Re-visiting the Health Care Luxury Good Hypothesis: Aggregation, Precision, and Publication Biases?

Joan Costa-Font

Marin Gemmill

Gloria Rubert

Re-visiting the Health Care Luxury Good Hypothesis: Aggregation, Precision, and Publication Biases?

Joan Costa-Font[a,b,c], Marin Gemmill[b] and Gloria Rubert[c]

[a]LSE Health, London School of Economics and Political Science, London, UK.
[b] Europan Institute, London School of Economics and Political Science, London, UK
[c]CAEPS, Universitat de Barcelona, Barcelona, Catalonia, Spain.

Contact Author: Joan Costa-Font, LSE Health, London School of Economics, London, UK. E-mail: j.costa-font@lse.ac.uk

Abstract

While a growing literature examining the relationship between income and health expenditures suggests that health care is a luxury good, this conclusion is contentiously debated due to heterogeneity of the existing results. This paper tests the luxury good hypothesis (namely that income elasticity exceed unity) using meta-regression analysis, taking into consideration publication selection and aggregation bias. The findings suggest that publication bias exists, a result that is robust to the meta-regression model employed. Publication selection and aggregation bias also appear to play a role in the generation of estimates. The corrected income elasticity estimates range from 0.4 to 0.8, which cast serious doubt on the validity of luxury good hypothesis. Nonetheless, due to the importance of aggregation, we cannot reject the luxury good hypothesis for aggregate time series data.

Keywords: meta-regression analysis, health care, luxury good, income elasticity, aggregate health expenditure, regional health expenditure

JEL Classification: I1, I10, I11, I18

1. Introduction

Beginning with the seminal paper by Newhouse (1977), a contentious debate has raged over the income elasticity of demand[1], the central question being whether health expenditures increase faster than per capita income. The general finding has been that income elasticity estimates exceed unity, implying that health care is a luxury good (Newhouse, 1987, Gerdtham and Johnson, 2000). Given the marked implications for the allocation of health care resources, the debate has often centered on the methodological robustness of elasticity estimates. The argument reads that if health care is a "necessity", this necessitates more redistribution of health care resources and arguably greater public involvement in health care. That is, the value of income elasticity provides insight into the optimal level of health expenditures in the economy and the efficient proportion of public and private health spending[2].

Some researchers suggest caution in interpreting the early results that health care is a luxury good as misspecification may be a possibility (Culyer, 1987). As a result, the methodological debate has focused on the existence of specific controls, such as health system controls (Gerdtham and Johnson, 2000), and the methods used, primarily the statistical properties of the data. Another source of variation is the heterogeneity of health care (Parkin, 1987; Gertham, 1992, Roberts, 2000), which depends on whether the data is measured at the national, regional, or individual level (Getzen, 2000; DiMatteo, 2003). The interdependence of several forms of health care implies that an

---

[1] We define the income elasticity of demand as the percentage change in health expenditures that is associated with a one percent change in income. The formula is given by: $e_I = (\partial HE / \partial I)(HE / I)$ where *HE* represents health expenditures and *I* represents income. If the income elasticity is less than one, then the health care expenditures are a necessity good. If the income elasticity is greater than one, then health care expenditures are a luxury good.

[2] If the income elasticity exceeds unity, some might argue that universal health coverage is unnecessary as the private market is more efficient in the provision of coverage. Alternatively, it may be that income inequalities are prevalent, although this interpretation rests on the assumptions that most health care consumption is necessary and that there is significant unmet demand among lower-income populations.

aggregate analysis is more intuitive, but there may be biases in employing aggregate estimates to infer individual behavior. Most research that utilizes aggregate data relies on country-level aggregation, mainly due to data availability, but interestingly, studies employing regional data do not necessarily find income elasticities below one. This difference might be due to aggregation bias or perhaps to the spurious regression of health care expenditure on income when aggregate time series is used (Getzen, 2000). Similar problems with aggregation have been found in the crime data (Glaeser et al, 2003; Glaeser and Sacerdote, 2007). However, the increasing availability of data and statistical methods implies the need for a paper that aggregates the existing studies on the basis of such effects. To date no study has investigated empirical biases in the income elasticity literature.

Publication selection is an important and commonly identified bias (Stanley, 2008; Stanley, 2005)[3]. Precision is another important given that there may be heterogeneity in the estimates given that all authors inevitably select different samples and employ different controls. Due to the inevitable heterogeneity of methods, samples, and classifications across different studies, meta-analysis or meta-regression analysis (MRA) is needed to model and estimate this variation and thereby to determine whether health care is a luxury good. Meta analysis integrates the existing estimates of a defined outcome variable (Farly, 1982) and assumes that the individual studies can be homogenized through a standard measure of empirical effect or effect size (Glass *et al* 1981), which is held constant across all studies. In economics, the effect size is usually an elasticity, a partial correlation coefficient, or regression coefficient thought to measure

---

[3] The existence of publication selection points toward a preference – typically among journal referees, editors, editorial boards, and even authors themselves - for statistically significant results that confirm the prevailing theoretical paradigm

some important underlying economic phenomenon (Stanley, 2001; Doucouliagos, 2005).

Given the clear heterogeneity in economics research, meta-regression analysis is the appropriate tool to explain the systematic heterogeneity of reported results and to obtain a flexible estimate that can be adapted to context-specific circumstances (Stanley and Jarrell, 1989). MRA entails a regression analysis of existing studies with controls for the study type, the sample characteristics, and the scope and precision of the elasticity estimate, allowing us to test the sensitivity of the parameter of interest to given objective characteristics. There have been dozens of MRA applications in the economics literature (Roberts, 2005), but there are only a few known applications to health care (Asensio-Boadi et al., 2007; Gemmill et al, 2007).

To address the debate regarding whether health care is a luxury or necessity good, this paper pools the existing aggregate income elasticity estimates from social science and economic related journals. We then apply MRA to obtain a corrected income elasticity estimate, accounting for the precision, publication selection, and aggregation nature of the included papers. The analysis is restricted to total health expenditures given that studies which consider specific expenditure types (e.g., pharmaceutical or inpatient) or employ individual-level data might not produce comparable estimates. Once we control for the relevant study-specific factors, it becomes clear that income elasticity estimates suffer from publication bias. After removing the publication bias, we can no longer conclude that health care is luxury good.

The paper is organised as follows: Section 2 provides an overview of the existing studies, distinguishing between those that employed national-level data and those that used regional-level data. Section 3 describes the methods employed in the analysis and offers more detail on the use of meta-regression analysis. Section 4 details the results, and Section 5 concludes with a discussion.

2. Brief overview of the literature

We focus on those studies that use national- and regional-level aggregate data to estimate the relationship between income and overall health expenditure. The reasons to do that lies in that health care is largely heterogeneous, and income elasticity's are to vary with each different type of service (inpatient, outpatient, drugs etc). However, the policy relevant decision question lies on predicting the overall size of health expenditures. Finally, even if we were to examine individual level studies, a comparability problem would not make them comparable. ,

2.1. Studies using country-level data

Two literature reviews that focused on country-level analyses of the relationship between income and health expenditures found that most papers reported income elasticity coefficients greater than one (Getzen, 2000; Gerdtham and Johnson, 2000). Getzen (2000) argued that while evidence indicates that health care is a necessity at the individual level, it is a luxury good at the aggregate level, although Hansen and King (1996) suggest that this relationship could be spurious.

In dealing with international health care expenditure functions, the availability of data has fostered a significant amount of empirical work. However, health care systems are

heterogeneously managed, regulated, and financed, and accordingly, there are sizeable differences in the health care packages among OECD counties. As a result, it is doubtful that data from different countries is measuring the same outcome. Another issue is that there might be a 'stability problem' when examining data over a large period of time (Jewell et al, 2003; Clemente *et al*, 2004).

As these methodological issues have led many to question the validity of the elasticity results (Clemente *et al*, 2004; McCoskey and Selden, 1998; Hansen and King 1996; Blomqvist and Carter, 1997; Karatzas, 2000; Roberts, 2000), some researchers have addressed specific methodological issues underlying the determination of the health care expenditure function. In particular, these studies account for the potential non-stationarity of the data, although there is no agreement on whether the data is co-integrated (Gerdthan and Lothgren, 2000; Clemente *et al*, 2004, Herwaetz and Theilen, 2003). Others have used panel data methods to account for potential differences in tastes and preferences in the health care expenditure function (Hitris and Possnett, 1992; Di Matteo and Di Matteo, 1998), but none of these analyses have considered spatial interactions, the existence of which might invalidate some of the existing conclusions. Some of the literature has focused on causality problems that occur when examining health expenditure and GDP, and this has been examined in the Spanish health care system (Devlin and Hansen, 2001). Okunade and Suraratdecha (2000) use a dynamic Engel specification of a Box–Cox expenditure model to account for the existence of inertia, especially in publicly financed health systems. It is important to note that they find that per capita real GDP's income elasticity behaves as a necessity in 20 of the 21 OECD countries.

In a further attempt to overcome some of the institutional heterogeneity issues, some studies have controlled for health system characteristics. Gerdtham et al. (1998) is one of the few studies that examines the influence of a set of institutional reforms. They find that health care systems where physicians serve as gatekeepers are consistently and statistically significant, and they are associated with lower health expenditures. Gerdtham et al. (1998), Hansen et al. (1996), and Roberts (1998) explicitly control for the percentage of public expenditures, but find mixed results.

2.2 Studies using regional-level data

Importantly, the controls for institutional context may be insufficient to overcome institutional heterogeneity (Di Matteo and Di Matteo, 1998), and thus some studies have used sub-national data to overcome this bias. There have been regional studies conducted in five countries (Canada, Italy, Spain, Switzerland, and the United States), and all of these studies find an income elasticity below one (Cantanero, 2005; Costa-Font and Pons, 2006; Crivelli et al, 2007; Di Matteo, 2003; Gionannoni and Hittris, 2002; Vater and Rüefli, 2003). As data at the regional level has only become available relatively recently, most of the studies examining health expenditures at the regional level are from the last ten years.

2.3 Aggregation effects

The bulk of evidence supporting the luxury good theory has been drawn from aggregate datasets, and there may be difficulties in drawing inferences about individual behaviour from aggregate data (Glaeser et al, 2002). Most studies using regional-level data have found elasticity values below one, while studies using national-level data find elasticity

values above one. The difference in results could be due to the aggregation effect

(Glaeser et al, 2002). In particular, the association between a country's income level and

health care expenditures can be affected by strategic complementarities; such as

preference or information spillovers due to information asymmetries. Furthermore,

individual-level income does not adequately capture the effect of technology, while at

the national level, income includes the technology effect. In practice, measuring the

technology effect is difficult because there is no accepted measure of technology

change. Another reason for casting doubts on behavioural inferences resulting from

aggregate data is that individual-level budget constraints differ from those at the

regional or national level, particularly in the presence of universal or extended insurance

coverage. The implication of this discussion is that aggregation effects may be

important, and they have a decisive effect on the luxury hypothesis.

3. Data and Methods

3.1. Meta-Regression Methods

The intent of this paper is to determine the corrected magnitude of the income elasticity

estimate derived from meta-regression analysis and to examine the extent to which the

predicted elasticity differs from one ($\beta_i \neq 1$). Specifically, the goal is to establish

whether the elasticity is greater than one (a luxury good) or less than one (a necessity

good) after controlling for study-specific characteristics.

Meta-regression analysis involves collecting the outcome variable and relevant study-

specific information from the existing literature in a systematic manner to determine

which factors influence the variability of the treatment variable (Stanley and Jarrell,

1989). These factors are then recorded as covariates, creating the meta-regression

dataset. The assumption is that each observation is drawn from an overall statistical population. Based on this compiled dataset, we can test our main hypothesis that the income elasticity of demand is greater or less than one and identify the factors that influence this treatment variable.

This technique has the distinct advantage of being less subjective than literature reviews where the researcher is interested in the average effect of a particular outcome variable. A literature review is subjective in that the researcher determines the inclusion criteria for the literature, the method of interpreting the results, and the potential reasons for varying results. Systematic literature reviews offer a methodological improvement and provide techniques for reducing the subjectivity, but researchers still have considerable leeway when deciphering the results and crediting various factors to variation in the outcome variable (Stanley, 2001). As a result, the ultimate aim of MRA is to overcome some of the pitfalls of literature reviews, allowing us to obtain an "estimate of estimates" with some acceptable precision.

a. Funnel-Asymmetry Test (FAT) the Precision Effects Test (PET)

The analysis begins with the collation of information from relevant studies, where we have *N* estimates of η_i (the dependent variable) and *i=1,…,N*. We identify the *k* characteristics of the diverse studies and integrate the findings as follows:

$$\eta_i = \beta + \beta_0 S_\eta + \sum_{k=1}^{K} \beta_k X_{it} + \varepsilon_i \qquad (1).$$

The reported income elasticity estimate of each i study (η_i) equals the real income

elasticity estimate (β) adjusted for the standard error of η_i $(S_{i\eta})$ and the k

characteristics (X_{ik}) of each published study. The X_{ik} are the independent variables and

account for the processes which affect the estimates of elasticity, while the parameters

(β_0, β_k) represents the biases associated with specific characteristics that lead to

misspecifications (Stanley and Jarrell, 1989). The covariates might be variables

measuring the quality of the study (particularly we use the impact factor of the journal),

numerical continuous variables accounting for the study size, and any other relevant

characteristics of the study Given that estimates are obtained by varying degrees of

precision, it is possible to control for publication bias by including the standard error of

the estimate in the regression. The $\beta_0 S_\eta$ term represents publication selection (Stanley

2005; Stanley, 2008). Studies with large standard errors will need to search harder and

longer to find the very large estimates of income elasticity that are statistically greater

than one.

Given that the model is based on estimates from previous regressions, it is important to

examine the distributional properties of the data. In the absence of publication selection,

estimates will vary randomly, hence *symmetrically*, around the "true" effect (Stanley,

2008).

Because the model is based on estimates from previous research, it is important to

examine the distributional properties of the data. While MRA coefficients should be

unbiased and consistent (Stanley and Jarrell, 1989), the fact that the revised studies are

drawn from different datasets, have differing sample sizes, and utilize different controls

and methods generally leads both to heterogeneity and to heteroskedastic error terms. thus, it is prudent to use a weighted least squares (WLS) regression. Here, WLS divides equation (1) by the standard error of η_i (S_η), causing the dependent variable to become the t-statistic.

$$t_i = \frac{\eta_i}{S_\eta} = \beta_o + \beta_1 \frac{1}{S_\eta} + \sum_{k=1}^{K} \frac{\beta_k X_{ik}}{S_\eta} + \mu_i \tag{2},$$

where t_i is the t-value associated with the ith reported estimate in the MRA dataset. The X_{ik} are the study specific controls including the impact factor of the journal (quality), institutional features of the data (health system type), type of data (public or private), and the level of aggregation (regional or national)—see Table 1. MRA (2) allows us to test for the presence of publication selection. The associated test is called the 'funnel asymmetry test' (FAT), because the asymmetry of the funnel plot is evidence of publication selection bias (Egger *et al.* 1997, Sutton *et al.* 2000 and Stanley 2005)). FAT has become common practice when performing meta-regressions in economics. Its null hypothesis is $H_0 : \beta_0 = 0$. When H_0 is rejected, we have evidence of the presence of publication selection.

Equation (2) also contains the Precision Effect Test (PET), which allows us to identify an empirical effect, regardless of publication bias (Stanley, 2005). The null hypothesis of PET is $\beta_1 = 0$ and it tests for the presence of an empirical effect (in our case, whether income affects health care expenditures) and is robust to the presence of publication selection bias (Stanley, 2008). However, the estimate of is biased downward when there is a genuine empirical effect(Stanley, 2008), as there is here. Thus Stanley and

Doucouliagos (2007) develop an additional MRA model that reduces the bias associated with correcting for publication selection (see PEESE below). PET can also suffer from inflated Type I errors if the existing heterogeneity is larger than the sampling error (Stanley, 2005). Therefore, it is important to perform further confirmatory tests using MSE and degrees if freedom which suggest the opposite.

b The Precision Effect Estimate with Standard Error(PEESE)

As an extension of model (2), a Heckman-like correction, the Precision Effect Estimate with Standard Error (PEESE) model, can be used to obtain an estimate that is robust to publication selection bias. For a complete derivation of this model, please see Stanley and Doucouliagos (2007). The PEESE equation starts from the premise that there is a nonlinear relationship between the observed outcome and its standard error, yielding the equation:

$$\eta_i = \alpha + \alpha_0 S_\eta^2 + \sum_{k=1}^{K} \alpha_k X_{it} + \varepsilon_i \qquad (4),$$

Assuming heteroskedasticity in the error term, we again apply the WLS correction to yield:

$$t_i = \alpha_0 S_\eta + \frac{\alpha}{S_\eta} + \frac{\sum_{k=1}^{K} \alpha_k X_{it}}{S_\eta} + \delta_i \qquad (5),$$

so that α estimates the magnitude of the empirical effect corrected for publication selection. As with any other empirical specification, as long as the model is not misspecified, it measures the specific meta-effects. One method of gauging the

sensitivity of the model to misspecifications is to vary the independent variables and measure the effects.

c Meta-significance testing

A second meta-regressin model serves as a robustness test—meta significance testing (MST). It exploits the fact that if there is a genuine underlying effect, there will also be a logarithmic relationship between a study's *t*-statistic and its degrees of freedom. Statistical theory predicts that the t-ratio will be related to the square root of the degrees of freedom, or :

$$E(\log|t_i|) = \gamma_o + \gamma_1 \log(df_i) \tag{6},$$

where $\gamma_1 = 0$ would confirm the null hypothesis of no effect, while an empirical effect implies that this coefficient is exactly 0.5 (Stanley, 2005). $0 < \gamma_1 < 0.5$ reflects the existence of publication bias.

d Homogeneity

Homogeneity, i.e. whether there is a common mean, is another aspect of the dataset that needs to be considered. We can test for homogeneity using the $Q = \sum (\eta_i - \overline{\eta}_{\mathrm{var}(\eta)})^2 / \mathrm{var}(\eta_i)$ statistic, where η_i is each elasticity estimate, $\overline{\eta}_{\mathrm{var}(\eta)}$ is a weighted average of each elasticity estimate corrected by its variance, and $\mathrm{var}(\eta_i)$ is the variance of each estimate. Under the null hypothesis of homogeneity, Q is distributed as χ^2_{N-1} where *N* is the number of studies. If the null hypothesis of homogeneity is rejected, this suggests that regression analysis is needed.

14

3.2 Data selection

Our search for all available evidence on the income elasticity of health care involved prescreening, selecting, and then classifying the income elasticity values and the associated study characteristics to create the MRA database. In developing the database, we identified and cross-referenced published studies using Econlit, Medline, and Sociofile up until 2006. An important point is that we restricted our sample to income elasticity estimates derived from aggregate datasets and published in social science journals. The intuition regarding aggregate estimates was explained previously. Additionally, an interesting sub-question to consider in the analysis is whether publication bias exists in this particular area of the literature.

There are many potential predictors of the income elasticity estimate—see Table 1 All of these potential independent variables can be identified from the specific papers collected for the analysis and are classified as follows:

(a) measurement and methods or study-specific characteristics (namely. the number of observations),

(b) institutional setting (namely the type of insurance coverage, the type of health system)

(c) publication or dissemination effects (which refers to whether published in a social science journal, quality or impact factor of the journal), and

(d) method or data specific controls (such as the presence of outliers).

Each of these predictors is intended to capture specific biases that influence the outcome variable. One of the most important considerations in the regression may be the coefficient on the standard error variable as a positive coefficient may be indicative of a

publication or dissemination effect. That is, some social science and economics related journals might be more interested in publishing studies with income elasticity estimates greater than one as this confirms the luxury good hypothesis. A further example is that estimates can vary significantly across study characteristics, such as the number of observations and the journal where the estimate was published. The institutional setting, such as whether the estimate was generated from a tax-based or social insurance-based country, may also be a key factor in determining the income elasticity as it reflects the distribution of income across the population and possible cultural factors. Finally, the presence of outliers related to specific studies that are of varied quality is another important effect.

In most cases both the income elasticity and the associated standard error were available in the paper. In a few cases, the standard error was not provided, but where possible we calculated this either from the given t-value or from the mean square error (MSE). In the case where the study reported the t-value associated with η_i rather than the standard error, we used the formula for the t-value:

$$t - value = \frac{\eta_i - 0}{s.e.(\eta_i)}, \tag{7}.$$

Thus, we substituted in the reported values of η_i and the reported t-value to solve for $s.e.(\eta_i)$.

4. Results

After the data collection process, our final sample consisted of 167 comparable elasticity estimates from a set of 48 published studies. Before proceeding to the meta-regression, we first considered the possibility of homogeneity in the sample by calculating the Q-statistic. At a value of 30,641 (p=0.000), extremely high, indicating that there is significant heterogeneity in our data. No single income elasticity can represent this data, and meta-regression analysis is needed to explain the systematic heterogeneity.

The next step was to visually examine the data to get a feel for any publication bias. Figure 1 is a funnel plot, which plots the elasticity estimates against a measure of precision (1/s.e). With the exception of a few outlier estimates, most of the income elasticity coefficients range from 0 to 2. The funnel plot also suggests that the value of one is likely at the centre of the distribution, and with the exception of few outliers, the distribution appears to be symmetrical around that value. Interestingly, if we look at the descriptive statistics for the income elasticity value (Table 1), there is significant variability as indicated by the values at the 10th and 90th percentiles.

[Insert Figure 1 and Table 1 about here]

Following the MRA methods outlined in Section 3, we then ran several meta-regressions. We first estimate the FAT-PET–MRA, equation (2). Table 2 show that our estimate of β_0 is significant and positive, meaning that we can reject the null hypothesis of no selection bias (according to FAT). In line with the interpretation of the funnel plot, the direction of the bias is positive. However, when more controls are

introduced, part of the selection or publication bias appears to be picked up by the controls, and FAT becomes insignificant.

[Insert Table 2 about here]

Next, consider the coefficient on $\beta_{1/s.e}$, which is also an estimate of the income elasticity of demand after correcting for selection bias (according to PET). The coefficient is positive statistically significant, with values ranging 0.26 to 0.71. Unfortunately, this coefficient is known to be biased downward when there is a genuine effect (Stanley, 2008). To reduce the bias in correcting for publication bias, Stanley and Doucouliagos (2007) have developed the PEESE variation on the FAT-PET-MRA (see Table 3).

Table 3 provides the estimates of the PEESE model (from equation (5)) where $\hat{\alpha}_{1/s.e}$ is the effect corrected for publication selection following Stanley and Doucouliagos (2007). The precision-corrected elasticity estimate lies between 0.38 and 0.84, depending on the specific study controls introduced. This provides a clear indication that health care may not be a luxury good. As expected, these estimates are larger than the corresponding PET coefficients and significantly less than one Wald tests only rejected the hypothesis that equals one for the first simple specification at the conventional 5% significance level.. These results are in line with previous FAT-PET-MRA, overall indicating that health is not a luxury good. Also, the coefficients on the moderator variables are quite consistent.

[Insert Table 3 about here]

Perhaps more important is the fact that two moderator variables are statistically significant and help to explain a great deal of the heterogeneity found in this area of research. Both the use of regional data (*region*) and the journal's impact factor (*impact*), were consistently significant. It appears that studies using regional data yield lower income elasticity values, with coefficients being negative and ranging from 0.664 to 0.51. This is consistent with the aggregation bias hypothesis, which remains irrespective of the introduction of additional controls. As for the second effect, there is a positive relationship between the impact factor and the income elasticity, suggesting that high impact factor journals have a preference for significant and larger elasticity estimates.

Finally, to insure the robustness of our findings, we also report the MST-MRA (Table 4), which tests the existence of a logarithmic relationship between the degrees of freedom and the t-value. Consistent with the previous results, we find a significant and robust effect that confirms the existence of selection bias. Post-estimation tests reject the null hypothesis of the coefficient being 0.5 ($F(1, 390)=7.73$), which can be interpreted as additional evidence confirming the intuition of publication bias.

[Insert Table 4 about here]

5. Conclusion

This paper has examined the existence of publication bias along with aggregation and precision effects to revisit the hypothesis of health care being a luxury good. Drawing from a battery of existing methodologies (FAT, PET, PEESE and MRT), our results suggest the publication bias does exist. Interestingly, we find that the income elasticity

19

of demand for health care lies between 0.4 and 0.8, which cast doubt upon the hypothesis that health care is a luxury good. This result is consistent with the proposal that health care is an individual necessity and an aggregate luxury (Getzen, 2000) and possibly of some spurious relationship might explain that income elasticities are larger than one. However, it is important to note that we do not assume that income elasticity remains constant over time, not that there is a "single income elasticity" but instead that health care income elasticity is not necessarily higher than one after undertaking relevant adjustments.

We also find that two study controls are consistently important predictors of the elasticity value. Studies using regional data yielded lower elasticity values, providing evidence for the existence of aggregation effects. Journal quality is also an important predictor, and it seems that journals with a higher impact factor, namely more prestigious journals especially within the health economics discipline, exhibit a systematic tendency to report larger elasticity effects.

Other controls such as institutional and methodological factors did not appear to influence the elasticity estimates. When more controls are introduced, part of the selection or publication bias appears to be picked up by the controls, and FAT becomes insignificant.

It is important to bear in mind the potential limitations of this analysis. Over time, as studies employ more controls, income elasticity estimates have declined markedly (Sen, 2006). It may this effect that is being reflected in our meta-regression results. In addition, there may be other important characteristics, such as indicators of health expenditure types (e.g. pharmaceutical, outpatient, inpatient), that explain heterogeneity

in elasticity estimates. However, given that expenditures are not independent – as more expenditure in drug treatments might result form a previous substitution from inpatient care to drugs-, and instead reflects an underlying demand for health channelled through agency relationship – as doctors are the privileged health care decision makers- , an income elasticity for health care appears to be more likely, even though it might vary across populations and institutional settings. Future research could account for other sources of heterogeneity using individual-level data.

References

Asensio-Boadi. F, Peter, T. J and Coast, J (2007). Exploring differences in empirical time preference rates for health: an application of metaregression. Health Economics, in press**.**

Blomquivst, A.G. and Carter, R.A.L. (1997) Is health care really a luxury good? *Journal of Health Economics* 16(2): 207-229.

Cantanero, D (2005). Decentralization and health care expenditure: the Spanish case. *Applied Economics Letters*, 2005, 12, 963–966

Clemente, J., Marcuiello, C., Montanes, A. and Pueyo, F. (2004) On the international stability of health care expenditure functions: are government and private functions similar? *Journal of Health Economic*s 23(3): 589-613.

Costa-Font, J; Pons-Novell, J (2007). 'Public health expenditure and spatial interactions in a decentralized national health system.' *Health Economics* 16 (3): 291-306

Crivelli, L , M Filippini, I Mosca (2007). Federalism and regional health care expenditures: an empirical analysis for the Swiss cantons. *Health Econ*. 15: 535–541 (2006)

Culyer AJ (1988). Health care expenditures in Canada: myth and reality; past and future. Canadian Tax Foundation paper no. 82..

Devlin, N. and Hansen, P. (2001) Health care spending and economic output: granger causality. *Applied Economics Letters* 8(8): 561-564.

Doucouliagos, H. and T.D. Stanley (2007).Theory Competitioin and Selectivity: A Meta-Meta-Analysis", Presented at the Aarhus Colloquium of Meta-Analysis in Economics, Sønderborg, Denmark, September.

Di Matteo, L. and Di Matteo, R. (1998) Evidence on the determinants of Canadian provincial government health expenditures 1965-1991. *Journal of Health Economics*, 17(2): 211-228.

Di Matteo, L. (2003) The income elasticity of health care spending. A comparison of parametric and nonparametric approaches. *European Journal of Health Economics* 4(1): 20–29

Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *Lancet* 350, 326-9.

Falancheti, D (2005). Aggregate health expenditures, national income, and institutions for private property. *Economics & Politics*, 17(3): 393-431.

Freeman, D.G. (2003): "Is health care a necessity or a luxury? Pooled estimates of income elasticity from US state-level data", *Applied Economics*, **35**, 495-502

Gerdtham, U.G., Jonsson, B. (2000), "International comparisons of health expenditure: theory, data, and econometric analysis", Handbook of Health Economics, A.J. Culyer, J.P. Newhouse, [eds.], Volume 1A, North Holland Publishing, Elsevier Science, Amsterdam

Gemmill, M; Costa-Font, J.; McGuire, A (2007). In search of a corrected prescription drug elasticity estimate: A meta-regression approach.' *Health Economics* 16, no. 6 (2007), pp. 627-643.

Gerdtham, U.G. and Löthgren, M (2000) On stationary and cointegration of international health expenditures and GDP. *Journal of Health Economics* 19(4): 461-475.

Gerdtham, U.G., Jonsson, B., Macfarlan, M. and Oxley, H. (1998) The determinants of health expenditure in OECD countries. In P. Zweifel (ed.) *Health, the Medical Profession, and Regulation.* Dordrecht: Klewer.

Gerdtham, U.G. (1992) Pooling international health care expenditure data. *Health Economics* 1(4): 217–223.

Gerdtham, U.G. and Jonsson, B. (1991) Conversion factor instability in international comparisons of health care expenditure. *Journal of Health Economics* 10(2): 227–234.

Gerdtham, U.G, Sogaard, J., Anderson, F. and Jonsson, B. (1992) An econometric analysis of health care expenditure: a cross-section study of OECD countries. *Journal of Health Economics* 11(1): 63–84.

Gerdtham, U.G. and Jonsson, B. (1991) Price and quantity in international comparisons of health care expenditure. *Applied Economics* 23(9): 1519–1528.

Getzen, T. (2000) Health care is an individual necessity and a national luxury: applying multilevel decision models to the analysis of health care expenditures. *Journal of Health Economics*, 19(2): 259-270.

Getzen, T and Poullier, J.P (1992) International health spending forecasts: concepts and evaluation. *Social Science and Medicine* 34(9): 1057-1068.

Gianonni, M and Hittris, T (2002) The regional impact of health care expenditure: the case of Italy. *Applied Economics* 34(14): 1829 – 1836

Greenwood, C.M.T, Midgley, J.P., Mattheu, A.J. and Logan, A.G. (1999) Statistical issues in a metaregression analysis of randomized trials: impact on the dietary sodium intake and blood pressure relationship. *Biometrics* 55(2): 630-629.

Glaeser, E.L., B.I. Sacerdote, and J.A.Scheinkman (2003). The social multiplier. The European Economic Association 1(2-3), 345-353.

Glaeser, E and Sacerdote, B (2007). Aggregation Reversals and the Social Formation of Beliefs. NBER Working Paper No. 13031

Hansen, P. and King, A. (1996) The determinants of health care expenditure: a cointegration approach. *Journal of Health Economic*s, 15(1): 127-137.

Herwaetz, H. and Theilen, B. (2003) The determinants of health care expenditure: testing pooling restrictions in small samples. *Health Economics* 12(2): 113-124.

Hitris, T. and Nixon, J. (2001) Convergence of health care expenditure in the EU countries. *Applied Economics Letters* 8(4): 223-228.

Hitris, T. (1997) Health Care expenditure and integration in the countries of the European Union. *Applied Economics* 29(1): 1-6.

Jewell, T., Lee, J. Tieslau, M. and Strazicich, M.C. (2003) Stationarity, health expenditures and GDP: evidence from panel unit root tests with heterogeneous structural breaks. *Journal of Health Economics* 22(2): 313-323.

Karatzas, G. (2000) On the determination of the US aggregate health care expenditure. *Applied Economics* 32(9): 1085-1099.

Barros, P.P (1998), "The black-box of health care expenditure growth determinants", *Health Economics*, 7: 533-803

Lybeck, J.A. (1988) Comparing government growth rates: the non-institutional vs. the institutional approach. In J.A. Lybeck and M. Henrekson (eds.) *Explaining the Growth of Government*. Amsterdam: North-Holland.

Levanggi, R and Zanola, R (2003). Flypaper Effect and Sluggishness: Evidence from Regional Health Expenditure in Italy. *International Tax and Public Finance*, 10: 535-547.

Macinko, J., L. Shi, B. Starfield, and J.T. Wulu, 2003. Income inequality and health: A critical review of the literature. *Medical Care Research and Review* 60(4), pp. 407-452.

McCloskey, S.K. and Selden, T.M. (1998) Health care expenditures and GDP: panel data unit root test results. *Journal of Health Economics*, 17(3): 369-376.

Newhouse, J.P. (1977) Medical care expenditure: a cross-national survey. *Journal of Human Resources*, 12(1): 115-125.

Newhouse JP (1987) Cross-national differences in health spending: what do they mean? J Health Econ 6:159–162

Newhouse, J.P. (1992) Medical care costs: how much welfare loss? *Journal of Economic Perspectives* 6(3): 3-21.

Okunade, A.A. and Suraratdecha, C. (2000) Health care expenditure inertia in the OECD countries: A heterogeneous analysis. *Health Care Management Science* 3(1): 31-42.

Okunade, A.A and Murthy, V.N.R (2002) Technology as a major driver of health care costs: a cointegration analysis of the Newhouse conjecture. J*ournal of Health Economics* 21(1): 147-159.

Parkin, D., McGuire, A. and Yule, B. (1987) Aggregate health care expenditures and national income: is health care a luxury good? *Journal of Health Economics* 6(2): 109-127.

Roberts, J. (2000) Spurious regression problems in the determinants of health care expenditure: a comment on Hitris (1997). *Applied Economics Letters* 7(5): 279-283.

Roberts, C.J. (2005) Issues in meta-regression analysis: an overview. *Journal of Economic Surveys* 19(3): 295-298.

Sen, A. (2005) Is Health care a luxury? New evidence from OECD data. *International Journal of Health Care Finance and Economics* 5(2): 147-164.

Stanley, T.D. and S.B. Jarrell, 1989. Meta-regression analysis: A quantitative method of literature surveys, *Journal of Economic Surveys*, 3: 54-67.

Stanley, T. D. (2001). 'Wheat from chaff: meta-analysis as quantitative literature review', *Journal of Economic Perspectives*, Vol. 15, pp. 131–150

Stanley, T. D. (2005). 'Beyond publication bias', *Journal of Economic Surveys*, Vol. 19, pp. 309–345

Stanley, T.D. and Chris Doucouliagos (2007). "Identifying and Correcting Publication Selection Bias in the Efficiency-Wage Literature: Heckman Meta-Regression." Deakin University Working Paper, 2007-11.

Stanley, T.D. (2008). "Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection," *Oxford Bulletin of Economics and Statistics*, 70(1): 103-127.

Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R., Jones, D.R., 2000. Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal* 320, 1574-77.

Vaseduva, M.N.R. and Okunade, A.A. (2000) Managed care, deficit financing and aggregate health care expenditure in the United States: a cointegration analysis. *Health Care Management Science* 3(4): 279-285.

Vatter, A., Ruefli, C. (2003): "Do political factors matter for health care expenditure? A comparative study of Swiss cantons", Journal of Public Policy, **23 (3)**, 301–323.
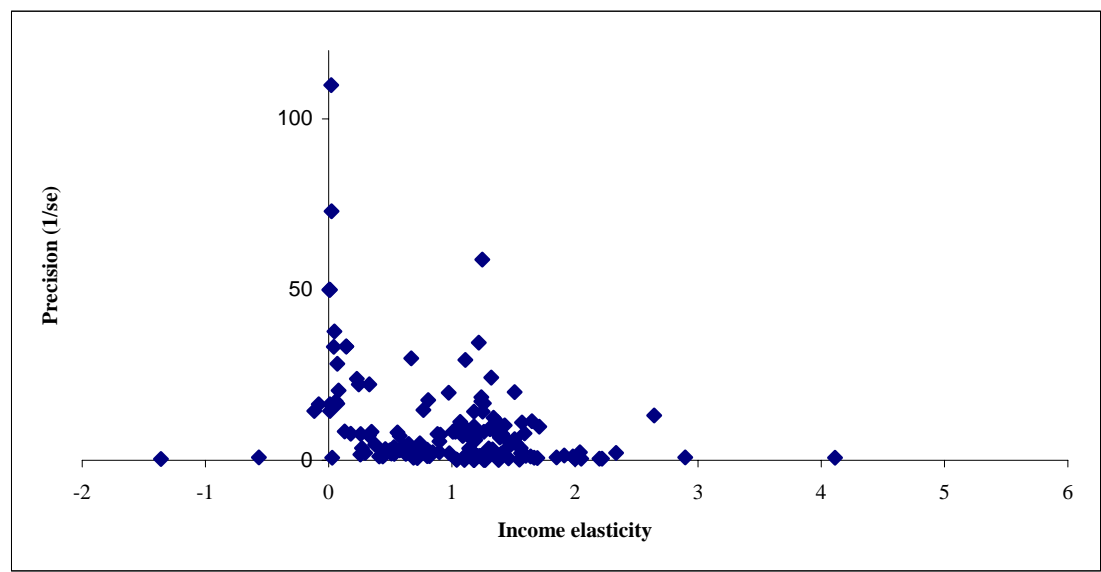
Figure 1. Funnel Plot

Table 1. Definitions of the variables and summary statistics (N=167)

Variable	Definition	Mean[a]	Median	10th percentile	90th percentile
inc_elasticity	Income elasticity of demand	0.999 (0.073)	0.908	0.0793	1.654
std_error	Standard error of the elasticity	1.215 (0.397)	0.290	0.0450	1.471
region	Indicates whether the data was regional (vs. national)	0.246 (0.033)	0.000	0.0000	1.000
df	Degrees of Freedom of each database	421.20 (36.11)	24.2	17.1	671.3
nhs	Dummy for the percentage of NHS observations in the study	0.532 (0.031)	0.500	0.0000	1.000
public	Dummy for public health expenditure	0.090 (0.022)	0.000	0.0000	0.000
impact	The impact factor of the medium where the paper was published	0.907 (0.075)	0.300	0.0000	2.500
panel	Indicates whether the study used panel data techniques	0.174 (0.029)	0.000	0.0000	1.000

[a]standard errors in parentheses

Table 2. Funnel Asymmetry Test (FAT) and Precision Effect Test (PET)

	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)
$\beta_{1/s.e}$	0.265[b] (0.148)	0.712[a] (0.321)	0.644[a] (0.320)	0.645[a] (0.321)	0.665[b] (0.412)	0.662[a] (0.331)
β_{region}		-0.634[a] (0.303)	-0.588[a] (0.293)	-0.613[a] (0.299)	-0.605[b] (0.324)	-0.515[b] (0.291)
β_{impact}			0.220[a] (0.099)	0.222[a] (0.100)	0.223[a] (0.094)	0.229[a] (0.087)
β_{public}				0.085 (0.094)	0.086 (0.089)	0.103 (0.069)
β_{NHS}					-0.030 (0.360)	
β_{panel}						-0.157 (0.208)
β_0	3.673[a] (1.259)	2.305[b] (1.402)	1.409 (1.051)	1.351 (1.054)	1.308 (1.076)	1.181 (1.159)
R^2	0.16	0.452	0.513	0.505	0.505	0.516

[a]significant at the 5% level, [b]significant at the 10% level

Table 3. Precision Effect Estimate with Standard Error (PEESE)

	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)
α_0	0.022 (0.020)	0.013 (0.013)	0.008 (0.008)	0.008 (0.008)	0.008 (0.008)
$\alpha_{1/s.e}$	0.387[a] (0.139)	0.824[a] (0.278)	0.691[a] (0.296)	0.691[a] (0.297)	0.742[a] (0.371)
α_{region}		-0.690[a] (0.290)	-0.610[a] (0.284)	-0.638[a] (0.287)	-0.613[a] (0.322)
α_{impact}			0.262[a] (0.119)	0.261[a] (0.119)	0.261[a] (0.121)
α_{public}				0.100 (0.100)	
α_{NHS}					0.103 (0.091)
α_{panel}					-0.086 (0.345)
R^2	0.340	0.591	0.645	0.648	0.680

[a]significant at the 5% level, [b]significant at the 10% level

Table 4. Meta-significance Tests

	MST					
	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)	coefficient (s.e.)		coefficient (s.e.)
γ_0	-0.03 (0.403)	0.006 (0.007)	0.002 (0.008)	-0.001 (0.008)		0.000 (0.022)
γ_1	0.380[a] (0.08)	0.285 (0.077)	0.225[a] (0.052)	0.223[a] (0.052)		0.224[a] (0.052)
γ_{region}		-0.077[a] (0.037)	0.048[a] (0.007)	0.049[a] (0.007)		0.049[a] (0.007)
γ_{impact}			-0.030 (0.210)	0.013 (0.008)		0.013 (0.007)
γ_{public}				-0.030 (0.205)		-0.002 (0.022)
γ_{NHS}						-0.033 (0.211)
R^2	0.16	0.32	0.40	0.41		0.41
F-test	11.46	20.56	25.67	21.3	44.7	3.8

[a]significant at the 5% level, [b]significant at the 10% level