

The effect of health care expenditure on patient outcomes: Evidence from English neonatal care

Samuel I. Watson^a, Wiji Arulampalam^b, and Stavros Petrou^a on behalf of
NESCOP

^aWarwick Medical School, University of Warwick, Coventry, CV4 7AL, United Kingdom

^bDepartment of Economics, University of Warwick, Coventry, CV4 7AL, United Kingdom

Keywords: Neonatal healthcare, healthcare expenditure, cost-effectiveness threshold

Word count (body): 4,846

Figures: 0

Tables: 6

Running Header: Health Care Expenditure and Newborn Outcomes

Funding: The Neonatal Data Analysis Unit is supported in part by a Programme Grant for Applied Research (RP-PG-0707-10010) from the National Institute of Health Research (NIHR) and has previously received funding from the Department of Health and unrestricted funding from Danone, AbbVie, and Abbott International. SW received funding from Bliss and the Department of Economics, University of Warwick and is currently funded by NIHR Collaborations for Leadership in Applied Health Research and Care West Midlands. The study funders had no role in study design, collection, analysis, and interpretation of data, writing of the report, or in the decision to submit the article for publication.

Conflicts of Interest: None declared.

Corresponding Author: Samuel I. Watson, A-155 Warwick Medical School, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, United Kingdom. S.Watson.1@warwick.ac.uk

Abstract

The relationship between health care expenditure and health outcomes has been the subject of recent academic inquiry in order to inform cost-effectiveness thresholds for health technology assessment agencies. Previous studies in public health systems have relied upon data aggregated at the national or regional level however there remains debate about whether the supply side effect of changes to expenditure are identifiable using data at this level of aggregation. We use detailed patient data derived from electronic neonatal records across England along with routinely available cost data to estimate the effect of changes to patient expenditure on clinical health outcomes in a well-defined patient population. A panel of 32 neonatal intensive care units for the period 2009-2013 was constructed. Accounting for the potential endogeneity of expenditure a £100 increase in the cost per intensive care cot day (sample average cost: £1,127) leads to a reduction in the risk of mortality of 0.38 percentage points (sample average mortality: 11.0%) in neonatal intensive care. This translates into a cost per life saved in neonatal intensive care of approximately £420,000.

1 Introduction

Recent work in the United Kingdom and elsewhere has attempted to estimate the relationship between healthcare expenditure and health outcomes in order to inform cost-effectiveness thresholds for health technology assessment (HTA) (Claxton et al., 2015, Vallejo-Torres et al., 2016). Beyond the consideration about investment in new technologies and services, such studies are also important to answer the question of whether the benefits of increased medical expenditure in general are worth the additional cost. In this study, we estimate the returns to expenditure in neonatal intensive care using individual patient-level data, which allows us to identify the level and type of care an infant received. Previous studies have demonstrated clinically and economically significant benefits of healthcare expenditure on patient outcomes (Martin et al., 2008, Cutler et al., 2006, Luce et al., 2006, Stukel et al., 2005, Claxton et al., 2015) and specifically in neonatal healthcare (Almond et al., 2010). However, there remains significant variability between these estimates, which may be due to differences in the methods used, variations in expressions of the outputs of healthcare, and use of aggregate versus individual level data. As an example of the observed variation, estimates derived from different models of the cost per statistical life saved in neonatal care range from \$550,000 (approximately £330,000) (Almond et al., 2010) to £15 million (approximately \$25 million) (Claxton et al., 2015). Similarly significant variation exists in the estimates of the returns to medical expenditure between alternative programmes of care within the same study (Claxton et al., 2015).

It is perhaps unsurprising that such variation exists both within and between research projects. Even if each study presented precise and unbiased estimates of the effect of interest, there would be heterogeneity between patient populations and the portfolio of technologies and labour inputs available to them. This has implications for the adoption of new medical technologies and decisions to increase expenditure. For such decisions, the estimated incremental cost-effectiveness is typically compared to a cost-effectiveness threshold below which new technologies or policies are recommended for reimbursement. Many authors have argued that the cost-effectiveness threshold should reflect the average changes in patient health outcomes when the healthcare budget has been con-

tracted or expanded (Claxton et al., 2015). New technologies with an incremental cost-effectiveness below this threshold will be more cost-effective on average than those services displaced and thus improve overall healthcare efficiency. A better understanding of which services are displaced and whether that displacement is within or between programmes of care can further improve efficiency, and hence reliable evidence of the returns to expenditure for different care types is required.

For health care systems free at the point of use, such as the National Health Service (NHS) in England and Wales, where individual expenditure data is not available, other methods such as comparing aggregate expenditure and outcomes are used, as demonstrated by (Martin et al., 2008, Claxton et al., 2015). Such methods are being applied in other European countries and elsewhere. However, it is difficult to identify the supply side effect of shifts in the budget using data at the regional or national level as patient populations are not well defined between regions and over time. Indeed, the aforementioned study by Claxton et al. (2015) was subject to a substantial critique (Barnsley et al., 2013). In this study we present an alternative approach using a hospital-unit level analysis and routinely collected cost data. We take a well-defined patient population, infants receiving neonatal intensive care, with well-defined providers, neonatal intensive care units. Using a panel of these units we identify the supply side effect of shifts to the value of inputs to care. This study provides methods that can be used for other care types while providing new estimates of the return to medical expenditure for at-risk newborns.

2 Previous Literature

The relationship between health expenditure and patient outcomes is the subject of a large and growing literature (see Gallet and Doucouliagos (2016) for a recent review). In many instances the aim is to provide evidence towards an empirically derived cost-effectiveness threshold for HTA (Gallet and Doucouliagos, 2016, Vallejo-Torres et al., 2016). In England and Wales, Claxton et al. (2015), as well as the preceding work by Martin et al. (2008), estimated the value of the cost-effectiveness threshold in terms of the incremental cost per quality adjusted life year (QALY) gained

using NHS Programme Budgeting Data. These data provide estimates of the expenditure on 23 programmes of healthcare within each of 152 local healthcare authorities. Using a number of socioeconomic variables as instrumental variables for local healthcare authority expenditure, such as deprivation score and levels of unpaid care, to identify the effect of expenditure on mortality rates, the average estimated incremental cost per QALY gained was £12,936. With regards to outcomes among neonates, Claxton et al. (2015) combined maternity and neonatal programmes of care into one category and assigned deaths to it on the basis of International Classification of Diseases, Version 10 (ICD-10) codes recorded on death certificates. Various cross-sections of data from 2002 to 2008 were used. The authors state there were some issues assigning patients and deaths to programmes of care, for example, 94% of infant deaths (below one year of age) were re-assigned to programmes of care other than neonatal and maternity care even if they were treated in a neonatal unit. The estimate of the incremental cost per life saved was approximately £3.4 million and incremental cost per life year gained was approximately £45,000¹—significantly larger than most other programmes of care, across which the average figures were £140,000 and £12,000, respectively. In contrast, individual level neonatal data (used in this study) for 2009 to 2013 reveal that 98% of infants that died in neonatal units were recorded as having neonatal ICD-10 codes.

Almond et al. (2010) estimated the marginal return to medical expenditure in a neonatal health-care setting. Using a regression discontinuity design, exploiting a discontinuity around treatment provision to newborns either side of a 1,500g birth weight threshold, the authors estimated that the cost of saving the life of a newborn with a birth weight around 1,500g was approximately \$550,000 (in 2006 US\$). Nonetheless, this result is arguably not generalisable to the wider newborn population given that only 4.1% of the admissions to neonatal specialist care are for infants with birth weights within the bandwidth utilised in the study (data from the National Neonatal Research Database (NNRD; see Section 4.1); bandwidth 1,415-1,585g).

¹These are the figures using 2006 expenditure data with 2006/7/8 outcomes data. The authors consider data issues extensively and provide estimates adjusted for factors such as differential coverage of expenditure and outcomes data. The ‘coverage adjusted’ figures for incremental cost per life saved and incremental cost per life year gained were approximately £28 million and £30,000, respectively.

3 Theoretical Framework

The health outcomes of a patient treated in hospital is a function of the labour (L) and capital (K) inputs to their care along with underlying health at admission H . The labour and capital inputs are themselves functions of the patient's health and vary between hospitals (α) and over time (t):

$$health = f(g(L, K), H)$$

$$L = L(H, \alpha, t)$$

$$K = K(H, \alpha, t).$$

Interest lies in the identification of the effect of increasing labour and capital inputs to a given patient's care in terms of their health outcomes. Since these inputs are typically not observed, total expenditure is often used as an overall measure of the cost of inputs. Since patients in worse health are likely to receive greater inputs to their care, and the fact that health at admission cannot be fully observed, the use of expenditure data raises the issue of how to identify the supply side effects of inputs in this model. Similar econometric issues arise if patients are transferred to hospitals with greater inputs to care. We focus on a well-defined patient population whose care requirements are clearly specified and relatively homogeneous.

This study focuses on two health outcomes of babies admitted to neonatal intensive care. These are first, mortality and second, morbidity free survival. Neonatal care in England and Wales is categorised into three principle levels of care. These are, from most to least intensive: intensive care (IC), high dependency care (HDC), and special care (SC). Neonatal units are classified according to the levels of care they provide, with only the highest level units providing IC. Units are organised into networks, which are comprised of units of all levels, and patients are transferred to units that provide higher levels of care if required (Gale et al., 2012). Patients do not have a choice over provider. The levels of care are defined by the treatments that the infant received (British Association of Perinatal Medicine, 2011). As described below, a cot day at each level of care is a defined 'unit' of treatment with known labour and capital inputs. The costs of these inputs are estimated by health

care providers in the NHS and used to produce a national tariff. Thus, changes to labour and capital inputs to an IC cot day should be more broadly reflective of changes to labour supply, for example, rather than shifts in the patient population.

4 Data and Sample

4.1 The National Neonatal Research Database

The National Neonatal Research Database (NNRD) was created by the Neonatal Data Analysis Unit (NDAU), a research unit based at Imperial College, London, and was established in 2006. The data are extracted from the individual, electronic patient records of infants treated within neonatal units in England. The NDAU holds national research ethics committee approval to create this database (reference REC 10/H0803/151) as well as the permission from the Caldicott Guardians of each National Health Service (NHS) Trust. The data are psuedo-anonymised by removing patient and maternal identifiers and encrypting the NHS number of each infant. The data include a large range of variables including static descriptive variables captured once per baby, such as birth weight and gestational age at birth, episodically, such as episodes of infection and other clinical outcomes, and daily items such as treatments and procedures as well as level of care.

4.1.1 Sample

From the NNRD, data were extracted on all infants born and discharged or died between January 1st 2009 and December 31st 2013, and who received at least one intensive care cot day. We restrict our analysis to units designated to provide intensive care (neonatal intensive care units); intensive care provided by lower designated units is therefore temporary or for stabilising an infant prior to transport and is likely to differ from ‘proper’ intensive care.

4.2 NHS Reference Costs

The NHS Reference Costs data are estimates of the unit costs of providing healthcare across all Healthcare Resource Groups (HRGs) in England. There are over 1,400 HRGs, which are classifications used in the NHS to categorise patients who consume the same level of resources, for example, a neonatal intensive care cot day. The estimated cost of providing one ‘unit’ of an HRG for each provider, i.e. NHS Trusts or NHS Foundation Trusts, is available as part of these data. The Department of Health uses these reference costs to set prices for NHS funded services in England. The raw data at the provider level, used to calculate the average costs, were obtained for this study. These data were not adjusted for local price differences. Each provider estimates their cost for each HRG by taking into account fixed costs (e.g. depreciation), semi-fixed costs (e.g. nursing staff), and variable costs (e.g. drugs and consumables) (Monitor, 2014). These costs represent different levels of factor inputs to the production of a predefined ‘unit’ of care and therefore represent an estimate of $g(K, L)$. Costs for an IC cot day do not include special or extraordinary treatments such as surgery or transfers between neonatal units.

5 Empirical Framework

The framework discussed in Section 3 suggests the following individual level model. The model is specified at the individual level initially since data at this level enable us to identify which infants received intensive care, their outcomes, and various covariates predicting these outcomes.

Consider baby i admitted to the neonatal unit in hospital j in year t . Let y_{ijt} be the health outcome, x_{ijt} a vector of exogenous characteristics explaining infant health, and ex_{jt} the cost per cot day in that unit. α_j , δ_{jt} and τ_t are unit, unit-year, and year fixed effects, and u_{ijt} is an error term. The equation considered is the following multi-level specification:²

$$y_{ijt} = x'_{ijt}\beta + \gamma ex_{jt} + \alpha_j + \tau_t + u_{ijt}. \quad (1)$$

²Expenditure is specified in levels rather than natural logarithm both to facilitate calculation of the cost per life saved and as it is approximately normally distributed.

The parameter of interest to this study is the average marginal effect of health care expenditure on infant health outcomes, γ . The labour and capital input data we use is the typical neonatal intensive care cot day per hospital for the NHS Reference Costs data (as discussed above). We aggregate the above model to the level of the variable of interest; unit level outcomes are those of interest.³

At the aggregated level there is still concern that the average cost per IC cot day may be correlated with average unobserved differences in patient casemix. Sicker infants are transferred to higher level units if required but it is plausible that the system of transfers is used to move infants to units that are better resourced (Gale et al., 2012). While there is little evidence to suggest this is taking place, we also consider an instrumental variables approach to estimation, to consider the causal interpretation of our results. The location of a patient’s residence and the distance to the nearest hospital is likely to be conditionally independent of the characteristics of that hospital given local socioeconomic differences—neonatal units located in more deprived areas may be more likely to have a sicker casemix, since deprivation is correlated with infant health at birth, and therefore greater inputs to care. The patient is most likely to be treated in their nearest hospital. Thus, location provides a source of exogenous variation in hospital characteristics, conditional on local socio-economic factors. This approach has been used in previous studies (Gowrisankaran and Town, 1999, Geweke et al., 2003, Watson et al., 2014). In Section 3, which presents the theoretical framework, the instrument can be seen as providing exogenous variation in the unit providing treatment, α . The average IC care day costs of the nearest hospital are used as an instrument for the costs at the place of birth. The first stage is also aggregated.^{4,5}

We present results from both the fixed effects (FE) estimator (i.e. within estimator) and the fixed effects instrumental variable estimator (FE-IV). Standard errors are clustered at the unit level.

³Infant-level variables are aggregated by taking the arithmetic mean within unit-years.

⁴The nearest neonatal unit, of any level, is determined by straight-line distance.

⁵Hausman and Taylor (1981) proposed a two-step method for the estimation of instrumental variable models where the baby-level covariates are potentially correlated with unobserved effect. However, we assume the baby-level covariates are exogenous in this model. The moment conditions required for validity of the instrumental variable method are assumed to be satisfied in this model, which implies they are satisfied at the aggregate level. Aggregation has the additional benefit of reducing the total number of moment conditions, the large number of which may bias the estimator at the individual level (Newey and Smith, 2004).

5.1 Variable Definition

The outcome we consider is the in-hospital mortality rate defined as a percentage in between 0 and 100. As a secondary outcome we consider a measure of ‘morbidity free survival’. A common condition among neonates treated in IC is bronchopulmonary dysplasia (BPD), which is a chronic lung condition associated with prematurity although normal term infants can suffer from this as well (Ehrenkranz et al., 2005). We define morbidity free survival as 100 minus the percentage of infants who survive to hospital discharge without BPD — in this way the outcome is also ‘negative’ like the primary outcome. The secondary outcome does not capture the full range of morbidities that an infant may suffer from but provides an indication of the relationship between mortality and morbidity rates and inputs to IC. BPD is defined using data extracted from the NNRD.

The measure of expenditure is the cost per IC cot day (HRG ‘XA01Z’ above) in hundreds of pounds. A number of exogenous determinants of in-hospital mortality are included. These are widely used in similar models (Medlock et al., 2011). The clinical variables included are gestational age, birth weight z-score,⁶ and dummies for whether an infant’s mother received antenatal steroids and female sex. Dummies for region of residence are included along with indicators for quintile of socio-economic deprivation for the mother’s home area derived from the Index of Multiple Deprivation (IMD). We also include the local market forces factor (MFF) as a covariate. The MFF is estimated by the Department of Health and represents the unavoidable cost differences in providing healthcare between areas, such as the cost of capital or labour inputs (Monitor, 2013). Local price levels are also likely to be correlated with differences in local socio-economic conditions and hence infant health at birth. As a secondary analysis we re-estimate the model excluding the MFF and region and deprivation quintile indicators.

We consider a range of sensitivity analyses and checks for the robustness of the results to consider the effects of the various assumptions of the model and possible issues with the data. These are reported in Section 6.4.

⁶Birth-weight normalised within gestational age week.

6 Results

6.1 Summary Statistics

Table 1 provides summary statistics for the sample of Neonatal Intensive Care (level three) neonatal units included in this study. Overall, 32 tertiary level neonatal units were included, which treated 7,979 infants who received at least one day of intensive care and from whom data was extracted. The mean (SD) cost per IC cot day across the sample was £1,198 (£367), which after adjustment for the MFF and in 2009 GBP was £1,127 (£315).

6.2 Instrumental Variable First Stage Results

Overall, 64% of infants in the sample were admitted to their nearest neonatal unit, and of those that were not, 71% did not have a neonatal unit designated to provide IC as their nearest unit. 89% of infants were admitted to a neonatal unit designated to provide IC in their nearest neonatal network.

Table 2 reports the results from the first stage regression. An F-test of the instrumental variable in the first stage regression provides strong evidence that the instrument is strongly correlated with the cost variable. The estimated coefficient (standard error) was 1.11 (0.04), which supports the hypothesis that increases in the costs per cot day at the nearest neonatal unit are associated with increases in the cost per cot day at the hospital of birth. Indeed, the coefficient is greater than one indicating that infants, on average, either receive treatment at their nearest neonatal unit or a unit with greater costs per IC care day. There was little evidence to suggest that changes in the average composition of the *casemix*, as measured by various clinical variables, affected the cost per IC care day. For example, the estimated coefficient for average gestational age was -0.24 (0.30), which indicates that for every week reduction in average gestational age of the casemix the average cost per cot day increases by £24. However, this is neither economically nor statistically significant. Taken together this evidence may suggest that there is little correlation between unobserved casemix differences and cost per IC cot day.

6.3 Effect of Expenditure on Patient Outcomes

Table 3 reports the main results for infants who received at least one day of intensive care and who were inborn and treated in a neonatal intensive care unit with the proportion of those infants who died in-hospital as the outcome. As indicated in the table, results from different models using different controls are presented. Columns 1 and 2 do not include any controls. Addition of casemix (unit level averages of gestational age, birth weight z-score, female sex, multiple birth, and whether a full or partial course of antenatal steroids was administered) and year controls (column 3) shows evidence of a reduction in the mortality rate associated with increases to IC cot day costs suggesting units with higher risk patient casemixes have higher costs per IC cot day. The point estimate further increases in magnitude with addition of area and socio-economic controls (column 4), implying positive correlations between socio-economic characteristics of an area, the risk of mortality, and the costs per IC cot day. Instrumenting the expenditure data (column 5) makes little difference to the point estimate; the estimated coefficient is -0.38 (0.18) implying a £100 increase in the cost per cot day leads to a reduction in the mortality rate of 0.38 percentage points, *ceteris parabis*. This is approximately 3.5% of the overall mortality rate for this group of infants.

Table 4 reports the results with in-hospital mortality and/or bronchopulmonary dysplasia as the outcome. These results are intended to provide an indication of whether the reductions observed in the mortality rate also translate into reductions in morbidity free survival. The point estimates for the models including all the controls are comparably smaller in magnitude than the equivalent estimates for mortality despite 47% of the sample experiencing BPD and/or mortality. This suggests reductions in the mortality rate are accompanied by equivalent rises in morbidity.

6.4 Robustness and Sensitivity Analyses

A number of sensitivity analyses were conducted to examine the robustness of results to the assumptions of the model. Firstly, the panel used in the main specification was unbalanced due to changes over years in the neonatal units contributing data to the NNRD. The unbalanced panel may bias the estimators if units dropping out or joining the panel differ meaningfully from the other

units. We therefore re-estimated the model using a balanced panel. Secondly, the time variable of the panel data is financial years, therefore some unit-year combinations have very few infants contributing if a unit's participation varies by calendar year. We therefore re-estimated the model dropping unit-years with fewer than ten infants. Thirdly, it is possible that the errors are serially correlated, we therefore re-estimated using a first differences estimator. Fourthly, an inspection of the data (see Table 1) suggests that the year 2009/10 appears to differ from the other years in terms of case mix and outcomes. This could be due to temporal variations in definition of intensive care. We re-estimated excluding this year as a further sensitivity analysis. Fifthly, there is a possibility that selecting only infants who received IC may be endogenous if units vary their threshold to provide IC in response to changes in the cost per cot day. We re-estimated the model using infants born extremely preterm (<29 weeks gestation). While not all of these infants receive IC, a high proportion do (approximately 82% received IC in the NNRD data). Finally, we provide results from the individual level model specified in Equation (1) with unit-year average cost per cot day to examine the effects of disaggregation.

Tables 5 and 6 report the results from the various sensitivity analyses. In comparison to the primary result reported in column 5 of Table 3, Table 5 shows that the use of a balanced panel, dropping observations aggregated from the data for a small number (<50) of infants, and using a first differences estimator makes little qualitative difference to the results. Turning to Table 6, re-estimating the model using data from the financial year 2010/11 onwards only leads to a small increase in the magnitude of the point estimate, potentially providing evidence to support the hypothesis for a shift in the definition or provision of IC between 2009 and 2010. Selecting the sample based upon gestational age rather than IC leads to a reduction in the magnitude of the point estimates, which is unsurprising given approximately 20% of these infants did not receive IC and so would be unaffected directly by changes in the costs per IC cot day. Nevertheless, the estimated coefficients remain negative. Finally, columns 5 and 6 in Table 6 report the results from the individual level model. The estimate from the FE model is qualitatively similar to that from the equivalent aggregated model (column 4 of Table 3). However, the FE-IV estimate is quite different in magnitude

and with greater uncertainty, but is still negative in sign. These results are reported for comparison but should be interpreted with the caveats discussed in Section 5.

7 Cost Per Life Saved

In this section, we derive a cost per life saved from the preceding results. These calculations are based on our principal result reported in column 5 of Table 3. This result is interpreted as a causal effect resulting from exogenous increases to labour and capital inputs to an IC cot day. This includes, for example, improved staff to patient ratios, which we have previously shown to affect the mortality rate in neonatal intensive care (Watson et al., 2016).

The neonatal units in the sample provided 122,305 IC care days over the course of the panel (Table 1). It is estimated that a £100 increase in the cost per IC cot day results in a reduction in the mortality rate of 0.36 percentage points [95% CI: -0.66, -0.07]. A £100 increase in the cost per IC cot day would cost £12,230,500 and result in approximately 29 [6, 53] fewer deaths thus suggesting that the cost per infant life saved at the margin is approximately £421,740 [£230,760, £2,038,420].

To convert the above estimates into an incremental cost per life year gained, the incremental cost per life saved can be divided by the number of life years gained. There are no suitable data on the life expectancies of infants admitted to neonatal units given that the survival rate of very preterm and preterm infants in the relatively recent past was very low. One option is to use the average life expectancy at birth for the English population today, which is approximately 81 years (Office for National Statistics, 2013). However, infants born in poor health, such as those born at a low birth weight, have below average health, education, and labour market outcomes (Black et al., 2007), which may suggest a reduced life expectancy for these infants. However, the life expectancy of an infant born today, and particularly one who survives the neonatal period, may be in excess of 81 years given reductions in mortality rates and improvements to public health and medical care. We therefore utilise the average life expectancy of 81 years.⁷ This suggests a cost per life saved of

⁷This is also the strategy of Claxton et al. (2015) in their calculations.

£5,210 per life year gained. The standard discount rate for benefits used by NICE is 3.5% (National Institute for Health and Care Excellence, 2013). Using this rate gives an equivalent incremental cost per life year gained of £15,200.

These results are comparable to those from previous studies. Almond et al. (2010) estimated that for infants weighing approximately 1,500g the cost per life saved was \$550,000 in 2006 USD. This is qualitatively similar to the results from this study. Around 5% of the sample of infants in our study were of the birthweights reported in Almond et al. (2010). Our estimates are lower than those reported by Claxton et al. (2015) for neonatal and maternity care, who reported an incremental cost per life year gained of approximately £200,000 using data from 2008 to 2010. However, they examined mortality outcomes for all infants who received neonatal care using aggregated data, the majority of whom are at a very low risk of mortality. Increases in inputs to neonatal care among infants at very low risk of mortality will likely be targeted at reducing morbidity and improving quality of life outcomes.

8 Conclusions

This study has examined the effect of neonatal unit expenditure on an intensive care cot day on the risk of mortality for infants receiving intensive care. Using these results, the incremental cost per life saved in neonatal intensive care was estimated at approximately £420,000. Assuming a 3.5% discount rate, this translates into an incremental cost per life year gained of £15,200. These results are comparable to a previous analysis of the returns to medical spending in at-risk newborns (Almond et al., 2010). However, they contrast with the higher estimates determined using regional expenditure and mortality rates in England for neonatal and maternity care (Claxton et al., 2015). This may suggest that there is wide variation between programmes of care within the healthcare service and indeed within a particular programme of care. With respect to the determination of an appropriate cost-effectiveness threshold this implies that whether a new technology would improve efficiency is strongly dependent on where funding is displaced from. Furthermore, the returns to

medical expenditure in areas where there is a very low mortality rate cannot be judged without data on relevant morbidity and quality of life outcomes.

This study has used data routinely available in the NHS to estimate the returns to medical expenditure for a well-defined patient population to identify the supply side effects of shifts in health care expenditure. The method used in this study could be used across other types of care and in other patient populations for which patient data is available. This could provide important information for both cost-effective investment and disinvestment strategies within the healthcare service. However, the final estimates of incremental costs per life saved and life year gained are relatively crude calculations. Life expectancy and quality of life outcomes of infants who received intensive care is lacking. Further research is also required to estimate the benefits of increased expenditure to infants not at the margin for the risk of mortality. Furthermore, reductions in the risk of mortality can also be achieved by increasing the technical efficiency of units, such as by changing the choice of factor inputs or the manner in which those inputs are employed. Further research is required to elucidate this. In the absence of published data we have not been able to adjust the incremental cost per life year gained estimated for health related quality of life to convert the results into the incremental cost per QALY gained, which would have been more useful for comparative purposes.

References

- Almond, D., Doyle, Jr., J. J., Kowalski, A. E., and Williams, H. (2010). Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns *. *Quarterly Journal of Economics*, 125(2):591–634.
- Barnsley, P., Towse, A., Karlsberg Schaffer, S., and Sussex, J. (2013). Critique of the CHE Research Paper on NICE Thresholds.
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2007). From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *The Quarterly Journal of Economics*, 122(1):409–439.

British Association of Perinatal Medicine (2011). Categories of Care 2011.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P. C., and Sculpher, M. (2015). Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technology Assessment*, 19:1–504.

Cutler, D. M., Rosen, A. B., and Vijan, S. (2006). The value of medical spending in the United States, 1960-2000. *The New England journal of medicine*, 355(9):920–7.

Ehrenkranz, R. A., Walsh, M. C., Vohr, B. R., Jobe, A. H., Wright, L. L., Fanaroff, A. A., Wrage, L. A., and Poole, K. (2005). Validation of the National Institutes of Health consensus definition of bronchopulmonary dysplasia. *Pediatrics*, 116(6):1353–60.

Gale, C., Santhakumaran, S., Nagarajan, S., Statnikov, Y., and Modi, N. (2012). Impact of managed clinical networks on NHS specialist neonatal services in England: population based study. *Bmj*, 344(apr03 1):e2105–e2105.

Gallet, C. and Doucouliagos, C. (2016). The impact of healthcare spending on health outcomes: A meta-regression analysis.

Geweke, J., Gowrisankaran, G., and Town, R. J. (2003). Bayesian Inference for Hospital Quality in a Selection Model. *Econometrica*, 71(4):1215–1238.

Gowrisankaran, G. and Town, R. J. (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, 18(6):747–767.

Hausman, J. A. and Taylor, W. E. (1981). Panel Data and Unobservable Individual Effects. *Econometrica*, 49(6):1377–1398.

Luce, B. R., Mauskopf, J., Sloan, F. a., Ostermann, J., and Paramore, L. C. (2006). The return on investment in health care: from 1980 to 2000. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 9(3):146–56.

- Martin, S., Rice, N., and Smith, P. C. (2008). Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of health economics*, 27(4):826–42.
- Medlock, S., Ravelli, A. C. J., Tamminga, P., Mol, B. W. M., and Abu-Hanna, A. (2011). Prediction of mortality in very premature infants: a systematic review of prediction models. *PloS one*, 6(9):e23441.
- Monitor (2013). A guide to the Market Forces Factor. Technical report, NHS England, London.
- Monitor (2014). Approved Costing Guidance. Technical report, Monitor, London.
- National Institute for Health and Care Excellence (2013). Guide to the methods of technology appraisal 2013. Technical report, National Institute for Health and Care Excellence, London.
- Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.
- Office for National Statistics (2013). Historic and Projected Mortality Data from the Period and Cohort Life Tables, 2012-based, UK, 1981-2062. Technical report, Office for National Statistics.
- Stukel, T. A., Lucas, F. L., and Wennberg, D. E. (2005). Long-term Outcomes of Regional Variations in Intensity of Invasive vs Medical With Acute Myocardial Infarction. 293(11):1329–1337.
- Vallejo-Torres, L., García-Lorenzo, B., Castilla, I., Valcárcel-Nazco, C., García-Pérez, L., Linertová, R., Polentinos-Castro, E., and Serrano-Aguilar, P. (2016). On the Estimation of the Cost-Effectiveness Threshold: Why, What, How? *Value in Health*.
- Watson, S. I., Arulampalam, W., Petrou, S., Marlow, N., Morgan, A. S., Draper, E., and Modi, N. (2016). The effects of a one-to-one nurse-to-patient ratio on the mortality rate in neonatal intensive care: a retrospective, longitudinal, population-based study. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 101(3):F195–F200.
- Watson, S. I., Arulampalam, W., Petrou, S., Marlow, N., Morgan, a. S., Draper, E. S., Santhakumaran, S., and Modi, N. (2014). The effects of designation and volume of neonatal care on

mortality and morbidity outcomes of very preterm infants in England: retrospective population-based cohort study. *BMJ Open*, 4(7):e004856–e004856.

Conflicts of Interest

None declared.

Ethical Approval

This study was carried out with the National Neonatal Research Database (REC 10/H0803/151) with permission from NHS Trust Caldicott Guardians.

Funding

The Neonatal Data Analysis Unit is supported in part by a Programme Grant for Applied Research (RP-PG-0707-10010) from the National Institute of Health Research (NIHR) and has previously received funding from the Department of Health and unrestricted funding from Danone, AbbVie, and Abbott International. SW received funding from Bliss and the Department of Economics, University of Warwick and is currently funded by NIHR Collaborations for Leadership in Applied Health Research and Care West Midland. The study funders had no role in study design, collection, analysis, and interpretation of data, writing of the report, or in the decision to submit the article for publication.

Acknowledgements

The authors would like to thank all of the neonatal units and unit leads participating in the Neonatal Economic, Staffing, and Clinical Outcomes Project (NESCOP) and members of NESCOP

for insightful comments and analysis. The other members of NESOP are Neena Modi, Neil Marlow, Elizabeth Draper, Andrei Morgan, and Andy Cole. The authors wish to thank Bliss and the Department of Economics, University of Warwick for generous financial support. In addition, the authors are grateful for helpful comments from Matt Sutton and Anne Spencer, two anonymous referees, as well as participants at the European Health Economics Association (EuHEA) conference in 2014. The authors also wish to thank Eugene Statnikov, Shalini Santhakumaran, Daniel Gray, Ahmadur Rahman, Cheryl Battersby, and Nicola Fitz-Simon at the Neonatal Data Analysis Unit for valuable insight and advice as well as data management and processing. Finally, the authors would like to thank two anonymous referees for helpful and insightful comments.

Tables

Table 1: Summary statistics of the sample

Variable	Financial Year				Whole Sample
	2009/10	2010/11	2011/12	2012/13	
N_j	30	26	26	21	
N_i	1,378	2,406	2,752	1,443	7,979
Casemix					
Birth weight (g)	1,743 (422)	2,106 (208)	2,143 (212)	2,142 (247)	2,017 (340)
Gestational age (weeks)	31.3 (2.3)	33.1 (1.1)	33.2 (1.1)	33.3 (1.2)	32.7 (1.8)
% male	46.4	42.5	44.4	45.0	44.7
Mortality (%)	7.5	13.4	11.4	13.0	11.0
Mortality and/or BPD	58.7	43.6	41.4	38.3	46.5
Costs					
IC cot day cost (£)	1,145 (348)	1,257 (388)	1,151 (356)	1,264 (388)	1,198 (367)
IC cot day cost (£; MFF and inflation adjusted ^a)	1,126 (285)	1,218 (368)	1,093 (324)	1,057 (283)	1,127 (315)
Annual IC cot days (mean (SD))	1,922 (1,031)	2,037 (1,067)	2,199 (1,238)	2,420 (1,157)	2,122 (1,119)
IC cot days (sample total)	24,530	35,554	29,279	22,943	122,305
Balanced Panel					
N_j	15	15	15	15	
IC cot day cost (£)	1,147 (439)	1,335 (400)	1,226 (289)	1,298 (425)	1,252 (389)
IC cot day cost (£; MFF and inflation adjusted ^a)	1,117 (265)	1,277 (356)	1,146 (214)	1,090 (315)	1,158 (288)

N_j is the number of neonatal units and N_i is the number of infants.

Mortality is any in-hospital mortality.

Birth weight, gestational age, and unit costs are mean (SD) values.

^a These figures are adjusted for local market forces factor and are further adjusted to 2009/10 GBP using the Health Services Cost Index (HSCI).

Table 2: Estimated coefficients from first stage regression

Coefficient	Estimate	Standard error
Gestational Age (weeks)	-0.24	0.30
Full or partial course of antenatal steroids	0.05	2.17
Birth-weight z-score	-0.52	0.77
2009/10	0.29	0.26
2010/11	0.18	0.18
2012/13	2.62	4.02
Female sex (%)	-1.25	2.68
Multiple birth (%)	-7.72	6.33
MFF	-30.40	45.40
Instrument: Average IC cot day cost nearest hospital	1.11***	0.04
F-statistics of excluded instruments	784.04	
Kleibergen-Paap rk LM statistic	10.70	
R^2	0.96	
N	104	

Variables are aggregated to unit unit level by year by taking the arithmetic mean. Regressions also control for IMD quintile and region of residence of the infants.

p-value: * <0.05; ** <0.01; *** <0.001

Table 3: Estimated effect of increases to intensive care cot day costs on mortality

	(1)	(2)	(3)	(4)	(5)
	Mortality				
	OLS	FE	FE	FE	FE-IV
Expenditure	0.04	-0.00	-0.26	-0.38*	-0.36*
Standard error	(0.16)	(0.15)	(0.16)	(0.18)	(0.15)
Casemix controls	No	No	Yes	Yes	Yes
Year controls	No	No	Yes	Yes	Yes
Area and SE controls	No	No	No	Yes	Yes
Observations	104	104	104	104	104

OLS = ordinary least squares, FE = fixed effects, FE-IV = instrumental variable fixed effects

Casemix controls are averages of gestational age, birthweight z-score, female sex, multiple birth, and whether a full or partial course of antenatal steroids was administered. Year controls are dummies for year. Area and SE controls are proportions of infants whose maternal residence was in each deprivation quintile, proportions from each region, and the market forces factor.

p-value: * <0.05; ** <0.01; *** <0.001

Table 4: Estimated effect of increases to intensive care cot day costs on mortality and/or BPD

	(1)	(2)	(3)	(4)	(5)
	Mortality and/or bronchopulmonary dysplasia				
	OLS	FE	FE	FE	FE-IV
Expenditure	0.07	-0.30	-0.13	-0.24	-0.26
Standard error	(0.36)	(0.44)	(0.17)	(0.21)	(0.17)
Casemix controls	No	No	Yes	Yes	Yes
Year controls	No	No	Yes	Yes	Yes
Area and SE controls	No	No	No	Yes	Yes
Observations	104	104	104	104	104

OLS = ordinary least squares, FE = fixed effects, FE-IV = instrumental variable fixed effects.

Casemix controls are averages of gestational age, birthweight z-score, female sex, multiple birth, and whether a full or partial course of antenatal steroids was administered. Year controls are dummies for year. Area and SE controls are proportions of infants whose maternal residence was in each deprivation quintile, proportions from each region, and the market forces factor.

p-value: * <0.05; ** <0.01; *** <0.001

Table 5: Results from sensitivity analyses

	(1)	(2)	(3)	(4)	(5)	(6)
	Mortality					
	Balanced panel		Dropped few obs.		First differences	
	FE	FE-IV ^a	FE	FE-IV	FD	FD-IV
Expenditure	-0.44	-0.37*	-0.32	-0.36	-0.38*	-0.38*
Standard error	(0.27)	(0.16)	(0.26)	(0.21)	(0.19)	(0.15)
Observations	60	90	73	69	66	66

FE = fixed effects, FE-IV = instrumental variable fixed effects, FD = first difference.

Dropped few obs. = unit-years are dropped if fewer than 50 infants contribute to the unit-year.

Casemix controls are averages of gestational age, birthweight z-score, female sex, multiple birth, and whether a full or partial course of antenatal steroids was administered. Year controls are dummies for year. Area and SE controls are proportions of infants whose maternal residence was in each deprivation quintile, proportions from each region, and the market forces factor.

^a Balanced panel has insufficient clusters to calculate robust variance covariance matrix. Units with four or three years of observations used.

p-value: * <0.05; ** <0.01; *** <0.001

Table 6: Further results from sensitivity analyses

	(1)	(2)	(3)	(4)	(5)	(6)
	Mortality					
	Drop 2009/10			Extremely preterm		Individual level
	FE	FE-IV ^a	FE	FE-IV	FE	FE-IV
Expenditure	-0.46*	-0.45	-0.88	-0.57	-0.30	-0.08*
Standard error	0.20	0.66	0.74	0.66	0.16	0.04
Observations	73	67	103	103	8,786	8,786

FE = fixed effects, FE-IV = instrumental variable fixed effects, FD = first difference.

Drop 2009/10 = First year of panel dropped. Extremely preterm = sample of infants is those born at <29 weeks gestation. Individual level model = model re-estimated at the individual level.

Casemix controls are averages of gestational age, birthweight z-score, female sex, multiple birth, and whether a full or partial course of antenatal steroids was administered. Year controls are dummies for year. Area and SE controls are proportions of infants whose maternal residence was in each deprivation quintile, proportions from each region, and the market forces factor.

^a Balanced panel has insufficient clusters to calculate robust variance covariance matrix. Units with four or three years of observations used.

p-value: * <0.05; ** <0.01; *** <0.001