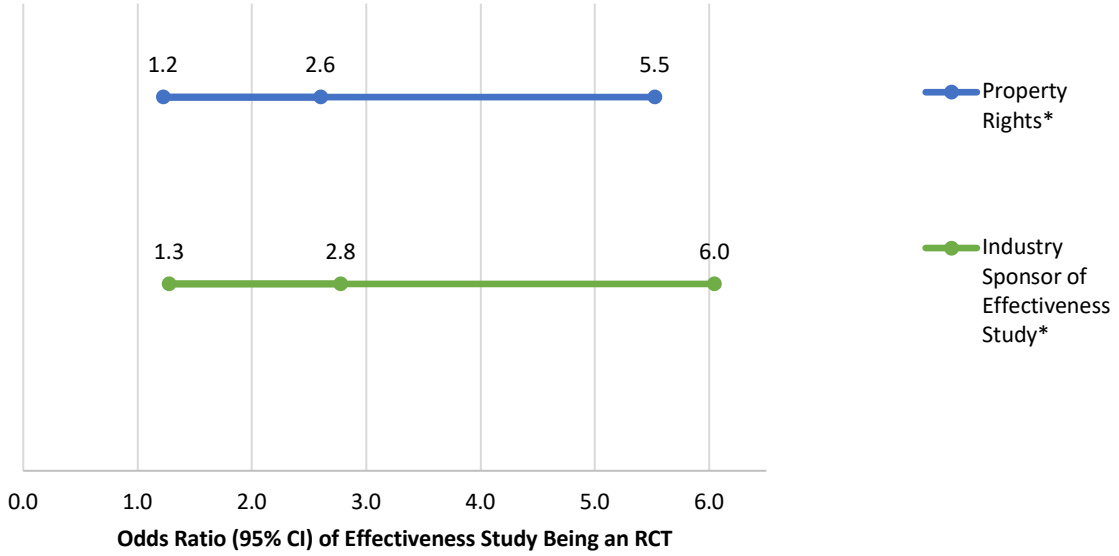
Having property rights and industry sponsorship significantly increased the likelihood that a medical/surgical effectiveness study used RCT data. The exact coefficients from this analysis and corresponding p-values are shown in Table 4. The corresponding odds ratios and 95% confidence intervals are OR=2.6 (1.2, 5.5), and 2.8 (1.3, 6.0) for property rights and industry sponsorship, respectively (Figure 12). In other words, holding industry sponsorship constant, having property rights results in a 160% increase in the odds of using RCT data. Holding property rights constant, having industry sponsorship increases odds of using RCT by 180%.

Table 4. Logistic Regression: Likelihood of Using an RCT (Medical/Surgical Studies Subset).

Parameter	Estimate	p-value
Property Rights	0.95	p=0.01
Industry Sponsor of Effectiveness Study	1.02	p=0.01

Note: Interaction terms were not significant

Figure 12. Logistic regression: Likelihood of Using an RCT (Medical/Surgical Studies Subset).



Notes: *while simultaneously controlling for other included variable; CI = confidence interval; RCT = randomized controlled trial

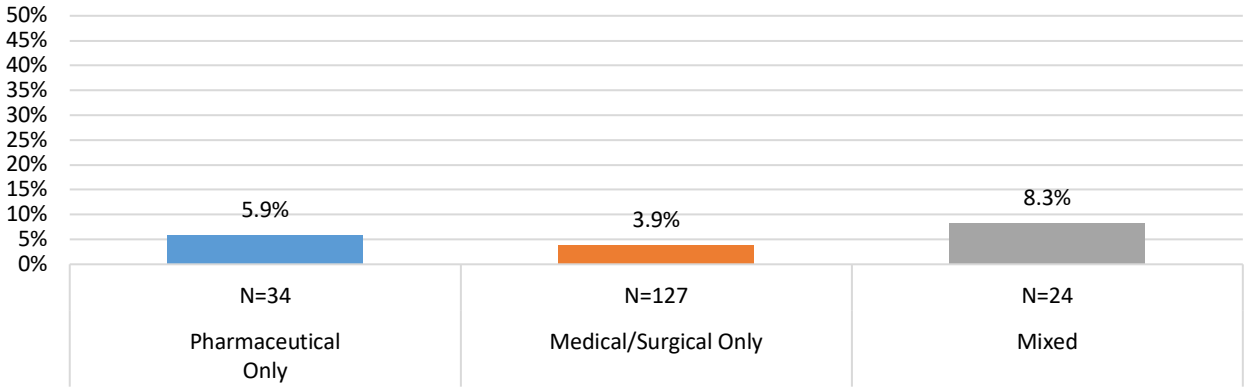
5. Use of Statistical Methods

5.1. Use of Advanced Statistical Techniques for Observational Data

Advanced statistical techniques are often required to get unbiased measures of effectiveness when using observational data. Use of methods such as instrumental variables, regression discontinuity, and difference-in-differences can eliminate the problem of endogeneity (often characterized as a case of unmeasured confounders). This set of analyses was intended to demonstrate the likelihood of using such techniques or propensity score matching among studies that used observational data and to look for any differences depending on whether the studies were used by pharmaceutical, medical/surgical, or mixed CEAs. Additionally, a set of regression analyses were performed to determine study characteristics that significantly influence the likelihood of using an advanced statistical technique. Despite not overcoming the endogeneity problem, propensity score matching was included in our main analysis, because in some cases it has been considered an advanced statistical technique used with observational data. [9]

Among effectiveness studies analyzing observational data, use of advanced statistical techniques did not significantly differ between pharmaceutical, medical/surgical, or mixed studies (Figure 13). Though utilization was low for all types of CEAs, mixed studies most frequently employed advanced statistical techniques (8.3%) while medical/surgical studies least often used these methods (3.9%). None of these differences was statistically significant.

Figure 13. Percent of Observational Studies Using Advanced Statistical Techniques, by Type of CEA



No pairwise comparisons are significant at p<0.05

Notes: CEA = cost effectiveness analysis

PSM accounts for all use of advanced statistical technique for pharmaceutical and mixed studies and 3.2% of the 3.9% for medical/surgical studies.

When we examined the likelihood of using advanced statistical techniques among observational studies in a multiple logistic regression, neither having property rights nor industry sponsorship significantly changes the likelihood that an observational study used advanced statistical techniques (instrumental variables, difference in differences, regression discontinuity and propensity score matching). We tested models where the predictor variables (property rights and industry sponsorship) were tested individually (Table 5), and another model in which we controlled for the effects of both predictors in one model (Table 6). In both cases we failed to reject the null hypothesis that the regression coefficients are zero ($p>0.5$). The corresponding odds ratios (Figure 14) and 95% confidence intervals (overlapping 1.0) show that neither property rights nor industry sponsorship affect the likelihood of using advanced statistical techniques. These results are not surprising given the observation that the techniques were rarely used – there was no variation to explain.

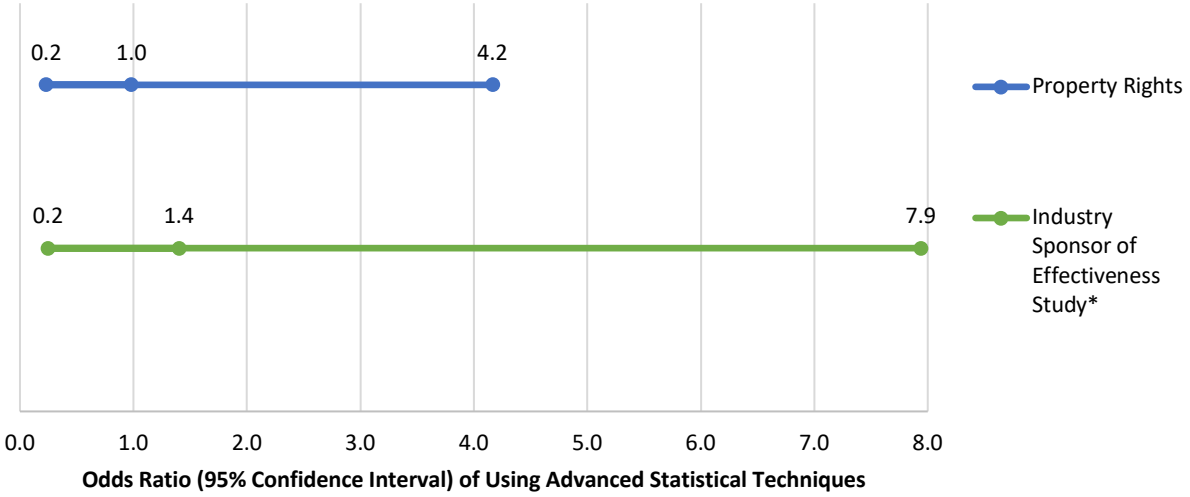
Table 5. Logistic Regression: Likelihood of Using an Advanced Statistical Technique (Two Separate Models)

Model	Parameter	Estimate	p-value
1	Property Rights	0.0745	P=0.9138
2	Industry Sponsor of Effectiveness Study	0.3296	P=0.6900

Table 6. Logistic Regression: Likelihood of Using an Advanced Statistical Technique (One Model)

Parameter	Estimate	p-value
Property Rights	-0.018	p=0.9806
Industry Sponsor of Effectiveness Study	0.337	p=0.7031

Figure 14. Logistic Regression Results: Likelihood of Using an Advanced Statistical Technique



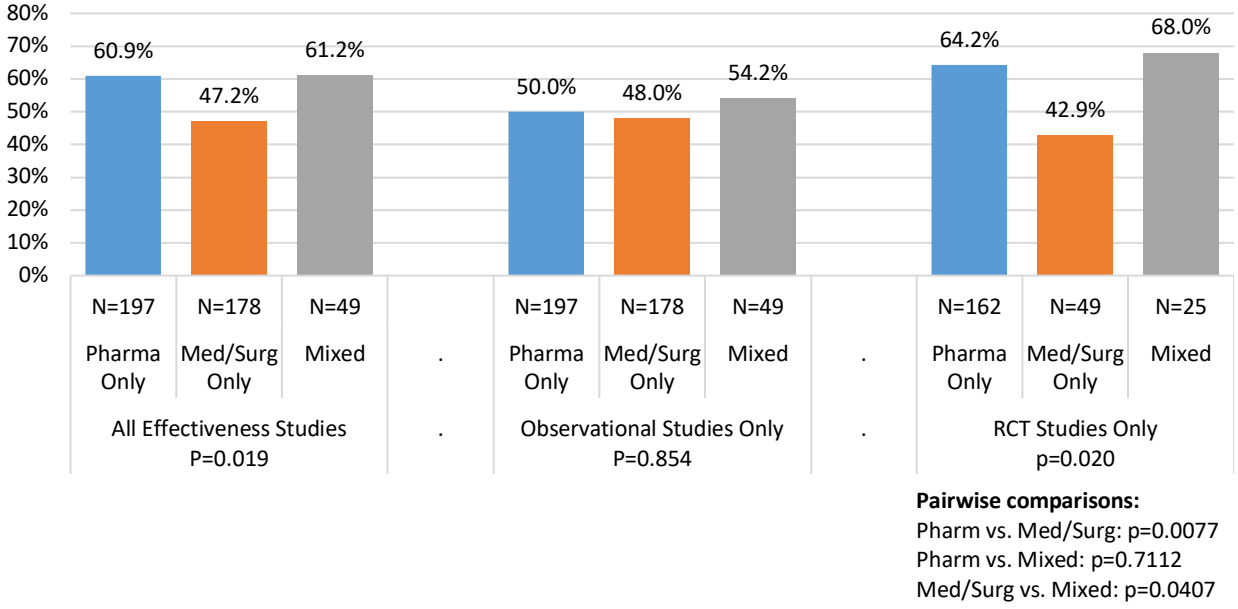
Note: *while simultaneously controlling for other variable; CI = confidence interval

5.2. Use of Multivariable Regression

Multivariable regression in its 'simple' form is not a technique that overcomes the endogeneity or confounding bias that can be eliminated by the techniques that we labeled as 'advanced;' however, multivariable regression was frequently used by effectiveness studies and we were interested to determine if it was applied disproportionately to pharmaceutical or medical/surgical studies, and how other relevant characteristics may influence the likelihood of a study employing this technique. We included both least squares regression, Cox proportional hazard models, and logistic regressions within our definition of multivariable regression as long as there was more than one right-hand side variable included (not counting a constant term).

When we performed a simple comparison of the use of multivariable regression by type of CEA, we found that this technique was significantly more likely among pharmaceutical and mixed studies compared to medical/surgical studies, and this difference in significance was primarily driven by a difference in its use among RCT studies (Figure 15). Among all effectiveness studies, we found that 60.9% of pharmaceutical studies and 61.2% of mixed studies used multivariable regression in their analysis while only 47.2% of medical/surgical studies used this technique (p=0.019). When analyzed by data type, 64.2% of the pharmaceutical RCTs used multivariable regression and 68.0% of mixed RCTs studied used this technique relative to 42.9% of medical/surgical RCT studies. Among studies that used observational data, no significant difference in use of multivariable regression was observed between different CEA types.

Figure 15. Percent of Effectiveness Studies Using Multivariable Regression, by Type of Effectiveness Study and Type of CEA



Notes: CEA = cost effectiveness analysis; Med/Surg = medical and surgical procedures; Pharm = pharmaceutical; RCT = randomized controlled trial

When we regressed individual predictors such as property rights, industry sponsorship, and use of RCT data on whether a study employs multivariable regression, we found that individually, industry sponsorship and RCT data significantly increase the likelihood that an effectiveness study employs multivariable regression ($p<0.05$) (Table 7).

Table 7. Logistic Regression: Likelihood of Using Multivariable Regression (Three Separate Models)

Model	Parameter	Estimate	p-value
1	Property Rights	-0.00671	p=0.9746
2	Industry Sponsor of Effectiveness Study	0.4167	p=0.0341
3	RCT data	0.4551	p=0.0211

Note: RCT = randomized controlled trial

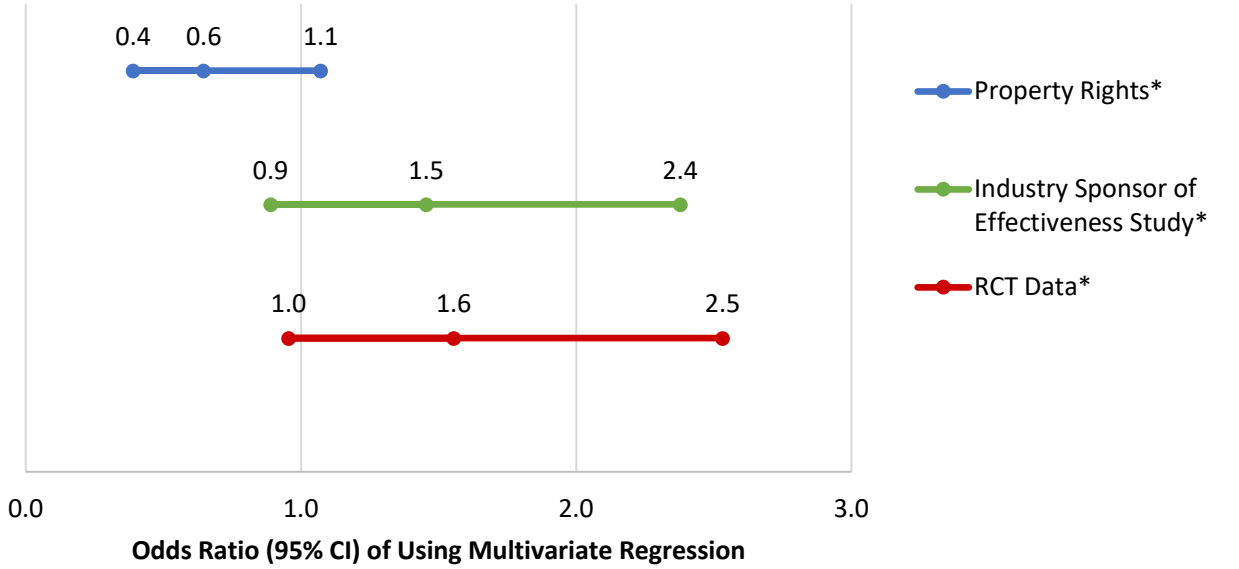
Including all three characteristics in a multivariable logit led to non-significant model coefficients (see Table 8 and Figure 16). As expected, the variables used as predictors in this model are highly correlated, with Pearson correlation coefficients of almost 0.5 for all of the following pairs of variables: property rights and industry sponsorship, RCT data and property rights, RCT data and industry sponsorship (see Table 9). All correlations are significant at $p<0.0001$. The observed multicollinearity may partially explain the non-significant findings in the multiple logistic regression, although it is true that despite the correlation between property rights and industry sponsorship both were found significant in explaining the use of RCT data back in Tables 3 and 4 and Figures 9 and 12.

Table 8. Logistic Regression: Likelihood of Using Multivariable Regression (One Multivariable Model)

Parameter	Estimate	p-value
Property Rights*	-0.4355	p=0.0908
Industry Sponsor of Effectiveness Study*	0.3749	p=0.1347
RCT data*	0.4413	p=0.0761

Notes: *while simultaneously controlling for other factors; RCT = randomized controlled trial

Figure 16. Logistic Regression: Likelihood of Using Multivariable Regression (One Multivariable Model)



Notes: *while simultaneously controlling for other factors; CI = confidence interval; RCT = randomized controlled trial

Table 9. Correlation Matrix among Independent Variables Included in Multiple Logistic Regression

	Property Rights	Industry Sponsor	RCT data
Property Rights	1.0		
Industry Sponsor	0.498 p<0.0001	1.0	
RCT data	0.476 p<0.0001	0.560 p<0.0001	1.0

Note: Correlation coefficients can take on values between 0 and 1, where 1 indicates perfect correlation.

6. Discussion

Our key findings emerged from examining the importance of sponsorship, the use of RCT data, the association with property rights, and the use of advanced statistical techniques. Industry sponsorship was found to be more common for effectiveness studies used in pharmaceutical CEAs, while government sponsorship was more prominent in studies used by medical/surgical and mixed CEAs. The use of RCT data was typical for pharmaceutical studies, while use of observational data dominated medical/surgical studies. Property rights and industry sponsorship increased the odds that RCT data were used. That relationship continued to hold when the subsample of only medical/surgical studies was analyzed.

Advanced statistical techniques for dealing with potential biases in observational data were rarely used, independent of whether the studies were for pharmaceutical, medical/surgical, or mixed analyses; however, that finding is more relevant to medical/surgical studies because of their greater reliance on observational data. Among a number of other findings, we found that characteristics like inclusion of time burden and non-health costs and conformity with recommendations of the Second Panel were similar across all types of CEAs. Use of multivariable regression, which might be considered more advanced than simple comparisons but does not eliminate endogeneity or confounding biases, occurred more often in cases with RCT data and industry sponsorship, and in mixed and pharmaceutical studies. We saw that government sponsorship is already evident in encouraging medical/surgical procedure studies, possibly filling the hole created by there being fewer property rights in that area. Even more government subsidization of medical/surgical studies may be needed if a more representative mix of CEAs is viewed as desirable.

Our evidence showed that RCT data were less often used for medical/surgical studies of effectiveness and the lack of such data may be preventing as many CEAs from being produced on the medical/surgical side. That lack of RCT evidence is likely because of different regulations, with RCTs required for approval of pharmaceuticals but not for procedures. A greater balance of CEAs might be attained if both government and private entities encouraged the production of more RCTs on the medical/surgical side. Private health plans could, for example, set co-pays in a fashion that favored goods and services for which RCT-based studies supported effectiveness.

Lack of property rights may also contribute to there being less RCT data for medical/surgical procedures. Among medical/surgical studies we found that when property rights existed, RCT data were more likely to be used. It was beyond the scope of the current study to determine whether the existence of RCT data for specific studies was required by regulation or not (medical and surgical devices may be required to undergo RCTs depending on their classification).

Are property rights or regulation the fundamental cause for the associations we found between pharmaceuticals, property rights, and use of RCT data? Or, were both necessary to get to where we are today? For example, regulation may be the prima facie reason that RCT data are far more available on pharmaceuticals, and that supply of 'good data' may facilitate and be part of the reason for a greater number of CEAs on pharmaceuticals. But the existence of property rights may also be a necessary part of the equation. If there were not property rights, there would have been little incentive to go through the regulatory process and produce the RCT data. Thus, increasing regulatory requirements on medical/surgical procedures by requiring production of RCTs might not result in a favorable outcome because no one person or entity would have the incentive to incur the cost of the RCT if they could not hold or enforce property rights to the procedure.

Whether there is a practical way to assign property rights for procedures is an interesting question, but if it could be done, that would encourage the holder of the right to 'prove' the effectiveness and thereby promote use of the procedure, with royalties returning to the

holder of the property right when the procedure was used. Such property right assignment may also encourage the owner to conduct RCTs because evidence from such studies may be considered more credible than evidence using observational data. Or, as discussed above, in the absence of greater government subsidization, both property rights and regulatory requirements might be needed to produce large increases in the amount of RCT data available for studying the effectiveness of medical/surgical procedures.

Although our analysis looked for the use of RCT data, it was beyond the scope of this study to assess the quality of the analysis of those data. For example, questions about whether appropriate comparisons and statistical testing occurred in studies using RCT data were not issues that we examined. Nor did we assess whether extrapolations made in these analyses were reasonable. In the case of observational data studies, we looked for use of certain statistical techniques, but made no attempt to judge whether those techniques had been appropriately applied in the very few cases in which they were used. Perhaps unexpectedly, we found that 'simple' multivariable regression analyses (excluding techniques like instrumental variables and propensity score matching) were more common in pharmaceutical and mixed studies, and in studies using RCT data and having industry sponsorship. A possible explanation is that with greater commercial interests at stake, there was a greater desire to conduct a robust analysis. For example, multivariable regression may reveal differential effects by subsets of the population.

Greater use of advanced statistical techniques for dealing with endogeneity/confounding in observational data appears to be called for. These techniques are more familiar among economists and econometricians, who have a long history of dealing with observational data. We found that use of advanced techniques like instrumental variables was rare, which implies that the effectiveness studies that used observational data – common among medical/surgical studies – may have endogeneity bias. Perhaps a concern for such biases has limited the number of studies done with observational data and, by extension, on medical/surgical procedures. Hence, an increased awareness and acceptance among the scientific community of the relevant statistical techniques could result in more studies and ultimately more CEAs on medical/surgical procedures, and a greater balance of CEAs across types of healthcare interventions.

7. Conclusions

We studied the characteristics of CEAs and their related effectiveness studies to shed light on the imbalance in the number of CEAs performed on pharmaceuticals versus medical/surgical procedures. Our analysis suggests that the existence of RCT data, sponsorship, property rights, and a lack of adoption of advanced statistical techniques all play a role in the relatively larger number of CEAs conducted on pharmaceuticals. To the extent that more CEAs on medical/surgical procedures are desirable, potential remedies could include greater government sponsorship of studies on medical/surgical procedures (including incentives to produce RCT data) and wider adoption of statistical techniques for overcoming potential biases in observational data.

8. References

1. Neumann, P.J., et al., The changing face of the cost-utility literature, 1990–2012. Value in health, 2015. **18**(2): p. 271-277.
2. Cost-Effectiveness in Health and Medicine. 2nd ed, ed. P.J. Neumann, et al. 2017, New York: Oxford University Press.
3. Neumann, P.J., C.-H. Fang, and J.T. Cohen, 30 years of pharmaceutical cost-utility analyses. Pharmacoeconomics, 2009. **27**(10): p. 861-872.
4. Centers for Medicare and Medicaid, Table 02 National Health Expenditures; Aggregate and Per Capita Amounts, by Type of Expenditure. 2016: Baltimore, MD, USA.
5. Califf, R.M., et al., Characteristics of clinical trials registered in ClinicalTrials. gov, 2007-2010. Jama, 2012. **307**(17): p. 1838-1847.
6. Sanders, G.D., et al., Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. Jama, 2016. **316**(10): p. 1093-1103.
7. Neumann, P.J. and J.T. Cohen, Tufts Medical Center Cost-Effectiveness Analysis Registry. 2017: Boston.
8. Baumgardner, J.R. and P.J. Neumann, Balancing The Use Of Cost-Effectiveness Analysis Across All Types Of Health Care Innovations, H.A. Blog, Editor. 2017.
9. Sox, H.C. and S.N. Goodman, The methods of comparative effectiveness research. Annual review of public health, 2012. **33**: p. 425-445.
10. Forthcoming:Kim, D.D., et al., The Influence of Time Horizon on Results of Cost-effectiveness Analyses. Expert Review of Pharmacoeconomics & Outcomes Research, 2017.
11. Kim, D.D., et al., The Influence of Time Horizon on Results of Cost-effectiveness Analyses. Expert Review of Pharmacoeconomics & Outcomes Research, 2017(just-accepted).
12. Weinstein, M.C., et al., Recommendations of the Panel on Cost-effectiveness in Health and Medicine. Jama, 1996. **276**(15): p. 1253-1258.
13. Centers for Medicare & Medicaid Services, State/County Table - All Beneficiaries. Baltimore, MD.
14. The Board of Trustees Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013 Annual Report of the Board of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. 2013: Washington, D.C. p. 10.
15. Bénard, A., et al., Comparative cost-effectiveness analysis of sacral anterior root stimulation for rehabilitation of bladder dysfunction in spinal cord injured patients. Neurosurgery, 2013. **73**(4): p. 600-608.
16. Lips, E., et al., Indicators of homologous recombination deficiency in breast cancer and association with response to neoadjuvant chemotherapy. Annals of oncology, 2010. **22**(4): p. 870-876.
17. Zeuzem, S., et al., Sofosbuvir and ribavirin in HCV genotypes 2 and 3. New England Journal of Medicine, 2014. **370**(21): p. 1993-2001.

9. Appendix

9.1. Estimation of the Percent of Medicare Spending on Medical and Surgical Procedures

We were interested in establishing that there is an imbalance in pharmaceutical versus medical/surgical procedure CEAs relative to their relative representation in total health spending. From the Tufts Registry and the National Health Accounts it is known that the fraction of CEAs that evaluate pharmaceuticals (46%) is over three times the proportion of national health care spending on pharmaceuticals (<15%). It is also known that only 22% of CEAs evaluate medical/surgical procedures. If spending on procedures exceeds spending on pharmaceuticals, then a relative imbalance of CEAs toward pharmaceuticals is more than established.

While determining spending on drugs is straightforward, since it is captured in the National Health Accounts, spending on medical and surgical procedures is not specifically collected and estimating it requires some additional analysis. Thus, we obtained data for Medicare spending for subcategories of spending in parts A and B from the Centers for Medicare and Medicaid State and County Report on All Beneficiaries [13] and data for parts A, B, and D from Table II.B1 of the 2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds [14]. The key calculations are captured in appendix table 1 below. Recognize that 'spending on physicians' services for procedures' captures only the payment to the physician and that additional facility costs are often incurred when procedures are conducted. Hospital outpatient costs and ambulatory surgical center costs are incurred when procedures are done at those types of facilities. In addition, it is certainly the case that a significant amount of inpatient hospital expenses are connected to the provision of medical or surgical procedures.

Appendix Table 1. Medicare Expenditure on Parts A, B and D by Source as Fraction of Total in 2012

	Spending	Percent of total A and B Spending	Percent of A, B, and D Spending
Total Spending Parts A and B	$324,436,363,960.71	100%	88%
Spending on Physician Services for Procedures	$21,261,122,136.95	7%	6%
Spending on Hospital Outpatient	$39,395,009,400.97	12%	11%
Spending on Ambulatory Surgical Centers	$2,781,548,441.38	1%	1%
Spending on Inpatient Hospital	$109,851,927,262.96	34%	30%

From the numbers in appendix table 1 note that even if no inpatient hospital spending is attributed to medical/surgical procedures, 17% of total spending in traditional Medicare is attributable to procedures (6%+11%+1%). If half of inpatient hospital spending is attributed to medical/surgical procedures, 32% of total spending in traditional Medicare is attributable to procedures (6%+11%+1%+(0.5*30%)). To the extent that the distribution of Medicare spending by service category is reasonably representative of overall national health care spending, then this establishes the imbalance of CEAs.

9.2. Examples of Assignment of the Quasi-RCT Variable

Illustrative Examples:

- A CEA comparing treatments for spinal cord injured patients with a neurological bladder utilized data from a non-randomized controlled clinical trial comparing two treatments. Because the study had two treatment arms, one of which was the true 'treatment', the other of which, was considered the 'control', yet lacked randomization, we considered this effectiveness study to be a quasi-RCT.[15]
- In another effectiveness study used by a CEA, patients were selected for a clinical trial based on various eligibility criteria. Following patient selection, one of three treatment regimens were assigned to each patient dependent upon the prior treatments they had received. Therefore, the study was controlled, however the treatment allocation was non-random.[16]
- A CEA directed us to an effectiveness study that was initially an RCT. However, after results of a different study were published demonstrating the large benefit of treatment, all patients in the placebo group were unblended and received the treatment group therapy. While this study was difficult to categorize, we ultimately decided that it should be considered non-random treatment with a control. For this reason, it is considered to be a quasi-RCT in this analysis. [17]

9.3. Examples of Assignment of the Property Rights Variable

Illustrative Examples:

- A CEA comparing different procedures for identifying a certain type of tumor/cells pointed to an effectiveness study that used a special dye/staining technique. Though neither funder nor manufacturer was mentioned, we assumed that the dye likely has a property right associated with it.
- A medical/surgical CEA directed us to an effectiveness study that specifies manufacturers of medical equipment used in the procedure, but does not report funding. We assumed that property rights exist for the equipment used in the procedure, and are owned by the stated manufacturer.
- In an effectiveness study examining two different methods of breast reconstruction, we assumed that no property rights exist as no specific companies were mentioned in association with equipment and no special equipment was needed.