A Comment on the Devereaux et al. Meta-Analysis of Mortality in Private American Hospitals

Prepared for the Senate Committee on Social Affairs , Science and Technology

Brian S. Ferguson
Department of Economics
University of Guelph
Guelph, Ontario, Canada
N1G 2W1

August 2002

A Comment on the Devereaux et al. Meta-Analysis of Mortality in Private American Hospitals

Brian S. Ferguson
Department of Economics
University of Guelph

I. Introduction:

The Devereaux et al. (2002)[1] paper is a meta-analysis of 14 sets of results from12 studies which have as their dependent variable mortality rates associated with hospital treatment and which include, among their explanatory variables, information on the ownership of the hospitals in which the patients were treated. The Devereaux et al. paper deals with relative mortality between private for-profit (PFP) and private not-for-profit (PNFP) hospitals in the United States, and does not include information from any other countries in which for-profit hospitals operate. It does not consider mortality in American for-profit hospitals relative to American public hospitals, on the grounds that 95% of Canadian hospitals are technically private not-for-profit institutions run by their own boards of directors and administrators. In this decision the authors seem to have been perhaps overly influenced by the appearance of independence.

It is true that the day to day administration of Canadian hospitals is in the hands of hospital boards, but those boards are free to manage their hospitals only so long as they reach decisions of which the provincial ministries of health approve. If a provincial ministry decides to close a service, or move a service from one hospital to another, there is nothing the hospital board can do about it except try to raise a political fuss that might cause the minister to back down. If a hospital board decides to close a service, the ministry can prevent it. The ministry can even close an entire hospital, should a restructuring commission recommend it.

In support of their argument, Devereaux et al. cite a paper by Deber[2] (2000) whose author

---

[1]Devereaux, P.J. et al. (2002) "A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals" Canadian Medical Association Journal 28 May, 166(11), 1399-1406 (plus appendix published on-line).

[2]Raisa Deber (2000): Getting What We Pay For: Myths and Realities about Financing Canada's Health Care System Department of Health Administration, University of Toronto

argues that most Canadian hospitals should be classified as private rather than public sector on the grounds that their employees do not work directly for governments, are not civil servants and that their management does not have to follow civil service guidelines. On this evidence, most provinces' school systems would not be classified as public sector, and the military would be classed as private enterprise. It is not a terribly convincing argument.

It is also worth noting that the image most Canadian have of American not-for-profit hospitals as providing uncompensated care to the uninsured is not strictly accurate. Most of that care is actually provided by public hospitals and a few teaching hospitals. For the most part, private not for profits provide no more uncompensated care than do private for-profits. Norton and Staiger (1994)[3] find that when for-profits and not-for-profits are located in the same area, they serve equivalent numbers of uninsured patients[4]. Differences in gross figures on the number of uninsured served by different types arise because PFP and PNFP hospitals are not distributed identically across the United States. Private American hospitals, both for-profit and not-for-profit, have considerably more freedom than Canadian hospitals have to choose the mix of patients they will serve. Canadian hospitals in this regard are much more like American public hospitals than they are like American PNFP hospitals.

While there are good reasons for wanting to compare the performance of American PFP and PNFP hospitals, most notably the fact that they are much more alike than either is like American public hospitals, the argument that Canadian hospitals are more like American PNFPs than they are like American public hospitals is not convincing.

[3]Edward C. Norton and Douglas O. Staiger (1994): "How hospital ownership affects access to care for the uninsured" Rand Journal of Economics 25(1), Spring, 171-185

[4]American PNFPs are supposed to provide community service, like uncompensated care, as the price of their tax exempt status. While many do, the behaviour of many more is such that many municipalities and states are beginning to doubt that the PNFPs are living up to their end of the deal.

The Devereaux et al. paper also presents a rather misleading picture of the financial situations of PFP and PNFP hospitals in the United States. The authors say, for example, that "Typically investors expect a 10%-15% return on their investment." (Pg. 1404). If that is so, investors in the for-profit hospital sector have consistently been sadly disappointed and presumably have never learned, since the average profit margin of for-profit hospitals is on the order of 5% and has been at that figure for several decades. In most years the average profit (or, as it is more commonly termed in the not-for-profit sector, surplus) margin of not for profit hospitals has consistently been on the order of a percentage point or two less[5]. In the past couple of years PNFP margins have been significantly below those of PFP hospitals, but that was because of losses the PNFPs took on their for-profit operations. They had purchased physician practices and insurance operators which they were running on a for-profit basis, perhaps in expectation of a 10-15% return, but discovered that for profit medicine is not actually all that profitable. According to Sloan (1998)[6] in 1993 the median margin for for-profit hospitals was 4.9%, compared to 3.6% for nonprofits, and 29% of for-profits had negative margins compared with 22.2% of nonprofits. Hoerger (1991)[7] analyzed the relative variability of profits between for-profit and not-for-profit hospitals[8] - in his data set (a panel data set drawn from the period 1983-88, permitting him to investigate the effect of the change in the way U.S. Medicare paid hospitals) the mean net profit[9] in total dollar terms of not-for-profit hospitals was 40% higher than that of for-profit hospitals[10].

Mark (1999)[11] looks at a data set containing information on all private acute care hospital conversions between 1989 and 1992, and at a comparison data set consisting of 3800 acute care

---

[5]MedPAC: Report to Congress: Medicare Payment Policy March 2001

[6]Sloan, Frank A. (1998): "Commercialism in Nonprofit Hospitals" Journal of Policy Analysis and Management 17(2), 234-252

[7]Thomas J. Hoerger (1991): "'Profit' variability in for-profit and not-for-profit hospitals" Journal of Health Economics 10, 259-289

[8]He found profits to be more variable at for-profit hospitals.

[9]Net profit was defined as "net patient revenue plus total non-patient revenue minus total operating costs and total other expenses".

[10]It might well, of course, be argued that what matters is not what the profits are so much as what is done with them. It is generally assumed that PFPs distribute their profits to their investors, while PNFPs use them to advance their social objectives. This neglects the fact that many PFPs use retained earnings for investment in plant and equipment. It is also worth noting that, if PNFPs were using the profits from profitable services to subsidise unprofitable ones, their profit figures would be lower. And, of course, they have to earn the profits in the first place.

[11]Tami L. Mark (1999): "Analysis of the Rationale for, and Consequences of, Nonprofit and For-Profit Ownership Conversions" Health Services Research 34(1, Part 1), April, 83-101

private hospitals which did not convert over the same period. Comparing for-profit and not-for-profit hospitals in general (ie not just the conversion group) she found that profit margins were higher for non-profits in 1989 and 1990, equal across the two types in 1991 and higher for for-profits in 1992-95. For six of the seven years she looked at, (again, in the broader group of hospitals, not the converting hospitals) average operating expenses and average Medicare expenses did not differ across the groups. Consistent with other work on how PFPs make their profits, she found that average revenues (calculated per discharge) were higher in PFPs than in PNFPs, in all seven years (although in one of those years, her Table 6 suggests no difference). She also found that in six of the seven years she looked at, total staff-to-patient ratios and registered nurse-to-patient ratios were significantly higher in for-profits than in non-profits, and that in five of the years, total nurse-to-patient ratios were higher in for-profits.

Even if the margins are much smaller than Devereaux et al. suggest, and even though in some years nonprofits have had higher margins than for-profits, on average (Sloan (1998)), it is true that for profits have consistently had higher margins than nonprofits, if only by a percentage point or two. There has been considerable research on for-profit and nonprofit hospital financial performance[12], and the most common result in that literature has been that for-profits achieve their extra margin on the revenue side, not the cost side. Empirically, for-profit hospitals do not make their extra margin by cutting costs, they make it by earning extra revenue.

Empirically also, not-for-profit and for-profit hospitals show very much the same responses to economic incentives[13] and, perhaps more tellingly for the argument that not-for-profits are fundamentally different, both tend to take advantage of being in a monopoly position by raising prices[14]. The

---

[12]See, for example Ramesh K. Shukla, John Pestian and Jan Clement (1997): ": A Comparative Analysis of Revenue and Cost-Management Strategies of Not-for-Profit and For-Profit Hospitals" Hospital and Health Services Administration 42(1), Spring, 117-134

[13]Duggan (2000) looks at the effect of the introduction of California's Disproportionate Share Program (DSH), which rewarded hospitals for providing care to certain types of uninsured patients, on the behaviour of PFP and PNFP hospitals. He found that both PFP and PNFP hospitals practised cream skimming, increasing the share of their patient mixes made up of the now more lucrative uninsured at the expense of uninsured groups not favoured by the DSH program, whom they left to public hospitals. Both PFP and PNFP hospitals significantly increased their revenue as a result of the introduction of the program, but neither type used the extra revenue to improve medical care quality for the poor or to increase the care they provided to groups which did not benefit from DSH. Both types used their DSH revenues primarily to increase their holdings of financial assets. Duggan concludes that there is no difference in the response of the two types of hospitals to financial incentives, and also concludes that PNFPs are no more altruistic than PFPs. Mark G. Duggan (2000): "Hospital Ownership and Public Medical Spending" Quarterly Journal of Economics CXV, November, 1343-1373.

[14]See the literature reviewed in Sloan (1998) and in Brian S. Ferguson (2002): Profits and the Hospital Sector: What Does the Literature Really Say? Working Paper, Department of Economics,

"nonprofit" defence against anti-trust actions, which argued that nonprofit hospitals should be immune to anti-trust action because their nonprofit nature indicates that they are public-service oriented, is under attack by legal scholars because of the behaviour of non-profit hospitals which have achieved monopoly or dominant market positions[15].

University of Guelph, and Atlantic Institute for Market Studies.

[15]Vaughn, Amanda J. (1999): "The use of the nonprofit "defence" under section 7 of the Clayton Act" Vanderbilt Law Review 52: March, 558-597

The Devereaux et al. article is also misleading with regard to the different sources of funds available to PFP and PNFP hospitals. The implication in their discussion of investor pressure is that not-for-profits, because they cannot access the stock market, do not have external pressures. While it is true that the non-distribution requirements on PNFPs means that they cannot sell stock, they can and do sell bonds. As Sloan, Hoerger, Morrisey and Hassan (1990)[16] show, PNFPs have seen a significant decline in philanthropy over time, with philanthropy tending to decline as insurance coverage increased. In 1993, for community hospitals generally, philanthropy, grants and interest income made up just 2% of hospital revenue (less than the average community hospital made from parking and the gift shop) while Medicare and Medicaid payments made up 55% and other insurance plans made up 33%[17]. Instead, PNFP hospitals make considerable use of bond markets - it was estimated in the early 1990s that there was at that time over a hundred billion dollars (U.S.) worth of nonprofit hospital debt outstanding[18]. Bond financing is different from stock financing in one (for our purposes) significant way: bond holders take priority over stockholders. Stockholders share in the profit of an enterprise, but also share in the risk. Bondholders tend to get a smaller return than stockholders in very good years, as the price of not having to share in the loss in bad years. Stock holders can be told that there will be no dividend paid this year; it is much harder to get out of making a bond payment. In fact, stockholders can be told that there will be no dividend payment because a bond payment must be made. The fact that PNFPs cannot have stockholders means that their external investors - their bondholders - do not

[16]Sloan, Frank A., Thomas J. Hoerger, Michael Morrisey and Mahmud Hassan (1990): "The Decline of Hospital Philanthropy" Economic Inquiry 28(4), October, 725-743

[17]Getzen, Thomas E. (1997): Health Economics: Fundamentals and Flow of Funds John Wiley & Sons, New York

[18]See Gerard J. Wedig, Mahmud Hassan and Michael Morrisey (1996): "Tax-Exempt Debt and the Capital Structure of Nonprofit Organizations: An Application to Hospitals" The journal of Finance 51(4), September, 1247-1283. Also Michael Grossman, F. Goldman, S. Nesbitt and P. Mobilia (1993): "Determinants of interest rates on tax-exempt hospital bonds" Journal of Health Economics 12, 385-410

have to share in the losses of a bad year. If anything, in a bad year, this would put greater financial pressure on a PNFP that was relying on the bond market than it would on a PFP that raised its funds from shareholders as well as through the bond markets.

Further, the interest rate a hospital, whether PFP or PNFP must pay on its bonds depends crucially on its credit rating[19]. For bond market purposes, PNFP hospitals are rated by the same credit rating agencies that rate investor-owned operations. A hospital's credit rating will depend on its expected ability to make its bond payments, which in turn will depend on its net margins. Quite consistently on an annual basis through the 1990s, more not-for-profit hospitals had their bonds downgraded than upgraded. Contrary to the impression created by the Devereaux et al paper, then, PNFP hospitals do have external, commercial investors and do face significant pressures from them. Langland-Orban et al. (1996)[20] argue that an increased need to rely on debt financing increases current interest expenses, which increase the carrying cost of new plant and equipment, and may discourage investment in plant and equipment. Even on a smaller scale, PNFPs which finance investment by borrowing from their local bank have to make interest payments, and whether the scale of borrowing is large or small, the people lending to the PNFPs are doing so in the expectation of making a profit comparable to that which they could earn elsewhere. As Getzen (1997)[21] notes, as community PNFP hospitals took on more and more debt and came under increasing financial pressure from borrowers, they had to become more and more business oriented. The notion that PNFP hospitals are somehow immune to investor pressure is nothing but a pleasant fantasy.

---

[19]Grossman et al (1993) estimate the impact of differences in credit rating on the interest rate a hospital must pay on its bonds.

[20]Langland-Orban, Barbara, Louis C. Gapenski and W. Bruce Vogel (1996): "Differences in Characteristics of Hospitals with Sustained High and Sustained Low Profitability" Hospital & Health Services Administration 41(3), Fall, 385-399

[21]Getzen, Thomas E. (1997): Health Economics: Fundamentals and Flow of Funds John Wiley & Sons, New York, pg 172-173

Further, the assumption that having profit as an objective must necessarily lead to cutting corners is not supported by the literature. Langland-Orban et al. (1996)[22] note that in a sample of 140 private Florida hospitals, including both PFPs and PNFPs, drawn in the early 1990s, a higher percentage of the high profit group of hospitals had a high Joint Commission on Accreditation of Healthcare Organizations (JCAHO) accreditation status than did the low-profit group. While they do not break the numbers out in detail by ownership status, they do note that 89% of the sustained high profit group were investor owned. (This does not mean that all investor owned hospitals were highly profitable; 55% of the low profit group were investor owned. Overall, 49% of their sample was investor owned. Not many investor owned hospitals fell in the middle profitability group. This is consistent with Hoerger's (1991) result that PFP hospitals had more profit variability than did PNFPs.)

[22]Langland-Orban, Barbara, Louis C. Gapenski and W. Bruce Vogel (1996): "Differences in Characteristics of Hospitals with Sustained High and Sustained Low Profitability" Hospital & Health Services Administration 41(3), Fall, pg. 397

Chen et al. (1999)[23] look at the 1997 edition of the HCIA 100 Top Hospitals report[24] in an attempt to sort out what makes a top hospital, at least according to that report's criteria. The HCIA report takes account of financial as well as medical performance, since it is at least in part intended as a guide for investors, but according to the HCIA report, the 100 top hospitals tend to have higher profit margins, lower mortality and complication ratings and a higher Medicare case mix index, meaning that they are treating a more complicated case load. The report does not break the numbers out into PFP and PNFP, but since investors can invest in PNFP hospitals through the bond market, it doesn't need to. Taking three years worth of data and counting a hospital as a "top 100" hospital if it had been listed in at least one of the 1994, 95 or 96 reports (so their "top 100" consisted of 224 hospitals) Chen et al generate their own performance indicators for the hospitals in the report, and for peer group hospitals (looking at a total of 4672 hospitals) in particular looking at AMI mortality (as opposed to the overall mortality figure that was used in the consultants' report) and find, after classing hospitals into four categories (small rural, small urban, non-teaching and teaching) that mean risk adjusted mortality in the top 100 hospitals did not differ significantly from their peer group hospitals, nor did the use of "guideline-based therapies in patients without contraindications" (pg. 64). In each of the first three groups of hospitals the PFP hospitals were over-represented in the 100 top hospitals to a statistically significant degree[25].

Other studies have found similar results: Hsia and Ahern (1992)[26] find that "Overall, not skimping on quality produces significantly higher profits despite addition of test costs and allowance for negative tests." (Pg. 24). Cleverley and Harvey (1992)[27] using a small sample of hospitals, and using mortality to judge quality, conclude that poor quality hospitals are less profitable. Annette Tomal (1998)[28] in a study which pools for-profit and not-for-profit hospitals finds that a higher prior year profit

---

[23]Jersey Chen, Martha J. Radford, Yun Wang, Thomas A. Marciniak and Harlan M. Krumholz (1999): "Performance of the '100 Top Hospitals': What Does the Report Card Report" Health Affairs 18(4), July/August, 53-68

[24]HCIA Inc. and William M. Mercer Inc. 100 Top Hospitals: Benchmarks for Success Baltimore and New York, annual

[25]In the Small Rural group, 20% of the "top 100" hospitals were PFP compared with 8% of the peer group population, in the Small Urban group 40% of the "top 100 " were PFP compared with 22% of the peers, in the Non-teaching group, 32.3% of the "top 100" were PFP compared with 12% in the peer group. Even in the Teaching group, where only 2.3% of the peer group population was PFP (13 out of 574), 5.1% (4 out of 78) of the "top 100" were PFP.

[26]David C. Hsia and Cathleen A. Ahern (1992): "Good quality care increases hospital profits under prospective payment" Health Care Financing Review 13(3), 17-24, Spring

[27]William O. Cleverley and Roger K. Harvey (1992): "Is there a link between hospital profit and quality" Health Care Financial Management 46(9), September

margin is associated with a lower current mortality rate.

Note that both the Cleverley and Harvey, and Tomal, studies look at the relationship between profits and mortality, yet neither was mentioned in the Devereaux et al. paper. The explanation is probably the restrictive nature of meta-analysis. Neither Cleverley and Harvey, nor Tomal, separated PFP and PNFP by ownership status, meaning that neither included a binary variable for ownership which could be used to generate a relative mortality risk figure. Thus, even though both studies are informative about the relation between earned profit and mortality, the methodology of meta-analysis demands that they be treated as if they contain no information at all, and therefore omitted. Given this, it seems reasonable to consider what other consequences the very strict filtering process a meta-analysis of the sort Devereaux et al. conducted might have had. Before we do this, though, we need to discuss the workings of meta-analysis.

II. Meta-analysis:

A meta-analysis is a formal approach to combining the results from several different studies of an issue, with the intention of coming up with a pooled estimate of the effect in question. Its most common application probably involves combining estimates of treatment effects. This may, for example, involve comparing the outcome of a new treatment for a particular condition with that of an older treatment for the same condition. The general approach involves combining estimates of treatment effects from a number of smaller studies in order to obtain an estimate of the effect that might have been found had one large study been done in place of a number of smaller ones.

The starting point for a meta-analysis is the hypothesis that there is an actual treatment effect and that the estimated effects found from individual studies are unbiased estimates of the true effect. This involves assuming that the data sets used in the individual studies are samples all drawn from the same population, so that the treatment effects estimated from the individual studies are estimates of the same population parameter. (Resampling from a single, large population is a widely accepted approach to estimating a population parameter, and is the basis of what is known as bootstrap estimation.)

Assume that M is the outcome of interest and T the treatment being investigated. Let T be a binary variable taking on the value one when the new treatment is used and zero otherwise - ie when the old treatment is used, T = 0. In a clinical trial setting, we apply the two treatments to a large number of individuals and then pool the outcomes data together, allowing us to estimate an equation of the form

$$(1) \qquad M_{ji} = a_0 + a_1 T_{ji} + a_X X_{ji} + ?_{ji}$$

---

[28] Anette Tomal (1998): "The relationship between hospital mortality rates and hospital, market and patient characteristics" Applied Economics 30, 717-725

where "i" refers to the individual being treated and "j" to the particular trial. (Within a single trial, "j" is the same for all observations and so the "j" notation would be redundant.) Here, "X" refers to a set of conditioning variables - variables which influence the effect of a treatment on the individual - such as age, sex and health status. (If there is reason to believe that the values of a_0 and a_1 differ across types of individuals - males and females, for example - the data would be analyzed separately for those types and separate estimates of the a coefficients obtained.) The $?_i$ term represents randomness in the outcome of treatment. We shall generally be able to neglect the ? term in what follows.

For an individual who received the older treatment, T = 0 and we have (neglecting individual randomness, ?)

$$(2) \qquad M_{ji} = a_0 + a_X X_{ji}$$

while for an individual receiving the new treatment, T = 1, and we have

$$(3) \qquad M_{ji} = a_0 + a_1 + a_X X_{ji}$$

so a_1 reflects the difference in outcome between an individual who receives the old treatment and an individual who receives the new treatment, conditional on the two individuals having the same values of the "X" variables. Obviously the value of M will vary as the value of X varies; the purpose of the analysis is to estimate the values of a_0, a_1, and a_X, so that we can determine how much of an observed difference in outcome can be attributed to differences in the values of the X variables and how much to differences in T. The estimated value of a_1, usually written $\hat{a}_1$, is an estimate of the differential effect of the new treatment on outcome, holding all other factors constant.

Assume now that a number of researchers have investigated the same treatment, and that each has found an estimate of the true population parameter, a_1 (note that the true parameter can be zero). Letting $\hat{a}_{1j}$ be the estimate of a_1 derived from trial "j", we can bring all of the different estimates of a_1 together into a single data set. In a meta-regression analysis, which is the approach to meta analysis most commonly used in the economic literature[29], we estimate an equation of the form

[29]See T. D. Stanley (2001): "Wheat From Chaff: Meta-analysis As Quantitative Literature Review" Journal of Economic Perspectives 15(3), Summer, 131-150

$$(4) \qquad \hat{a}_{1j} = ß_1 + ß_S S_j + e$$

where the S_j are features of study design which might have affected the estimated value of a_1 and e represents the random error which is always present (but unfortunately unobservable) in coefficient estimates. It is not unusual to include in the S things like the year the data were drawn from (to allow for the possibility that the true value of a_1 changed over time), the size of the study, the type of data used and other information on how the study was undertaken. If, for example, it is suspected that the true value of a_1 actually differs between males and females, one of the S variables could be the proportion of males and females in each sample population. The estimated $ß_S$ coefficients represent the effect of study design on the estimated treatment effect. Ideally, if there are no study design effects, all of the $ß_S$ coefficients will equal zero, and we are left with

$$(5) \qquad \hat{a}_{1j} = ß_1 + e$$

giving $\hat{ß}_1$ as the overall estimate of the treatment effect (ie $\hat{ß}_1$ is the pooled estimate of a_1) .

If all of the studies have the same design, as might well be the case with a set of estimates derived from clinical trials, there will be no differences in the S variables across studies, and a regression (perhaps with the observations weighted by trial size to eliminate the statistical problem of heteroscedasticity) of the

a

$\hat{}_{1j}$ on a constant will yield an estimate of the true population treatment effect parameter a_1. Even in the case of a meta analysis of clinical trials it is a good idea to include locational and time variables in the set of "S" variables, and run a regression equation of the form (4). If, for example, the individual trials being pooled were conducted at widely separated points in time, advances in general treatment technology (improvements in equipment, to take an example which has been suggested might explain differences in estimates of the effectiveness of mammography) could translate into differences in estimated treatment effect. DerSimonian and Laird (1986)[30], in a paper setting out the methodology of the random effects meta-analysis which Devereaux et al. adopt, recommend the use of covariate information where possible. Devereaux et al. do not do this formally.

If it turns out that none of the "S" variables has a statistically significant effect on

a

$\hat{}_1$, we can have more confidence in

ß

$\hat{}_1$ as estimated from (4) as an estimate of the true treatment effect. If any of the S variables do have a significant effect, it raises difficult issues. If differences in study design turn out to be significant, it means

[30]DerSimonian, Rebecca and Nan Laird (1986): "Meta-analysis in Clinical Trials" Controlled Clinical Trials 7, 177-188

that our estimate of the treatment effect is very sensitive to the way we go about trying to determine it, which means at the very least that we should be cautious about claiming too much for our results. If there is a time variable in "S", indicating when the data were collected, and it proves to be significant, it indicates that the treatment effect has been changing over time in a systematic manner, so again we must be cautions about interpreting our results[31]. In the present case, estimation of an equation of the form (4) on relative risks calculated from the articles Devereaux et al. include in their survey, using generalized least squares with patient number as the weighting factor to allow for heteroscedasticity in the estimates, suggests that the estimate of the relative risk is sensitive, to a statistically significant degree, to the number of hospitals used in the estimation in the original articles. This is an interesting indicator of a possible sample design effect, which would probably be worth following up.

While a meta-analysis can, in principle, be a useful way of summarizing the results of a collection of studies, the information requirements of a properly-done meta-analysis are considerable. To begin with, the choice of dependent variable - estimated treatment effect - can exclude perfectly respectable papers simply because their results were not presented in a manner convenient for the meta-analyst. It may also be the case that the effect being meta-analyzed was estimated as a minor part of another study whose focus was a completely different effect. While clinical trials are reasonably standardized, studies of observational data are not, typically controlling for a range of different factors ("X" variables in equation (1) above). This can lead to problems of omitted variable bias in the estimated coefficient, which would pass through to the meta-analysis.

Omitted variable bias is a statistical problem which arises in an equation like (1) above when an X variable which does in fact play a role in determining M is left out of the analysis. Just because the researcher has left an explanatory variable out of his study does not mean that its effect is dropped into the disturbance term, ?. That would be the case only if none of the other, included, explanatory variables were correlated with the omitted variable. If one or more of the included explanatory variables is correlated with the omitted variable, the regression procedure will assign as much of the omitted variable's effect as it can to the correlated included variables, biassing the estimated values of their coefficients. In regression analysis of equations like (1) it is better to include irrelevant variables than to omit relevant ones, since the inclusion of an irrelevant variable will not systematically bias the estimates of the coefficients of the other variables.

---

[31]On these general issues, see Simon G. Thompson and Stephen J. Sharp (1999): "Explaining Heterogeneity in Meta-Analysis: A comparison of Methods" Statistics in Medicine 18, 2693-2708 and Maria Blettner, Willi Sauerbrei, Brigitte Schlehofer, Thomas Scheuchenpflug and Christine Friedenreich (1999): "Traditional reviews, meta-analyses and pooled analyses in epidemiology" International Journal of Epidemiology 28, 1-9

If the estimated coefficients in the original studies do suffer from omitted variable bias, they can not be regarded as drawings from a distribution centred on the true value of the parameter. When the original studies being combined into a meta-analysis used different lists of explanatory variables, a meta regression analysis like equation (4) above should always be performed, including among the "S" variables binary variables indicating whether a key "X" variable was present or absent from the j-th original study.

III. Deriving Relative Mortality Risks for a Meta-Analysis:

Most of the papers in the Devereaux et al meta-analysis approach the question of the determination of mortality risk through hospital-level regression analysis. This means that their dependent variable is hospital-level mortality, either total or for a specific condition or set of conditions. (They use U.S. Medicare data since those data sets were the most detailed large scale data sets available.) Despite reports in the media that the studies included in the Devereaux review analyzed thirty-eight million patient records among them, only two of the studies actually analyzed individual-level data.

The general pattern of the studies Devereaux et al consider is to estimate a regression equation of the form

$$M = a_0 + a_N N + a_F F + a_X X + \varepsilon \quad (6)$$

where M is the hospital's mortality rate, N is a binary variable taking on a value of 1 when the hospital in question is private not-for-profit (PNFP) and zero otherwise, F is a binary variable taking on the value 1 when the hospital is private for-profit (PFP) and X is a vector of other explanatory variables, including patient illness severity measures. The elements in X differ from study to study. When both N and F are equal to zero, we have

$$M_G = a_0 + a_X X + \varepsilon \quad (7)$$

where the "G" subscript on M indicates that this is the mortality equation for public, or government, hospitals (Devereaux et al did not include public hospitals in their calculations of relative mortality risks across hospital types despite the fact that many of the studies they analyzed included information on public hospitals). When the hospital in question is PNFP, N = 1 and F = 0 and we have

$$M_N = a_0 + a_N + a_X X + \varepsilon \quad (8)$$

which is the mortality equation for PNFP hospitals, and when the hospital in question is PFP we have

$$M_F = a_0 + a_F + a_X X + \varepsilon \quad (9)$$

Equations (7) - (9) are derived from estimates of equation (6), which has been estimated at the hospital level, meaning that the number of observations used in the estimation equals the number of hospitals. For all but the two individual-level studies, this means that the number of observations, while large by the standards of economics, is in the thousands, not the millions.

These estimated equations are used to investigate differences in mortality between PFP and PNFP hospitals, as well as to investigate other factors which affect hospital mortality rates - in most cases the effect of ownership was not the primary focus of the studies.

Differences in mortality can be expressed in a number of ways. Relative risk, which was the measure used by Devereaux et al, reports mortality rates in PFP hospitals relative to mortality rates in PNFP hospitals - hence the reports that the Devereaux paper found that mortality was 2% higher in PFP than in PNFP hospitals. In principle, this is calculated as M_F / M_N . An alternative to measuring differences by relative risk is to measure the absolute difference in risk: $M_F - M_N$. While both contain the same basic information, it is generally desirable to report both rather than just one, since they can sound quite different. An M_F of 2 per cent and an $M_N = 1$ per cent yields a one percentage point absolute risk difference but a 100% risk increase in relative terms.

The relative mortality risk associated with a PFP hospital as compared to a PNFP hospital is actually calculated from (8) and (9) as

$$(10)\quad R = [M_G + a_F] / [M_G + a_N]$$

where M_G must be evaluated at particular values of the X variables, typically the mean values. The same value of M_G appears in both numerator and denominator of (10), but it is worth noting that the value of R varies with the value of M_G at which R is being evaluated. In the present case the effect of changing M_G on R will be small, but it is worth noting that, strictly speaking, there is no single, unconditional value of R but that the value of R depends on the point at which it is being evaluated. Since the dependent variable in (6) is a measure of the hospital mortality rate, the two elements in (10) are generally in rates; either percentages or in deaths per thousand. The analysis reported by Devereaux et al looks at whether R is greater than one, in which case M_F is larger than M_N, less than one, meaning M_F is smaller than M_N, or equal to 1, in which case the two mortality rates are equal. Since the calculations are being done using coefficients estimated from data samples, there is a degree of uncertainty attached to each of the estimated as, which passes through to any variable generated using the estimated a values: hence the need to calculate confidence intervals around R. Standard practice, which Devereaux et al follow, is to calculate a 95% confidence interval around the value of R calculated from the estimated coefficients, and to say that R is significantly different from 1 if 1 does not fall within that confidence interval.

Devereaux et al calculate the confidence intervals around R for each of the studies which they included in their review by using the estimated mortality rates to generate a 2X2 table of the form

(11)

	PFP	PNFP	
D	D_F	D_N	T_D
A	A_F	A_N	T_A
	T_F	T_N	Total

In (11), Total is the total number of patients treated in the hospitals in the study being considered, T_F is the number treated in PFP hospitals and T_N the number treated in PNFP hospitals. In half of the studies included in the Devereaux et al review this information was not available and had to be approximated on the basis of mean numbers of Medicare admissions to general acute care hospitals in 1989[32]. D_F and D_N are the number (not the rate) of patients in each type of hospital who are estimated to have died in the time period of the study - while a few of the studies used in-hospital mortality, most used thirty day mortality rates. Similarly, A_F and A_N are the number of people treated in each type of hospital who are estimated to have been alive at the end of the study horizon. We say "estimated' because Devereaux et al. very seldom have actual counts. Instead, they estimate these numbers by starting from the total number of patients, distributing them between PFP and PNFP hospitals, then applying the estimated mortality rates, M_F and M_N to those distributed numbers to estimate the numbers in each cell of (11). As we noted above, in the case of half of the studies which they consider in their review they had to use mean Medicare admission numbers to generate (11).

Devereaux et al. then use the numbers in (11) to calculate confidence intervals around the relative risk figures reported in the articles they are reviewing. It should be noted that the confidence interval found is based on an estimate of the standard error of the relative risk. They then test whether the relative risk differs between the two types of hospital by looking at whether the 95% confidence interval generated from (11) contains 1; if it does there is no statistically significant difference between mortality rates in the two types of hospitals.

There are a couple of points worth making about this procedure. First, note that the elements in (11) are being calculated using the coefficients estimated from (6). We have already seen that the relative mortality risk can be found from (11), which is a nonlinear combination of coefficients from (6). The hypothesis that there is no difference between mortality rates in PFP and PNFP hospitals can be written as

(12) $[M_G + a_F] / [M_G + a_N] = 1$

[32]See Appendix 1 to Devereaux et al. This Appendix was posted on-line but was not included in the paper journal version of the article.

which can be tested in a regression framework as a Wald test. More importantly, (12) is equivalent to

$$(13) \qquad [M_G + a_F] = [M_G + a_N]$$

which, since M_G is the same on both sides of (13), is equivalent to

$$(14) \qquad a_F = a_N$$

In other words, the hypothesis that the mortality rate is the same in both types of hospital can be tested as a linear hypothesis using the estimated a_F and a_N coefficients. The question obviously arises as to whether this hypothesis could have been tested directly from the information contained in the original articles, without having to generate (11) above. The answer is that, in several cases, it could, at least to a reasonable approximation. Ideally we would want to have the full Variance-Covariance matrix of the estimated coefficients to test a hypothesis involving the estimated values of two of those coefficients, and we have at most the standard errors of the estimated coefficients, which permit us to recover the diagonal elements of the Variance-Covariance matrix but not its off-diagonal elements. In practice, though, it is very unlikely that the off-diagonal elements would dominate the diagonal elements in a hypothesis test, so we can safely draw inferences based on the diagonal elements. As an approximation to the true test it is certainly no worse than the one derived from (11). Since Devereaux et al. used the procedure based on (11) to generate the confidence intervals even when data were available to permit them to test (14) directly, the next question must be whether the two approaches give the same result. The answer is that, in the case of the Devereaux review, they sometimes do not. We shall return to this point in our discussion of the Pitterle et al. and Bond et al. papers below.

IV. Selection of Results for Inclusion:

One drawback to a strict meta-analytic approach, as distinct from the approach of a traditional literature review or a mixture of the two, is that nuances in the articles being reviewed can easily be missed. Devereaux et al., for example, seem to conclude that for-profits achieve cost savings by cutting corners on patient care. In one of the articles included in their systematic review, Sloan et al. (2001)[33] conclude that (pg. 19) "the hypothesis that for-profits engage in cost cutting at the expense of quality does not receive support.". In a traditional literature review, the disagreement between the conclusion reached by Sloan et al. and that reached by the authors of the review would at least be highlighted, and

[33]Frank A. Sloan, Gabriel A. Picone, Donald H. Taylor Jr., Shin-Yi Chou (2001): "Hospital ownership and cost and quality of care: is there a dime's worth of difference?" Journal of Health Economics 20, 1-21

probably discussed. Similarly McClellan and Staiger (2001)[34], in the article which is included in the Devereaux review, in their discussion of their results qualify their finding that (pg. 111) "the performance of not-for-profit hospitals in treating elderly patients with heart disease appears to be slightly better than that of for-profit hospitals" with the observation that (pg. 110) "Many not-for-profit hospitals are below average, many for profit hospitals are above average, and these relationships vary enormously at the market level." And (pg. 111) "this small average difference masks an enormous amount of variation in hospital quality within the for-profit and not-for-profit hospital groups." They go on to suggest that the entry of for-profits into markets where quality of care had generally been poor might provide the impetus for quality improvements in those markets. Their argument rests on the fact that for-profit hospitals will tend to enter markets where they perceive profit opportunity, and those markets will not necessarily be ones in which quality of care had previously been extremely good. Getzen (1997)[35] notes that (pg. 173) through the 1980s, financial market pressure on PNFP hospitals, resulting from their having taken on considerable debt, made it harder and more costly for them to maintain and upgrade their capital equipment. Many of these facilities were acquired by PFP chains, which had access to the funds necessary to refurbish the plant[36]. Whether one agrees with McClellan and Staiger's suggestions, in a traditional literature review it would be necessary to discuss them; this often does not happen in meta-analysis.

The McClellan and Staiger article is also interesting as an illustration of the selection process

---

[34]Mark McClellan and Douglas Staiger (2001): "Comparing Hospital Quality at For-profit and Not-For-profit Hospitals" in David M. Cutler, (ed.) The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions National Bureau of Economic Research Conference Report, University of Chicago Press, Chicago

[35]Thomas E. Getzen (1997): Health Economics: Fundamentals and Flow of Funds John Wiley & Sons, New York

[36]Tami Mark (op. cit.) found that low profit margins were a good predictor of conversions both from not-for-profit to for-profit and from for-profit to not-for-profit, and that hospital profits margins increased both after PNFP to PFP conversion and after PFP to PNFP conversion.

necessary in a meta-analysis. Devereaux et al calculate one relative risk ratio from this study for inclusion in their analysis, but McClellan and Staiger report nine, based on three different years data and three different estimation approaches. Devereaux et al. have, not unreasonably, chosen to report a relative risk based on the most recent data subset reported by McClellan and Staiger, but it is worth noticing that, while McClellan and Staiger's results do favour PNFPs, this is also the subset which yields the largest gap (1.15 percentage points).

Normal practice in a meta-analysis, when an author reports the results of several different estimation approaches, is to use the one which the author argues is preferred. This approach works reasonably well when the effect being meta-analysed is the focus of the original investigations. When, however, it is estimated as part of the process of investigating another effect, the rule of thumb may not apply. Suppose, for example, that we are considering two sets of hospitals, both of which have the same total level of inputs, but that one class of hospital devotes more of its resources to quantity and less to quality than does the other. This is essentially what Devereaux et al. seem to believe would distinguish PFP from PNFP hospitals. If we regress a quality measure, like mortality, on total inputs (both fixed and variable) and a quantity measure of output (eg total number of discharges) and add a binary variable for ownership type, then, if the whole of the difference between the two types is in how they use the same total number of resources, inclusion of the quantity variable controls for that effect, and the ownership variable is likely to be nonsignificant. In this case we would prefer to regress mortality on an ownership variable and either the total levels of the inputs but not the discharge quantity, or on the ownership variable and the discharge quantity and not a full set of variable inputs (although including bed size as a scaling variable would probably not cause problems).

Looking at the three sets of estimation results which McClellan and Staiger report for each of 1985, 1991 and 1994, the first difference among them is in the way mortality was adjusted for patient illness severity : McClellan and Staiger's chief interest is in investigating an alternative approach to generating hospital-level risk adjusted mortality for AMI patients. For each year, they create two hospital-level risk adjusted mortality series - one using what they term the actual risk adjusted mortality, the other their preferred filtered risk adjusted mortality. They run three regressions using these rates as dependent variables. Each regression contains variables for ownership status, including government ownership, and a variable indicating teaching status. In addition, the equations using the standard risk adjusted mortality rate include hospital volume as an explanatory variable as do one of each pair estimated using the filtered rates; the second filtered rate equation for each year omits volume, since it is potentially endogenous. Assuming we use an equation with McClellan and Staiger's filtered risk adjusted mortality as the dependent variable for the basis of comparison there is still the question of whether it should include the hospital volume measure. If we take the view, as some econometricians do, that omitted variable bias is a potentially more serious problem than endogeneity bias, we should use the equation which includes volume. For 1994, the year whose estimated effect was chosen for inclusion in the Devereaux analysis, the difference in mortality between PFP and PNFP still favours PNFP hospitals, and is statistically significant, but falls from 1.15 percentage points to 0.87 percentage points. McClellan and Staiger also ran equations including county level fixed effects variables, to control for regional differences which would affect PFP and PNFP hospitals alike (these would include things

like degree of rurality and factor prices)[37]. Unfortunately they do not report the estimated coefficients from the equations including the fixed effect terms, but simply state that adding county-level fixed effects cuts the difference between PFP and PNFP mortality rates roughly in half. If this applies to the coefficients from the equations controlling for volume, we are now down to a difference of roughly 0.44 percentage points. We presume the difference is still statistically significant, on the assumption that they would have mentioned the fact if it were not. Since an equation with risk adjusted mortality as dependent variable, including among its explanatory variables controls for ownership, volume and regional effects but not including the levels of the hospital's variable inputs, would be a reasonable form to use in testing for ownership effects, we can see that the effect is very sensitive to the specification of the original equation.

V. Choosing Among Alternative Sets of Results cont'd:

The papers by Pitterle et al (1994) and Bond et al (1999) provide interesting illustrations of the judgement calls which have to be made with regard to the inclusion of articles and results in meta-analytic studies. Each article considers determinants of overall Medicare mortality, Pitterle et al analyzing HCFA data tapes for 1988 and Bond et al for 1992. Devereaux et al (2002), in their Figure 2, indicate[38] that both articles favour private not-for-profit (PNFP) hospitals over private for-profit hospitals (PFP), the Bond article slightly more so than the Pitterle. In both cases, though, the confidence interval around the estimated relative risk which is drawn in Figure 2 in Devereaux at al

[37]They do this to control for the effect of location on hospital outcome, since PFP and PNFP hospitals are not identically distributed geographically, and since there is evidence that factors like rurality do affect outcomes. See, for example, the discussion in another of the included papers - Emmet B. Keeler, Lisa V. Rubenstein, Katherine L. Kahn, David Draper, Ellen R. Harrison, Michael J. McGinty, William H. Rogers and Robert H. Brook (1992): "Hospital Characteristics and Quality of Care" Journal of the American Medical Association 268(13), October 7, 1709-1717.

[38]This is on the assumption that the order of the observations on the graph in Figure 2 is the same as the order of the articles in the list in the same figure.

does not include 1, meaning that in both cases mortality risk is significantly higher in PFP than in PNFP hospitals.

Each of these articles report results of more than one specification of equations which include hospital ownership as an explanatory variable. In each case[39] they report multivariate and what they term univariate regression models with mortality measured as deaths per thousand admissions as the dependent variable. The univariate regression models (referred to in Bond at al (1999) as simple regressions) are not, strictly speaking, textbook simple regressions. They were actually conducted in two steps - first, equations were estimated with severity of illness indicators as explanatory variables. Then the other indicators, including the ownership variables, were entered individually into equations which already contained the severity of illness indicators as explanatory variables. The coefficients on the other indicators, then, were from adding those variables in as explanatory variables in equations which already controlled for severity indicators. According to Pitterle et al (pg. 623) "Thus, all the other univariate results were adjusted for the presence of severity of illness indicators, creating a more accurate analysis of individual measures of association with mortality rates." Basically, then, the univariate regressions in both of these papers were conducted using severity-adjusted mortality rates.

Devereaux et al, in their discussion of their selection criteria, indicate that, when a paper reported several adjusted analyses, they included the results from the analysis with what they regarded as the most appropriate adjustment. They considered it appropriate to adjust for patient characteristics, including severity of illness, but preferred, if possible, to avoid the use of results from equations where the explanatory variables included factors such as number of registered nurses per bed which were under the control of hospital administrators. In economic terms, in the short run factors of production fall into two broad categories - fixed and variable. Fixed factors are ones whose levels cannot be varied in the short run while variable factors can be adjusted. Broadly speaking, labour inputs tend to be more variable, and capital inputs are more likely to be fixed factors in the short run (in the long run, all factors are variable).

The reason for Devereaux et al's judgement as to what features of the hospital it is appropriate to adjust for is not set out, but presumably runs something like this: The production function for health outcomes for individual conditions is presumably the same in all types of institutions, in the sense that equal application of equivalent inputs should produce (on average) equal outputs. This means that if an institution were to convert from one status to the other and continue to use the same mix of inputs as before, its outcomes should remain unchanged. Differences in outcomes between institutions of different ownership types, then, are likely to be due to differences in input use - in economic terms to differences in (by definition) the use of variable inputs. If we estimate a production function for mortality with both fixed and variable hospital inputs as explanatory variables (including in the equation, of course, controls for patient illness severity), much if not all of the difference in mortality between PFP and PNFP institutions would be accounted for by differences in variable input use. This would not be a problem if the intent was to estimate the production function, since the greater the variation in input and output

[39]The two articles were actually two separate investigations on different data sets by the same team of researchers.

levels in the data the more precisely the function can be estimated. It would not, however, be directly informative about systematic differences in outcome between ownership types. If we add an ownership dummy to a fully-specified production function, there will be nothing left for it to pick up, even if PFPs do systematically use fewer inputs than do PNFPs.

As we noted above, Figure 2 in Devereaux et al. seems to indicate that the Pitterle paper favours not-for-profit hospitals, in the sense that the point estimate of relative risk is greater than one and the 95% confidence interval does not include 1. This indicates that the mortality risk associated with being in a for-profit hospital is significantly greater than that associated with being in a not-for-profit hospital.

Pitterle et al present three tables of regression results, found from estimation of equations like (6) above. Their first table reports estimates from what they term univariate regression (although as we noted, the severity of illness variables are included in each equation, so the coefficient on each of the other variables is found by estimating an equation in which it is the only explanatory variable other than the severity of illness variables which is included in the equation) and a multivariate regression, which includes a number of explanatory variables including variables for the ratio of board-certified physicians to all physicians, and the ratio of registered nurses to all nurses. These latter variables would seem to fall in the category of "under the control of hospital administrators", which Devereaux et al indicated they would prefer not to have included as adjustment factors. In the univariate regression results, the estimated a_F coefficient is -0.008, with an estimated standard error of 0.0007, and the estimated a_N coefficient is -0.003, with an estimated standard error of 0.0010. Since the Pitterle equation is of the form of equation (6) above, there coefficients indicate that, at any common values of the other explanatory variables, both PFP and PNFP hospitals have significantly lower mortality rates than do public hospitals. More importantly, the PFP mortality rate is less than the PNFP rate, and the difference between them, is, on the basis of the estimated standard errors reported in the paper, statistically significant at the 5% level. On the basis of the univariate results reported in Pitterle's table 2, then, it appears that mortality in for-profit hospitals is significantly less than mortality in not-for-profit hospitals.

As we noted, table 2 also reports coefficients from a multivariate regression equation, and the introduction of the additional variables does change the estimates of the ownership coefficients to change quite dramatically. The coefficient on for-profit ownership is now -0.0039 with an estimated standard error of 0.0007, which means that the coefficient estimate is significantly less than zero, while the coefficient on not-for-profit ownership is -0.0033, with estimated standard error 0.001, so that coefficient, too, os significantly different from zero. When we use the reported estimated standard errors to test whether the two coefficients are significantly different from each other, we get a t-statistic of 0.49, meaning that the difference between them (which favours for-profits still) is not statistically significant. On the basis of Pitterle's Table 2, then, either for profits have a mortality rate which is significantly less than that of not-for-profits, or there is no significant difference between them.

Pitterle's Table 3 reports the results of a similar exercise, but with a slightly different set of severity-of-illness variables. The equations reported in Table 2 included a hospital-level predicted mortality variable among the severity indicators while those reported in Table 3 omit it. In Table 3, in the univariate equation, the estimated a_F coefficient is -0.008, with standard error 0.0010, making it

significantly different from zero, while the estimated value of a_N is -0.005 with standard error 0.001. The difference between the two coefficients is still significant at the 10% level.

It is in the multivariate equation in table 3, which includes data on the mix of medical personnel, that we have a_N being larger in absolute value than a_F : a_N is -0.0051 with standard error 0.0013 and a_F is -0.0039 with standard error 0.001. The difference between them is, on the basis of the reported standard errors, nowhere near statistical significance. Thus even though this set of coefficients seems to favour not-for-profit hospitals, the absolute difference in mortality risk is not statistically significant.

Pitterle's Table 5 reports an additional set of regression results, but these are from a smaller data set - a subset of the data used in Tables 2 and 3 - and even here the univariate equation favours for profits, although the difference is not statistically significant, and none of the ownership variable survive the selection process for the multivariate equation (which includes additional variables which might be regarded as under the control of hospital administrators), meaning that none of them had any significant effect on mortality. All told, then, none of the results reported in the Pitterle paper seem consistent with its being classed as favouring not-for-profit hospitals to a statistically significant degree[40].

Turning to the Bond et al. paper, which Devereaux et al. show as favouring private not-for-profit hospitals to a statistically significant degree, the authors of the original paper again report univariate results in which severity of illness has been controlled for and other variables are then added individually, and then report a multivariate regression equation. This time, in the univariate equations, a_N is -0.0035 with standard error 0.0011 and a_F is -0.001 with standard error 0.0014, so the differential mortality rates do favour the not-for-profit hospitals. A t-test of the difference between the a_N and a_F coefficients, however, shows the difference between them as non-significant. In the multivariate equation a_N is -0.0049 with standard error 0.0011 and a_F is -0.0043 with standard error 0.0013, a difference which comes nowhere near significance.

Clearly there is a question as to why different approaches to testing the same hypothesis seem to lead to such drastically different conclusions, as in the case of the Bond et al. paper. While testing a hypothesis in non-linear form will generally lead to different results than testing it in linear form, the difference here appears to arise from a different source. Both the Pitterle et al. and the Bond et al. papers were among those for which Devereaux et al. indicate that they were unable to obtain the numbers for the cells in a 2X2 table like (11) above, and for which values had to be based on other figures. While it is not unreasonable to apply estimated relative risks to populations other than those from which they were estimated, to see the implications of the point estimates for alternative populations, caution has to be employed in using numbers other than those which were used to find the point estimates of coefficients to yield standard errors for those coefficients or for combinations of those

[40]Preliminary results suggest that the choice of the values from the Pitterle et al. paper which are used in the meta-analysis may well affect whether the final, pooled estimate of relative risk is significantly different from one.

coefficients. The standard error of an estimated coefficient (and the variance-covariance matrix of a set of coefficients from a multiple regression) reflect the uncertainty inherent in the estimated values, given the information contained in the data set which was used to estimate them. Any statistical testing should ideally be done on the basis of that uncertainty.

Using a different data set to generate confidence intervals means generating an artificial data set on the basis of estimates derived from the original data, and performing statistical inference on the basis of that artificial data set. Since the artificial data set (for use in the 2X2 table) was generated with the point estimates from the first data set, there is actually no uncertainty in it. For that artificial data set, the estimates are precisely equal to the population values. Allowing for the fact that we are dealing with approximate formulae for variances, we would still expect the confidence intervals around the relative rates in the 2X2 table to be very tight. While there are cases in which the authors of the original papers provide no information on standard errors - as when the report only p-values, for example - whenever the information is present in an article which is being reviewed it would be preferable to use it as the basis for tests of hypotheses about coefficient values estimated in that article. If we do that in the present case, both the Bond et al. and Pitterle et al. articles move from the "favours not-for-profit" category to the neutral category, if not the "favours for-profit" category.

VI. Selection of Literature to Exclude:

The selection criteria for a meta-analysis also include decisions about what articles to exclude. Exclusion of an article because it does not meet all of the criteria for inclusion is equivalent to saying that it contains no information relevant to the question at hand. This might be an acceptable approach when the literature is large and the methodology of the original papers is reasonably well standardized, but it is a much riskier procedure when the literature is small and a wide variety of specifications has been used in the original articles[41]. Devereaux et al decided, for example, to exclude an article by Taylor, Whellan and Sloan (1999)[42] because its data set was a subset of the data used in the included paper by Sloan et al. This should not, however, be taken to mean that the 1999 paper contained no additional information. In the 2001 article, Sloan et al. estimated a single equation for probability of death, with primary diagnosis as an explanatory variable. In the 1999 paper they estimated relative mortality risks

[41]Ideally we would want to include only articles which applied the standard specification tests to their estimated equations - testing for heteroscedasticity and normality of the residuals and for the functional form of the estimated equations. (And whose equations passed the tests, of course.) If we applied that rule to this literature we would be left with very few articles.

[42]Donald H. Taylor, David J. Whellan and Frank A. Sloan (1999): "Effects of Admission to a TeachingHhospital on the Cost and Quality of Care for Medicare Beneficiaries" New England Journal of Medicine 340(4), January 28, 293-299

for different hospital ownership classes separately for hip fracture, stroke, coronary heart disease and congestive heart failure. This is of interest since the correlation between mortality rates across different conditions within individual hospitals tends to be very low. Mortality for different conditions have, in economic terms, different production functions. In addition, all hospitals face budget constraints so that as part of the process of deciding what resources to acquire even general hospitals may choose to specialize (or find that they have accidentally wound up specializing) in a limited number of areas, so it is of some interest to see if there is evidence that the failure in their 2001 paper to find a significant effect of ownership on pooled mortality was due to the fact of pooling.

In the 1999 paper, testing mortality in PNFP relative to PFP hospitals, Sloan et al. find no significant difference in mortality risk for any of the individual conditions. For two of the conditions, the relative risk is less than one, favouring PNFP hospitals and for two it is greater than 1, favouring PFP hospitals. The smallest p-value is .08, (in the hip fracture equation, in favour of PNFP hospitals) the other three range from 0.33 to 0.58. In other words the failure to find a significant effect of ownership is consistent across conditions.

Among the other papers Devereaux et al. considered and excluded is one by Ettner and Hermann (2001)[43]. Because their interest was in treatment of patients with psychiatric diagnoses, Ettner and Hermann exclude from their study patients who died in hospital or within 30 days after discharge, noting only that mortality rates were similar for for-profits and (private) not-for-profits. Obviously a study which does not consider mortality as an outcome is not a candidate for inclusion in a meta-analysis of studies of mortality rates, but a more traditional literature review of evidence on hospital quality might note that Ettner and Hermann did include a quality proxy, the 30 day rehospitalization rate, and found no significant difference by ownership type. Nor did they find a significant difference in cost, nor any evidence that PFPs selected easier-to-treat patients.

A paper by Burns and Wholey (1991)[44] did consider mortality rates and was excluded perhaps because its results could not be fit into the meta-analytic framework. Burns and Wholey report logistic regressions of mortality equations for each of 16 procedures or conditions. They include, as explanatory variables, besides ownership, age, sex, comorbidities, teaching status, hospital patient volume, rural/urban and a set of physician characteristics. It is a paper for which there seems to be a

[43]S. L. Ettner and R. C. Hermann (2001): "The role of profit status under imperfect information: evidence from the treatment patterns of elderly medicare beneficiaries hospitalized for psychiatric diagnoses" Journal of Health Economics 20(1), 23-50

[44]Lawton R. Burns and Douglas R. Wholey (1991): "The Effect of Patient, Hospital and Physician Characteristics on Length of Stay and Mortality" Medical Care 29(3), 251-271

strong case for inclusion in a systematic review of mortality differences, but it is not included probably because Burns and Wholey follow the common practice of only reporting, in their estimation results, the values of the statistically significant coefficients. For all but one of the estimated equations the coefficient identifying a hospital as investor-owned as opposed to PNFP is nonsignificant - ie in 15 out of the 16 equations the difference between mortality in PFP and PNFP hospitals is not statistically significantly different from zero - and therefore is not reported. In one case - hip fracture - the coefficient on "investor owned" is negative, meaning that PFP hospitals have lower mortality than PNFP, and is statistically significant. Even in that case, however, their result is not mentioned in the Devereaux paper.

Beyond the fact that it estimated logistic equations on micro-level data, so that in the one case in which a statistically significant effect was found the coefficient is a log odds ratio, the Burns and Wholey paper would have a strong claim for inclusion in a traditional literature review because it includes a variable virtually no other study does. Because their data came from a single state, Burns and Wholey were able to construct a physician volume index, using the volume of patients with each condition seen by a physician in the sample year in each hospital in which that physician practised. Most studies are not even able to include a variable for the number of physicians practising in each hospital. In discussing the inclusion of this variable, Burns and Wholey note evidence that the volume of cases handled by a physician may be more important than the volume handled by a hospital for some conditions.

Physician volume is significant in eight out of the 16 equations, more than any of the hospital characteristics. In seven out of the eight higher physician volume has a negative effect on the patient's probability of death (in the case of dehydration it is significant and positive, which is not easy to explain). There is clearly good reason to think that if there is a practice effect, with greater experience translating into greater skill, it is probably stronger at the level of the physician than the level of the hospital. The fact that Burns and Wholey found a negative and significant effect for so many conditions is definitely of interest, if only because it suggests that other studies which exclude a physician volume measure suffer from omitted variable bias. And even though Burns and Wholey do not report the data necessary to calculate relative risk ratios, the fact that in fifteen out of sixteen equations there is no significant difference, and in the one in which the effect is significant the effect is negative, would merit mention in a traditional literature review even though those findings cannot be quantified in the precise manner called for by a random effects meta-analysis.

There are other articles in the literature which, while they do not meet all of the precise requirements for inclusion in Devereaux et al's meta-analysis, are none the less informative. For example, Geweke, Gowrisankaran and Town (2001)[45], in a study of mortality among Medicare pneumonia patients admitted to 114 hospitals in Los Angeles County in the 1989-92 period, and using a Bayesian approach to control for selectivity bias in hospital selection, find PFP and PNFP hospitals to

[45]John Geweke, Gautam Gowrisankaran and Robert J. Town (2001): Bayesian Inference for Hospital Quality in a Selection Model National Bureau of Economic Research Working Paper 8497, October

be similar in quality with a slight edge for PFP: moving a patient from a PNFP to a PFP hospital results in a 2% decrease in mortality risk. They also find that mortality risk tends to be lower in larger hospitals.

Gowrisankaran and Town (1999)[46], using a larger sample of Medicare pneumonia patients, investigate the sensitivity of mortality risk estimates to approaches used to control for selectivity bias. They compare results of two estimation procedures - one Instrumental Variables (IV) approach which controls for endogeneity of hospital choice, and a Generalized Least Squares (GLS) approach which does not. Under GLS estimation there is no significant difference between PFP and PNFP (in the GLS equation which allows mortality to depend on bed size, average length of stay, and a number of other factors as well as ownership they find that PNFP hospitals have higher mortality risk but the t-statistic is 1.50, which is not significant at conventional levels). In the IV estimation, the coefficient on PNFP (which compares mortality at PNFP with PFP hospitals) is negative and close to significant, but in the equation which allows mortality to depend on bed size, they find that the quality of PFP hospitals increases with bed size (ie the mortality rate in PFP hospitals decreases as bed size increases) while the quality of PNFP hospitals does not. An increase of 10 beds reduces the daily mortality rate in PFP hospitals by 0.7 percentage points. While the negative coefficient on PNFP ownership indicates that mortality is 1.3 percentage points lower in PNFP than PFP hospitals, when bed size is factored in, mortality is lower in PFP than PNFP hospitals for any hospital with more than about 20 beds.

McGinty (1993)[47] looks at the effect of ownership on excess (ie severity adjusted) mortality for several conditions combined in 1985-6 and finds that government hospitals have higher than expected mortality while the difference between PFP and PNFP is not statistically significant.

Kessler and McClellan (2001)[48] in a paper which, like a number of other recent papers, investigates ways of controlling for selectivity bias resulting from the endogeneity of hospital choice, analyze longitudinal data on all nonrural elderly Medicare patients hospitalized for a new heart attack in the period 1985-96. Their results would be difficult to fit into a meta-analysis, since they calculate the ratio of PFP to PNFP hospitals in an area and report the effect on mortality of different densities of PFP/PNFP hospitals. They find no significant effect of living in an area with above median density of PFP relative to PNFP hospitals, and no significant effect on mortality of being in an area which is above the 75th percentile of PFP to PNFP density, although there is a significant positive effect on mortality of living in an area with above the 75th percentile of density of public relative to PNFP hospitals.

In other articles germane to the issue of hospital quality but moving farther from the use of

---

[46]Gautam Gowrisankaran and Robert J. Town(1999): "Estimating the quality of care in hospitals using instrumental variables" Journal of Health Economics 18, 747-767

[47]Michael J. McGinty (1993): Medicare Reimbursement and the Quality of Hospital Care RAND Graduate School Doctoral Dissertation, RAND Corporation paper N-3409-RGSD, Santa Monica CA.

[48]Daniel Kessler and Mark McClellan (2001): The Effects of Hospital Ownership on Medical Productivity National Bureau of Economic Research Working Paper 8537, October

mortality as the index of quality, Brook et al. (1990a)[49] combine incidence of death, heart attack and stroke into an index of adverse events for patients undergoing carotid endarterectomy[50] in 1981. The inclusion of stroke and MI as outcomes means that the Brook study does not fit neatly into a meta-analysis of studies focussing only on mortality, but the finding that there is no

[49]Robert H. Brook, Rolla Edward Park, Mark R. Chassin, Jacqueline Kosecoff, Joan Keesey and David H. Solomon (1990a): "Carotid Endarterectomy for elderly patients: Predicting Complications" Annals of Internal Medicine 113, 747-753

[50]Some authors prefer combined indices of adverse events simply because mortality rates for so many conditions are so low.

significant difference between PFP and PNFP hospitals in the incidence of any of the three adverse events seems worth mentioning. Also using 1981 data, Brook et al (1990b)[51] find no significant difference between PFP and PNFP hospitals in appropriateness of three procedures. Also in the adverse events literature, Brennan et al (1991)[52], using 1984 data from New York state, using record review to identify adverse events, found proprietary hospitals to have significantly fewer adverse events which were judged as probably being due to negligence.

VII. Caveats on Meta-Analyses:

A meta-analysis can be a useful device for summarizing and to a degree extending the results of a number of studies. So long as the study design is appropriate and does not differ greatly across the studies being reviewed, a meta regression equation can add information beyond that present in any single one of the studies. As we have noted, the estimated treatment effect from a single study can, assuming the estimator is unbiased, be regarded as a drawing from a distribution centred on the true treatment effect. Suppose that we collect together fifty of such studies and that in each of them the treatment effect, although not statistically significantly different from zero, is positive. Assuming the data sets used in the studies to be statistically independent of each other, the likelihood that each of the estimates would being positive when the true effect was not is extremely small. A meta regression analysis of those studies can give an estimate of the true treatment effect, and it is legitimate to pool the information in the individual studies to get an idea of whether that treatment effect is statistically significantly different from zero. We can also get an idea of whether the effect is likely to be of practical, as well as statistical, significance. With a large enough sample, virtually every estimated coefficient will be statistically significant - when the sample is extremely large the non-significant coefficients are, in some ways, more interesting than the significant ones. The fact that a coefficient from an extremely large data set is statistically significant does not mean that it matters for policy or treatment purposes.

The fact that a meta-analysis can provide useful information does not, however, mean that every meta-analysis does provide useful information. The information requirements of a good meta-analysis are stringent, and the smaller the number of studies being included in the analysis the greater the chance that bias present in the estimate derived in one study will contaminate the meta-analysis. When the

[51]Robert H. Brook, Rolla Edward Park, Mark R. Chassin, David H. Solomon , Joan Keesey and Jaqueline Kosecoff (1990b): "Predicting the Appropriate Use of Carotid Endarterectomy, Upper Gastrointestinal Endoscopy and Coronary Angiography" New England Journal of Medicine 323, October 25, 1173-1177

[52]Troyen A. Brennan, Liesi E. Hebert, Nan M. Laird, Ann Lawthers, Kenneth E. Thorpe, Lucian L. Leape, A. Russell Localio, Stuart R. Lipsitz, Joseph P. Newhouse, Paul Weller and Howard H. Hiatt (1991): " Characteristics Associated with Adverse Events and Substandard Care" Journal of the American Medical Association June 26, 265(24), 3265-3269

literature is small it is probably better at the very least to accompany the meta-analysis with a detailed literature review, which includes a discussion of the articles which were not included in the final statistical pooling.

It should also be noted that there is disagreement in the literature about the track record of meta-analysis as it has been practised (as distinct from its theoretical potential). LeLorier et al (1997)[53] compared the results of twelve large, randomized controlled trials with 19 earlier-published meta-analyses dealing with the same topics. They concluded that thirty-five per cent of the time the meta-analyses did not predict accurately the results of the large trials. Naylor (1997)[54] argued that LeLorier et al. overlooked the results of certain important studies (which, since the LeLorier et al. article was itself essentially a meta-analysis, serves to emphasize the potential sensitivity of the statistical analysis to the selection of articles being included) and that another review of meta-analysis suggested that "Directionally, 80% of meta-analyses agreed with the results from the larger trial, although concordance for statistically significant findings was much less." If meta-analyses only get the direction of effect right 80% of the time, we should clearly be cautious in deciding how much weight to give their conclusions. Bailar (1997)[55] argues on the basis of his own review of meta-analyses that

> "..... problems were so frequent and so serious, including bias on the part of the meta-analyst, that it was difficult to trust the overall "best estimates" that the method often produces. On present evidence, we can generally accept the results of a well-done meta-analysis as a way to present the results of disparate studies on a common scale ..... But any attempt to reduce the results to a single value, with confidence bounds, is likely to lead to conclusions that are wrong, perhaps seriously so."

---

[53]Jacques LeLorier, Geneviève Grégoire, Abdeltif Benhaddad, Julie Lapierre and François Derderian (1997): "Discrepancies Between Meta-Analyses and Subsequent Large Randomized, Controlled Trials" New England Journal of Medicine 337(8), August 21, 1997, pp. 536-542

[54]C. David Naylor (1997): "Meta-analysis and the meta-epidemiology of clinical research" British Medical Journal 315, September 13, 617-619

[55]John C. Bailar (1997): "The Promise and Problems of Meta-Analysis" New England Journal of Medicine 337(8) August 21, 1997, pp. 559-561

Problems with meta-analyses are even more pronounced in behavioural research, where agents cannot be randomized into groups and behavioural responses must be controlled for statistically rather than in study design. The omitted variable bias introduced by a failure to control adequately can distort estimates of treatment effects to a serious degree. To take one example relevant to the current case, there is considerable evidence in the literature that for at least some conditions, the outcome of treatment depends on the volume of cases that a hospital treats[56] and that for some conditions this carries through to hospital size, with larger hospitals having better outcomes than smaller ones. Since PFP and PNFP hospitals are not identically distributed across bed size, with PFP generally being smaller, if the hospital size effect is genuine, and no scale variable is included in the estimated mortality equation, the effect of hospital scale will tend to be picked up by the ownership variable.

This type of problem led Blettner et al. (1999, op. cit.), in a discussion of the potential of meta-analysis as a tool of epidemiology, to conclude that "Meta-analyses from published data are in general insufficient to calculate a pooled estimate since published estimates are based on heterogeneous populations, different study designs and mainly different statistical models."

This does not mean that we cannot draw conclusions about treatment effects from various types of literature reviews. We do, however, have to give much more attention to the design of the original studies than is necessary in a meta-analysis of clinical trials.

VIII. Structuring an Investigation of the Effect of Hospital Ownership on Mortality:

How, then, would we go about investigating the effect of hospital ownership on patient outcome in the American context? The basic hypothesis to be tested is fairly straightforward. We assume that both types of hospital, PFP and PNFP, face the same production technology, meaning that given the same level and mix of inputs they are capable of producing the same mix of output.

We say mix of output because we are thinking of a hospital's output as having two components - quantity and quality. Both are produced by application of inputs - capital, labour and materials. Clearly, if there is to be anything to investigate, quantity and quality must be choice variables - for a given total level of inputs it must be possible to allocate the inputs to produce different mixes of quantity and quality. If we could observe this allocation of input effort there would be no problem answering the

[56]See, for example, John D. Birkmeyer, Andrea E. Siewers, Emily V.A. Finlayson, Therese A. Stukel, F. Lee Lucas, Ida Batista, H. Gilbert Welch and David E. Wennberg (2002): "Hospital Volume and Surgical Mortality in the United States" New England Journal of Medicine 346(15), 1128-1137, who estimate, for example, that the risk of death associated with coronary-artery bypass graft was roughly 14% higher in a low volume hospital than in a high volume hospital (the difference between a death rate of 4.8% and 5.5%).

basic question - we would simply look at the proportion of input effort allocated to quality in PFP hospitals as compared with the proportion allocated to quality in PNFP hospitals. The only reason that there is anything to study is because we can only observe total input levels (at best - data on the physician input is generally lacking) and not how inputs are allocated towards the two dimensions of output.

The fact that the quantity-quality mix is determined by resource allocation means that each hospital, whatever its ownership, is operating on a quality-quantity trade-off curve (in economic terms a production possibility frontier). If we think of the curve as drawn on a graph with quality on the vertical axis and quantity on the horizontal, the distance of the curve from the origin depends on the total quantity of inputs which the hospital has at its disposal. The more resources it has, the further out from the origin the curve on which it is operating[57]. Given the same total level of inputs, two hospitals, one PFP and one PNFP, will lie on the same trade-off curve, since those curves are determined by the technology of medicine. The hypothesis under consideration here is basically that, given two groups of hospitals lying on the same trade-off curve, the PFP hospitals will choose to operate at a point with a higher quantity coordinate and a lower quality coordinate and the PNFPs will choose a point with a higher quality and lower quantity coordinate. If we visualize the two groups of hospitals as clusters of points on the curve, the hypothesis being tested is that the two groups will form two distinct, but possibly overlapping clusters, and that the centres of the two clusters will differ, with the centre of the PFP cluster being associated with a lower level of quality of care than the centre of the PNFP cluster. On the assumption that lower quality means higher mortality, this would translate into the PFP cluster being centred on a higher mortality level. Note that, even if the two clusters are centred at different points they can still overlap. The location of the centre tells us nothing about the variance or skewness of the distribution.

In order to ensure that we are comparing like with like, then, we need to control for the trade-off curves the hospitals are on[58]. Any mortality equation, then, should include among its explanatory variables not just patient severity but also some kind of scale variable to identify the trade-off curve a particular hospital lies on. Absent that, we will be confusing scale effects with resource allocation effects. The larger a hospital is, in terms of the resources at its disposal, the further out from the origin is the trade-off curve it lies on, and the more of both quantity and quality it can produce. The important question here is how much of each it chooses to produce given its potential. If we do not control for

[57]The quantity of inputs a hospital employs in total, the quantity of output it produces and the severity mix of the patients it treats are all choice variables in the hospital's problem, which means that they are, at least potentially, statistically endogenous. This means that when we are actually estimating the equations discussed here we should use a technique which allows for this endogeneity - perhaps Instrumental Variables as opposed to Ordinary Least Squares. Choice of estimating technique is an econometric issue - in this discussion we limit ourselves to modelling issues.

[58]It can be shown that no hospital will deliberately operate below its trade-off frontier. For a PFP hospital, to do so would be to forgo potential profit, and for a PNFP to do so would be to pass up the chance to treat more patients.

scale, and if the hospitals in our two groups differ in
size, the group (call it group A) with generally smaller resource endowments will tend to show up as producing poorer quality output even if, when we look at any one trade-off curve and compare the quantity-quality choices made by the subsets of the two groups of hospitals which lie on that curve, group A hospitals consistently choose higher quality points.

Consider, for purposes of illustration, Figure 1 below. Suppose we have two broad classes of firms, type A and type B, with type A firms being smaller than type B and therefore being on the lower (closer to the origin) trade-off curve. Assume that all type A firms are tightly grouped around point A on Figure 1 and all type B firms are tightly grouped around point B. As the diagram is drawn, if we consider the ratio of quality to quantity produced by each type of firm, we will find that type A firms have a higher quality/quantity ratio than type Bs. We can see this if we draw a line from the origin through point A, and a similar line through B. the slope of the two lines represent their quality/quantity ratios. Because the line through A is steeper than that through B, A represents a higher quality/quantity ratio than does B.

Type B firms produce more of both quality and quantity than do type A firms, simply because they are larger and have more resources at their disposal. If we ignore the size differences, it would look as if type B firms have a stronger preference for quality than do type A firms. If, however, we control for the firms' resources, by controlling for the curve they are on, we will find that, given their resource levels, type A firms have a stronger preference for quality than do type B firms.

The same issue arises when our two types of firms are both distributed across the two trade-off curves, so that each type includes some larger and some smaller firms, but the type B firms are on average larger than the type As. Again, failing to control for the curve they lie on can bias our estimate of the strength of their preference for quality[59].

A carefully constructed study of the relation between hospital ownership and mortality, then, will regress mortality on patient severity variables and a scale variable of some sort. This could be the total

[59] In fact, the two ownership types are quite different in a range of characteristics. Investor-owned hospitals tend to be smaller: in 1997 investor owned community hospitals had an average bed size of 144 compared to 197 for not-for-profit community hospitals, and an average daily census of 75 compared to 125. Their geographic distribution is also significantly different: while 16% of all US community hospitals are investor owned, they are very much a southern phenomenon: in the South Atlantic census division, 26% of all community hospitals are investor owned, accounting for 25% of all investor owned community hospitals. The South Atlantic district has 13% of all US not-for-profit community hospitals. In the East South Central census district 26% of all community hospitals are investor owned, accounting for 15% of all for-profit hospitals. The East South Central district has 6% of all not-for-profit community hospitals. In Florida, 46% of all community hospitals are investor owned. Since hospital size and location can have an impact on quality (there are procedures for which outcomes are clearly better in high volume hospitals, for example) analyzing mortality rates as if hospitals had been randomized to ownership status is not indicated.

input level, or it could be the hospital volume (the quantity measure), but it should not include both. Including the level of all inputs identifies the trade-off curve a particular hospital is on. Including the quantity of output a hospital has chosen to produce in addition to the level of inputs identifies the point that particular hospital has chosen to operate at on the curve. Given those two items, there is nothing left for the ownership variable to identify.

Ideally we would prefer to have the full input set as our set of explanatory variables. If we are using quantity as an explanatory variable, we should at least omit the variable (primarily labour) inputs - these are the ones most likely to be easily reallocated between quantity and quality. We should also include variables for other, regional factors which might affect PFP and PNFP hospitals alike, but which differ across geographic areas, since PFP and PNFP hospitals are not identically distributed geographically. Rurality is one obvious factor which, judging by the literature, can affect outcomes but is likely to affect both types of hospitals in a similar manner. We are interested in identifying the degree to which ownership affects outcome after controlling for common factors.

In estimating a mortality relation it is important to keep in mind the underlying theoretical basis of the hypothesis that ownership affects quality and therefore mortality. The presumption in much of the literature is that PFP hospitals can increase their profits by cutting back on quality. This presumes either that quality is unobservable (in which case the empirical investigation should follow the literature on the econometric estimation of production functions involving unobservable endogenous quality[60]) or that patients are unresponsive to quality.

[60]There are articles in the economics literature dealing with the econometric analysis of cost functions for firms whose output contains an unobserved quality element. See B. Friedman and M.V. Pauly (1981): "Cost functions for a service firm with variable quality and stochastic demand: the case of hospitals" Review of Economics and Statistics 63, November, 630-634; P. J. Gertler and D. M. Waldman (1992): "Quality-adjusted cost functions and policy evaluation in the nursing home industry" Journal of Political Economy 100(6), 1232-1256, and Jos L.T. Blank and Evelien Eggink (2001): "A quality-Adjusted Cost Function in a Regulated Industry: the Case of Dutch Nursing Homes" Health Care Management Science 4, 201-211, for example.

Empirically neither hypothesis is likely to be correct (and both are testable). In particular, patients are likely to be concerned with the quality of care they expect to receive. Various factors will affect the demand for quality - quality is a normal good, so higher income communities are likely to be more concerned with quality, after they have paid for a given quantity of care. Areas with better insurance coverage will also be more sensitive to quality. Insurance, as structured in the United States, tends to make the patient insensitive to the price of care. The more extensive, and more comprehensive, insurance cover in an area, the less sensitive patients are going to be to the price of care. Historically this has meant that well-insured patients have tended to prefer to obtain their care from the provider with the highest perceived quality. (This was what underlay what was known as the "medical arms race" in the United States.) In order to attract patients, then, any hospital will have to offer higher quality services in higher income, better insured areas than they would in lower income and less well insured areas. The degree of the resultant increase in quality will depend on the sensitivity of demand to variations in quality, which will depend on the income and insurance factors which we have already mentioned, but also on the degree of competitiveness of the local market. If there is only one hospital in a local area, even patients with a strong preference for quality have little choice about the care they get. If the market is competitive, hospitals will have to compete much more vigorously in the quality dimension. Kessler and McClellan (2000)[61] investigated the effects of competition and concluded on the basis of differences in AMI mortality between the bottom and top quartiles of competitiveness that increased competition had the potential to improve AMI mortality by 1.46 percentage points, or 4.4 per cent. This is a larger percentage difference than Devereaux et al. ascribe to the ownership effect, and roughly on a par with van Walraven and Bell's estimate of the extra risk of death or readmission associated with being discharged from a Canadian hospital on a Friday[62].

IX. Conclusions:

There is a tendency among health care analysis to assume that quality must be the enemy of profitability. This is not necessarily the case, for several reasons. In any industry, doing a good job of producing a product can reduce the costs associated with later repairs - in health care, reducing the rate of complications is likely to be cost saving, and process improvements which reduce complication rates are likely also to reduce mortality. In addition, having a reputation for quality, especially in a field associated with life or death, can be a powerful generator of demand. Certainly hospitals do not hesitate to get the word out when they get high quality rankings[63]. In the United States, both individual

[61]Daniel Kessler and Mark McClellan(2000): "Is Hospital Competition Socially Wasteful" Quarterly Journal of Economics , May, 577-615

[62]Carl van Walraven and Chaim M. Bell (2002): "Risk of death or readmission among people discharged from hospital on Fridays" Canadian Medical Association Journal 166(13), 25 June, 1672-1673

[63]See the discussion in David Dranove(2000): The Economic Evolution of American Health

patients and managed care organizations are showing increased sensitivity to quality as the number of hospital report cards being produced increases.

There is also a tendency to assume that the mean of a distribution is completely representative of the entire distribution. To know the risk of winding up in a poor quality hospital it is also necessary to know the spread of the distribution. In Pennsylvania, The Pennsylvania Health Care Cost Containment Council produces outcomes reports on hospitals performing Coronary Artery Bypass Grafts (CABG). In 1994-95, 43 hospitals in Pennsylvania performed CABG. None were for-profit, all were non-profit of various types. The mean unadjusted mortality rate was 3.19%. However, the hospital with the highest mortality rate had an unadjusted mortality rate which was twice the average and over four times the mortality rate of the hospital with the lowest unadjusted rate. If we adjust by dividing each hospital's actual mortality rate by its expected mortality rate, the worst-performing hospital had an actual mortality rate 70% higher than its expected rate, and the best performing hospital had an actual rate less than half of its expected rate.

Using the data in the Pennsylvania reports, we regressed the actual mortality rate on a number of explanatory variables. The most powerful predictor of actual mortality was expected mortality. By itself it could explain about 42% of the variation in CABG mortality among the 43 hospitals. The number of operations performed in a hospital had a significant negative effect when it was the only other explanatory variable (besides expected mortality) in the equation, but it lost its significance when we introduced a variable for the share of a hospital's operations performed by the surgeon who performed the largest number of operations in that hospital. The coefficient on that variable was negative. The relative significance of the physician and hospital volume indicators suggests that the practice effect was more important at the surgeon level than at the hospital level. Adding a variable for the number of operations performed by the surgeon with the maximum output in each hospital increased the explanatory power of the equation to 52%.

The analysis of behavioural models is much more difficult than the analysis of traditional epidemiological models. An economic agent's behaviour will be driven by a number of objectives, and he or she will have a number of instruments which can be used to try and attain those objectives. To test hypotheses about human behaviour we must first derive predictions from a plausible model of that behaviour, taking account of how individuals will respond to policy interventions. The combination of modelling uncertainty with econometric issues, especially the issues associated with omitted variable bias, selectivity bias and endogeneity bias means that for most problems there will never be a definitive answer, just incremental movements towards better understanding. The smaller the effect being studied, the greater the ratio of noise to signal in the results of any estimation, and the more attention that must be given to details of modelling and estimation.

Care Princeton University Press, Princeton, New Jersey

A hospital is a multi-product firm, one of whose key products, quality of care, is difficult to observe. Most studies of hospital quality use a limited number of indicators: in many cases just mortality. While mortality is obviously an important outcome of hospital care, there is considerable debate about the validity of risk adjusted mortality rates as indicators of quality[64].

Efforts to improve hospital quality need to take two forms. First is the need to develop incentives that will encourage quality regardless of factors like ownership. While designing incentive systems which will encourage quality without being excessively susceptible to gaming is not easy, it can be done, as the work of Mingshan Lu at the University of Calgary indicates[65]. Second is the need to make detailed information about hospital performance widely available. Canada lags behind other countries in developing and publishing hospital report cards, and while the administrators of hospitals which rank badly can always find some reason why the rankings are particularly inaccurate in their particular case, there is considerable work, in the United States and elsewhere, on report cards and on technical issues pertaining to profiling medical care providers, and complaints that report card indicators are not absolutely perfect should not be used to justify not producing them.

[64]See, for example, David L. Zalkind and Steven R. Eastaugh (1997): "Mortality Rates as ans Indicator of Hospital Quality" Hospital and Health Services Administration 42(1), Spring.

[65]See, for example, Mingshan Lu, Ching-to Albert Ma and Lasheng Yuan (2001): Risk Selection and Matching in Performance-Based Contracting Working Paper, Department of Economics, University of Calgary