Decomposing the Effect of Workload on Patient Outcomes: An Empirical Analysis of a Maternity Unit

Michael Freeman
Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom mef35@cam.ac.uk

Nicos Savva
London Business School, Regent's Park, London NW1 4SA, United Kingdom nsavva@london.edu

Stefan Scholtes
Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom s.scholtes@jbs.cam.ac.uk

Extant literature provides strong evidence that service quality (e.g., patient outcomes) deteriorates at high levels of workload. In this paper, we use a detailed dataset from the delivery unit of a major teaching hospital to better understand *how* workload impacts quality. We consider two mechanisms: *direct* impact through deterioration in the quality of task execution and *indirect* through the choice of care pathway. First, we demonstrate that workload has an effect on the care pathway – discretionary but resource-intensive interventions, such as pain relief, are less likely to be administered at high workloads. Second, we show that workload increases the propensity of some adverse outcomes, such as perineal tears, and reduces others, such as emergency cesarean sections. The effect on perineal tears is predominantly direct, while the effect on cesarean sections is indirect. We also examine post-birth length of stay and find that the direct and indirect effects work in opposite directions; while both effects are individually significant, they cancel each other out on aggregate. Our results provide a better understanding of how workload affects quality in service settings and we demonstrate how they can be used to predict changes in outcomes and costs under different demand and staffing scenarios.

Key words: empirical; healthcare operations; service quality; workload management

---

1. Introduction

In recent years, hospitals have been under sustained pressure to reduce costs. In the US, for example, the 2010 Affordable Care Act contained specific measures aimed at slowing or reversing the growth of healthcare costs (ACA 2010). Similarly, the UK's National Health Service (NHS) was instructed to make savings of 4% per year over 2011-2014 (Hurst and Williams 2012). Since labor cost constitute more than half of hospital expenses (Guerin-Calvert 2011, Hurst and Williams 2012), the almost uniform response has been to cut front-line staff in hospitals and re-prioritize patient care so that only the most urgent cases receive hospital treatment. These structural changes continue to have a significant impact on the workload of individual hospital-based healthcare providers such as nurses and doctors.

It can be argued that running acute hospitals at high utilization is desirable as it improves efficiency. However, due to the stochastic nature of both emergency demand and service duration, hospitals running close to full capacity will occasionally have to deal with an unmanageably high level of workload. Hospitals use a number of strategies to cope with such periods of excessive demand, including ambulance diversions (Deo and Gurvich 2011), cancelations of elective surgical procedures (Nasr et al. 2004), premature discharges (KC and Terwiesch 2012), and admission to a less appropriate level of care (Chan et al. 2012). Nevertheless, there is empirical evidence to suggest that increased workload has an adverse impact on operational performance and service quality. For example, as workload increases the accuracy of hospital discharge coding deteriorates (Powell et al. 2012), nurse absenteeism rises (Green et al. 2013), and, perhaps most importantly, patient outcomes deteriorate (Needleman et al. 2011, Kuntz et al. 2014).

The above body of research treats the hospital environment largely as a "black box" and seeks to estimate the aggregate effect of system load, usually measured by bed occupancy, on outcomes such as mortality, readmission, length of stay, reimbursement, and absenteeism. While these studies provide evidence that workload is a quality-critical factor, they have done little to identify the mechanisms through which workload impacts outcomes. The goal of this study is to address this limitation by opening the "black box" and examining one mechanism through which workload affects outcomes: the chosen care pathway for a given condition.[1] More specifically, our aim is to first understand how workload affects the care pathway decisions made by providers and then to establish to what extent a workload-induced deterioration in outcomes is the result of a change in the quality of execution along the chosen care pathway – which we call the *direct* effect of workload – or due to a change in the care pathway itself – which we call the *indirect* effect of workload.

Our attempt to identify the mechanism by which workload affects outcomes, and specifically to disentangle the direct effect from the indirect effect, is econometrically challenging for two reasons. The first challenge has to do with the level of data granularity required. We require not only detailed staffing and patient-level outcome data but also information on the possible care pathways for a condition and the actual path taken in each case. This information is usually not available in the standard discharge records utilized in the literature. The second challenge stems from the fact that patients are not randomly assigned to care pathways. Instead, this decision takes into account patients' medical history, complications, and comorbidities (some of which may not be observable to us as researchers). Therefore, patients who may be considered more likely to

[1] Care pathways are discrete treatment paths distinguished by the type of care provided and the amount of resources dedicated to patient care.

experience a particular outcome may also be more likely to be sent down a particular care pathway. This endogeneity complicates our attempt to identify the impact of the choice of care pathway on outcomes, and through that the indirect effect of workload.

To overcome the first challenge, we assembled a detailed dataset from a maternity unit of a large UK-based teaching hospital over a four-year period, amounting to over 18,000 births. The delivery unit (DU), which we describe in more detail in §3, offers several econometric advantages. First, once we exclude scheduled cesarean sections (C-sections), the arrival process is random (homogeneous Poisson), and since women in labor are emergent patients, they cannot be selectively canceled or re-prioritized. This provides statistically exploitable exogenous variation in workload. Second, the DU offers a set of well-defined and easy-to-observe care pathways that differ substantially in terms of the resources required for patient care. The least resource-intensive pathway is a natural birth. Interventions that are more resource intensive include pain relief, such as epidural analgesia, and instrumental deliveries. Third, the data contains a number of well-defined outcome measures that are frequently used for cross-hospital comparisons. For our study we use three outcomes, two of which are clinical and one operational: third- or fourth-degree perineal tears, emergency C-sections,[2] and post-birth length of stay (LOS).

To address the second challenge, we specify and estimate econometric models that account and correct for non-random assignment: recursive bivariate probit, Heckman treatment effect, and endogenous switching regression (Tobit-5). These models estimate simultaneously (i) the propensity of a patient being assigned to a specific care pathway and (ii) the outcome given the selected care pathway. The simultaneous estimation takes into account that the error terms might be correlated, i.e., that there exist unobservable factors that affect both the assignment and the outcome. To estimate these models consistently it is desirable to identify at least one variable that affects the selection equation but has no direct effect on the outcome equation: a condition known as "exclusion restriction" (Wooldridge 2002). We identify such a variable – the average operating theater use in the hours prior to delivery – which takes advantage of the fact that decisions that affect the care pathway are typically separated by a significant time lag from outcomes.

We obtain two main results. First, workload has a pronounced impact on the care pathway. Women delivering when the DU is at the 90th percentile of workload are 6.3% less likely to receive an intervention than those at the 10th percentile. Decomposing this further we find that this workload effect is concentrated at discretionary decision points, such as whether to administer

[2] An emergency C-section could be viewed as an intervention rather than an outcome. In this analysis we use the term intervention for less severe treatments such as epidural analgesia and instrumental delivery. This is discussed further in §3.

epidural analgesia for pain relief; interventions for which there are strong clinical indications, such as instrumental deliveries, are not significantly affected by workload.

Second, the impact of workload on outcomes is also significant but more nuanced.

1. In the case of perineal tears, we find that workload has a positive and significant effect. We observe a 20.1% increase in the incidence of tears at the 90th percentile of workload as compared to the 10th percentile. Furthermore, the effect is primarily direct, i.e., due to midwives having to look after more patients simultaneously rather than a change in care pathway.

2. The impact of workload on the incidence of emergency C-sections is negative, suggesting that at high workload fewer emergency C-sections are performed. At the 90th percentile of workload, an emergency C-section is 2.5% less likely to occur than at the 10th percentile. We find that this effect is predominantly indirect, i.e., is a result of a change in the care pathway.

3. The impact of workload on post-birth LOS is, on average, statistically indistinguishable from zero. Nevertheless, our analysis suggests that workload does have a positive direct effect, but only for women who do not receive an epidural: Those delivering at the 90th percentile of workload stay 4.4% longer than those at the 10th percentile. Furthermore, workload has a negative indirect effect on post-birth LOS: Women who receive an epidural stay 44.3% longer and workload reduces the epidural rate.

These results have important managerial implications. First, we find that the impact of workload is easily misunderstood if not properly decomposed into direct and indirect effects. This is most obvious in the case of post-birth LOS, where one might erroneously conclude that workload does not matter. While this is true on average, workload does matter for some patients. In particular, managerial interventions that target the positive effect of increased workload on LOS for those patients that do not receive epidural analgesia would lead to an overall reduction in LOS. Furthermore, without decomposition it could be understood that there is a workload-related reduction in the provision of emergency C-sections. Based on this, recommendations to management might include examining the robustness of the emergency C-section decision-making process and adding C-section capacity to avoid this apparent rationing at busy periods. These recommendations would be incorrect. The decomposition shows that the workload-related reduction in the incidence of emergency C-sections is indirect, i.e., is caused by a reduction in more benign interventions, such as epidural analgesia, made earlier on in the care process.

Second, our findings can be used to complement capacity planning by predicting how outcomes, as well as the cost of patient care, are likely to change with different staffing levels or demand scenarios (for an example in a similar context see Green and Liu 2013). For instance, through a counterfactual analysis we demonstrate that if the hospital we study increased staffing to provide the desired one-to-one care 95% of the time, up from the 30% currently achieved, then the epidural rate would increase by 4.6% leading to a 1.7% increase in the incidence of emergency C-sections, while the rate of perineal tearing would decrease by 10.8%. Using recent cost figures (Curtis 2012, NICE 2012), this increase in staffing, which would cost the hospital approximately £1,460,000 p.a. to implement, would also lead to a net *increase* in the cost of patient care by approximately £70,000 p.a. (or 4.8% of the additional staffing cost), due mainly to the increase in the number of emergency C-sections.

In summary, opening the organizational "black box" and studying workload effects at the level of a sequential process with appropriate econometric methods has uncovered a surprising richness of phenomena. Beside the insights we obtain in our context, our methodology is generally applicable to sequential processes in healthcare and beyond and can help managers identify where "quality" bottlenecks occur.

2. Literature Review

Our research contributes to the literature on the effect of system load on system performance in contexts where servers have some discretion over service provision. Oliva and Sterman (2001) were among the first to systematically study how workers begin to cut corners when workload pressure increases. Whether and how this happens matters because systems with discretionary completion times can exhibit surprising and even paradoxical effects. Hopp et al. (2007), for example, demonstrate that in such systems an increase in capacity can lead to an increase in waiting times. Much of the subsequent empirical literature has focused on healthcare systems, where data is increasingly available and the implications, specifically in terms of morbidity or mortality, can be particularly serious. In the context of patient transport, KC and Terwiesch (2009) show that in the short term the system responds to an increase in workload by reducing service times, while in the longer term a sustained increase in load has a detrimental effect. By contrast, Berry-Jaeker and Tucker (2013) show that in the context of in-patient care, workload can prolong service times and increase patient LOS and that there are spillover effects across patient types. In the context of emergency care, Batt and Terwiesch (2012) show that simultaneous speed-up and slow-down mechanisms come into play as workload changes, with task reduction being counterbalanced by a general slowdown in common treatment processes. These studies provide clear evidence that service times change in response to workload, albeit that the direction of the change depends on the context.

In addition to service times, researchers have also studied the relationship between workload and other operational, financial, and service quality metrics. Green et al. (2013) show that nurse absenteeism rates are linked to anticipated workload. Powell et al. (2012) find that patients discharged at higher workload levels are less likely to be classified as "high severity" for reimbursement purposes, which can have a significant detrimental effect on a department's revenues. In addition, several studies have focused on the impact of workload on quality of care. KC and Terwiesch (2009) estimate the effect of workload on the mortality of cardiothoracic patients and find that load, as measured by bed occupancy, has no significant effect, while long periods of increased load are correlated with higher mortality rates. Kuntz et al. (2014) suggest that the lack of an occupancy effect in KC and Terwiesch (2009) may be due to the assumed linear relationship and argue that organizational workload affects mortality in a non-linear fashion. Using a large data set of 85 German hospitals, they provide evidence for a tipping point effect, where mortality is unaffected by occupancy variation until occupancy crosses a threshold, beyond which it affects mortality significantly. A similar effect of workload has been identified by Tan and Netessine (2012), who find a non-linear effect between the number of diners assigned to waiting staff and staff sales performance in the context of a restaurant chain. In this case, sales initially increase with load as staff become more motivated but ultimately decline as staff become overworked.

These workload studies in the operations literature are complemented by several studies in the medical literature. Needleman et al. (2002) find that higher nurse staffing levels can reduce the prevalence of poor nurse-sensitive patient outcomes such as urinary tract infections and "failures to rescue." In a study of 232 Californian acute care hospitals, Cho et al. (2003) calculate that a one-unit increase in the ratio of registered nurses per patient-day can reduce the odds of acquiring pneumonia by 8.9% among surgical patients. More significantly, higher load has been consistently shown to increase the mortality rate (see e.g., the review in Kane et al. 2007). In one of the most comprehensive studies to date and using a sample of nearly 200,000 patient records, Needleman et al. (2011) find a significantly increased mortality risk when staffing levels fall below their acuity-adjusted targets and when patient turnover is high.

The aforementioned literature has focused primarily on extricating the aggregate effect of workload on services and outcomes without explicitly distinguishing between the exact mechanisms behind the effect. The exception to this is recent work on intensive care units (ICUs) by Chan et al. (2012) and KC and Terwiesch (2012). At higher levels of ICU occupancy the former study finds that there is a decrease in the chance of ICU admission, while the latter shows that there is an increase in the chance that a patient will be discharged early. Together these papers indicate that workload affects patient physical routing decisions and that this has an adverse effect on patient outcomes; e.g., they find that re-routed patients are more likely to require readmission to the ICU.

In summary, our work integrates two streams of research: work on the aggregate effect of workload on outcomes and research on the effect of workload on decisions regarding the type of care a patient receives. Specifically, we propose a methodology that allows us to study how workload affects outcomes through its effect on both the decision of which care pathway to send a patient down and the quality of execution along the chosen care pathway.

3. Clinical Setting

The setting for this study is the delivery unit (DU) in the maternity department of a large UK teaching hospital. The DU is the primary location for childbirth and immediate post-natal care and is made up of standard delivery rooms, clinical rooms for higher-risk patients, obstetric theaters, and a recovery bay. It sits within a typical maternity hospital that also contains an ante-natal unit, which provides care for patients with problematic pregnancies, a post-natal unit, in which women and babies are observed in the period post-birth but before discharge, and a neonatal unit, which specializes in providing additional care for babies (DH 2013). This setting is important as childbirth is the most common cause of hospital admission and accounts for 2.8% of all healthcare expenditure in the UK (NAO 2013).

The DU deals essentially with two types of patients: scheduled and unscheduled. Scheduled patients, who make up 14.6% of the total observations, are those admitted for an elective C-section. For these patients the date of delivery is pre-booked and the care pathway is locked-in in advance. Elective C-sections are performed in a dedicated operating theater, attached to the DU, by a dedicated team of specialists. As a result, these elective procedures are not likely to be affected by workload conditions at the DU. We therefore exclude these deliveries from our main sample. Unscheduled patients, who make up the remaining 85.4% of total observations, consist of those who either anticipate a natural birth or have arrived before their scheduled elective C-section (e.g., because of the early onset of labor). For these patients the arrival date cannot be accurately predicted and the care pathway is not fixed in advance. The unpredictable nature of these births, as evident from Figure 1, makes the arrival process largely random and uncontrollable, without any significant seasonal or daily variations. This unpredictability, which also makes it difficult to plan staffing levels, gives rise to econometrically useful variation in workload.

Setting aside elective C-sections, there are three main routes into the DU: (i) directly if in labor on arrival, (ii) transfer from the ante-natal unit, and (iii) transfer from a waiting lounge. The main route out of the DU is into the post-natal unit. This routing is illustrated in Figure 2. When an unscheduled patient arrives at the DU she is assigned a primary midwife who is responsible for her well-being, as well as that of the baby, throughout labor and childbirth. The midwife must attend the patient regularly in order to observe the frequency of contractions, monitor the fetal and maternal heart rate, record temperature and blood pressure, determine whether there is need for a doctor to intervene, and perform other related activities. For an uncomplicated birth, the midwife will also perform the delivery, carry out an initial examination of the baby, and provide the immediate post-natal care for the mother.

Depending on individual patient circumstances, there are a range of interventions that can be used. The most common of these are induction of labor, epidural analgesia, and instrumental

Figure 1 **Number of births by hour of day and month of year (mean with 95% confidence intervals and 25th and 75th percentiles), and time series of number of births per day in 2011.**

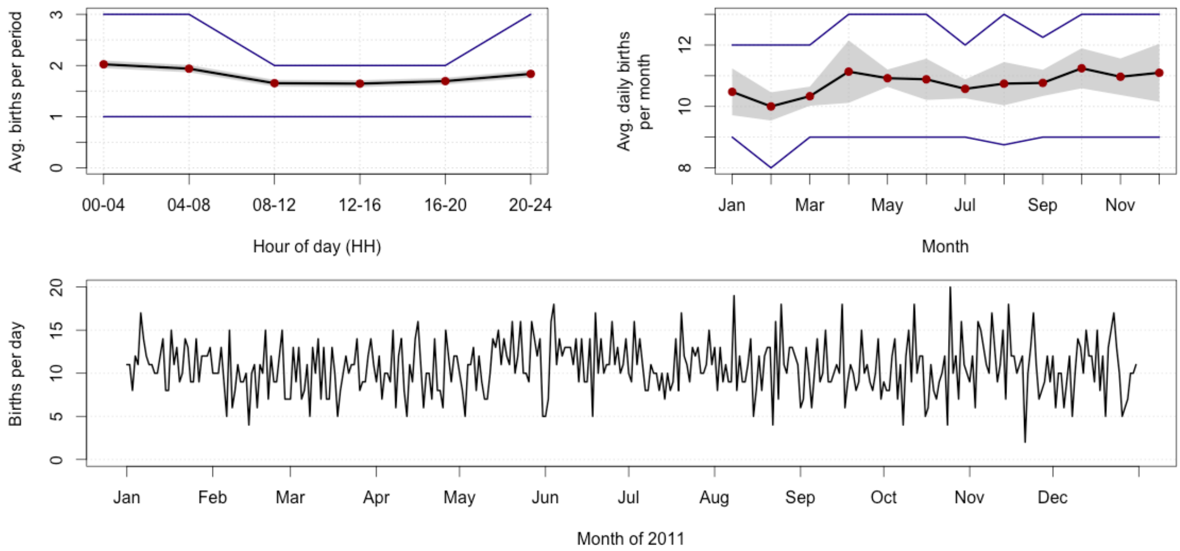
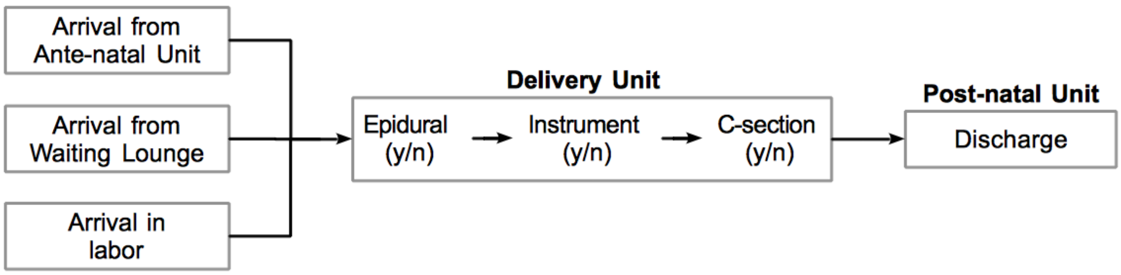


Figure 2 **Standard patient flow through the maternity hospital.**



delivery. A patient that experiences any of these procedures is said to have deviated from the "normal delivery" pathway and outcomes for births with intervention are frequently compared to those for natural births (MCWP 2007).

Induction of labor is the use of a vaginally administered drug to start contractions. It is most commonly performed when the baby is overdue, although other factors, such as maternal health or waters breaking early, may indicate induction. The decision to induce is almost always taken in advance of the patient arriving at the hospital and is performed in the ante-natal unit, with the patient being transferred to the DU when labor begins. This process typically takes more than 10 hours (Griffiths et al. 2002), meaning that DU workload conditions are unlikely to influence the induction decision. We note that inductions are occasionally canceled due to workload-related pressure; however, this is due mainly to a lack of available beds in the ante-natal unit rather than DU workload. For this reason, we control for induction in our empirical analysis but do not explicitly examine whether the induction rate is affected by DU workload.

Epidural analgesia is usually administered to improve patient experience when gas and other painkillers provide insufficient pain relief. It involves the injection of painkilling drugs into the lower back, which aims to block the nerves and reduce or eliminate labor pain. This form of intervention is highly discretionary; it is typically given no later than two hours before delivery and must be administered by an anesthesiologist at the recommendation of an obstetric doctor, who assesses suitability based on the progress of labor and any presence of contraindications. The procedure takes approximately 20 minutes to perform, and post-provision there must be a midwife with the patient at all times in order to take her blood pressure at regular intervals and to monitor the baby's heart rate (continuously for at least 30 minutes) to ensure that no complications arise. The need for specialist doctors and post-procedure supervision makes an epidural highly resource-intensive. From a medical perspective, an epidural can also have disadvantages, such as reducing maternal blood pressure (which may affect the flow of oxygen to the baby), the potential for drugs to cross the placenta (which can affect the baby's breathing and cause drowsiness), slower labor, and increased risk of further interventions such as augmentation (speeding up of labor) and instrumental delivery or an emergency C-section (Anim-Somuah et al. 2011).

Instrumental deliveries and/or emergency C-sections are carried out if labor is prolonged or if there is an observed deterioration in the health of the mother or baby that indicates that a natural birth is unlikely. For an instrumental delivery the baby is delivered manually using instruments such as forceps or a vacuum pump. The decision to assist will be made either on arrival at the DU or at some point during labor when medical circumstances demand it. The intervention itself is carried out by an obstetric doctor usually in the operating theater (although instrumental deliveries may occasionally take place in the delivery room). The procedure takes on average 45 minutes to perform and requires the presence of at least one midwife and, on occasion, a team of specialists, including an anesthesiologist and a pediatrician.

Emergency C-sections are performed when the physician believes that the delivery will not occur vaginally without placing the woman or baby under undue risk. An emergency C-section is clinically essential in the light of the perceived risk to the patient's health, which distinguishes it from an instrumental delivery. Emergency C-sections are major surgeries carried out under regional or, occasionally, local anesthetic and take 30-45 minutes to perform. Furthermore, they are one of the most resource-intensive and costly processes in the DU; they carry significant risk for the patient – such as hemorrhage, infection, thrombosis, and an increased risk of complications in subsequent pregnancies (Bragg et al. 2010) – and have a significant effect on both post-birth recovery times and patient experience (Henderson et al. 2001). After delivery, women who have had an instrumental delivery or emergency C-section remain in the recovery bay for a short time to be looked after by specialist nurses before being transferred to the post-natal unit.

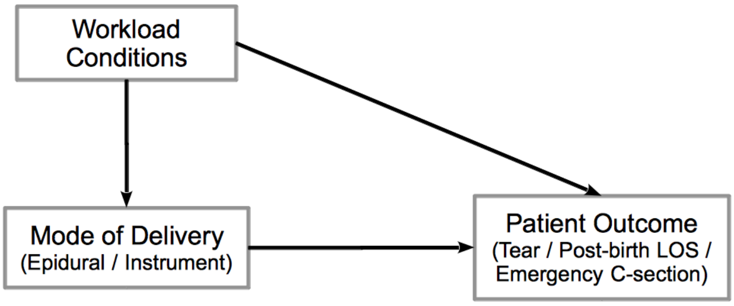
In addition to well-defined interventions (which constitute the care pathway), we can also observe a number of well-defined outcome measures. For our study we use two clinical outcomes, (i) severe (third- or fourth-degree) perineal tears and (ii) emergency C-sections, as well as one operational outcome: (iii) post-birth LOS. The rate of perineal tearing is commonly used as an indicator for cross-hospital comparison and features on the list of 26 hospital safety indicators compiled by the US Department of Health & Human Services Agency for Healthcare Research and Quality (AHRQ 2013). There is also extensive evidence on the associated risk factors (see e.g., Handa et al. 2001, de Leeuw et al. 2001, Eskandar and Shet 2009). Our second outcome metric is delivery by emergency C-section. In contrast to instrumental delivery, which we consider to be a form of intervention, we consider emergency C-section as an outcome in its own right. The considerable variation between emergency C-section rates across hospitals, even after patient risk-adjustment (Bragg et al. 2010), suggests that organizational factors, such as workload, may explain the occurrence of this outcome. Our third outcome metric, post-birth LOS, is an operational measure of quality of care that captures the aggregate effect (if any) that DU workload has on the utilization of hospital resources further downstream in the post-natal unit.

Finally, we note that hospital staff in the UK that work for the NHS receive a fixed-rate salary; remuneration is not linked to performance or results (either clinical nor financial). Therefore, staff have no personal financial incentive to recommend or warn away from any particular course of treatment on the grounds of cost (Lilley 2003).

4. Hypothesis Development

For a healthy woman and baby, a natural, unmedicated birth is the most desirable outcome. Since the arrival process for high-risk women and babies is exogenous, workload conditions in the DU would be expected to have no impact on the rate of natural births. However, in practice DU workload may affect the choice of care pathway – natural or interventional birth – in two ways: First, at low-workload conditions there may be an over-treatment effect, with patients receiving unnecessary intervention as a result of excess resource availability. Simply put, unoccupied staff may create work to fill their time: an idea that gives rise to the idiom “the devil makes work for idle hands” (Tan and Netessine 2012). Moreover, low-workload conditions may result in over-monitoring, leading to false alarms and a higher likelihood of adopting an “act-early” rather than a “wait-and-see” approach, which triggers unnecessary medical interventions. Second, at high workload levels resources become stretched and so rationing must occur in order to cope with the increased demand. For example, more time-consuming treatment options may become infeasible to administer due to a shortage of staff. Combined, these two effects suggest that interventions are less likely to occur as workload increases, leading to our first hypothesis:

Figure 3 **Hypothesized effect of workload on patient outcomes.**



(H1a) As workload increases the likelihood that a patient will follow a less resource-intensive care pathway, such as a natural birth, increases.

The main interventions in a DU (induction, epidural analgesia, instrumental delivery, and emergency C-section) are described in §3. To examine the workload effect on processes more closely, we focus specifically on epidural analgesia and instrumental delivery.[3] From hereon in, the term "intervention" refers solely to these two procedures. An intervention is highly resource-intensive and places additional demands on midwives' time in the form of additional monitoring and one-to-one care. At the same time, there is a degree of flexibility and discretion in the decision to intervene. Epidural analgesia is administered primarily for comfort purposes and is not typically indicated by clinical concerns over the health of the woman and baby. At times of low workload, the option to have an epidural is made more available, which may result in higher uptake by patients. Conversely, although the decision to deliver instrumentally is not always clear-cut, it is less discretionary. Indications for instrumental delivery are outlined in clinical guidelines and include fetal compromise, medical conditions (e.g., cardiac disease) that require a shortening of the second stage of labor, and inadequate progress based on clearly defined progress times (RCOG 2011). As a result of the discretionary element, we anticipate that the decision to administer an epidural is affected more by workload than the decision to deliver instrumentally.

(H1b) Workload affects largely discretionary interventions, such as epidural analgesia, more than less discretionary interventions, such as instrumental delivery.

Hypotheses 1a and 1b are concerned with workload effects on pathway decisions – whether or not to intervene in the natural birth process. We next consider the effect of workload on three outcomes: perineal tears, emergency C-sections, and post-birth LOS. The potential adverse effect of workload on outcomes is discussed in a series of papers (Needleman et al. 2011, KC and Terwiesch 2009, Kuntz et al. 2014). In contrast to these papers, we will consider *how* workload affects outcomes.

[3] This is due to the fact that DU workload is not expected to affect or be affected by inductions and emergency C-sections are considered an outcome, as discussed in §3.

Specifically, we wish to understand whether a workload effect on outcomes is due primarily to a deterioration in the quality of execution over a patient episode or to a discrete decision made during care delivery, i.e., the chosen care pathway.

Since we posit in Hypothesis 1b that workload affects the rate of epidural analgesia more than that of instrumental delivery, and since the epidural decision has been identified by midwives in the DU as the critical decision that "set women off on a path to further intervention," we focus specifically on the effect of epidural analgesia on outcomes. We argue that the incidence of perineal tearing and emergency C-sections is affected differently by workload. Consider the path diagram shown in Figure 3. We claim that the effect of workload on the incidence of tearing is positive and predominantly direct, i.e., increased incidence is the result of deteriorating quality of execution brought about by higher workload. Precautionary action against tearing is difficult owing to the challenge in predicting if and how it will occur prior to labor (Williams et al. 2005); however, there is evidence that appropriate midwifery care during labor can reduce the likelihood of a tear (Aasheim et al. 2011). Furthermore, medical literature on the effect of epidural analgesia on the propensity of tearing is sparse and the findings are mixed (Albers et al. 2007, Donnelly et al. 1988). This leads us to our next hypothesis:

(H2a) Higher workload increases the propensity of perineal tears. The effect is primarily direct, i.e., related primarily to deteriorating quality of execution brought about by increased workload, rather than indirect.

Emergency C-section rates are less likely to be directly affected by DU workload for two reasons: First, senior obstetricians are involved in the decision-making process for an emergency C-section; and second, an emergency C-section is a serious intervention in response to severely compromising or potentially life-threatening clinical indications (NICE 2012). However, it is possible that an earlier decision as to whether to administer an epidural will affect the likelihood of an emergency C-section: epidural analgesia slows down contractions and prolongs labor, which may cause additional distress for the woman or baby and lead to an emergency C-section. In addition, women who receive epidural analgesia are connected to a fetal heart rate monitor, which has been shown to increase the likelihood of an emergency C-section (Alfirevic et al. 2013). Therefore, in contrast to perineal tears, the effect of workload on the rate of emergency C-sections is hypothesized to be largely indirect. Since we have already hypothesized that the epidural propensity is reduced with increased workload, we can infer that workload has a negative indirect effect on the propensity of emergency C-sections due to the presumed increase in the likelihood of further escalation that epidural analgesia brings.

(H2b) Higher workload reduces the propensity of emergency C-sections. The effect is primarily indirect, i.e., related to the earlier decision to administer epidural analgesia, rather than direct.

Turning to the impact of DU workload on post-birth LOS, we expect that workload will have a positive direct effect. As argued in Hypothesis 2a, higher workload leads to a deterioration in the quality of execution, which manifests as an increase in the propensity of complications such as tears. A woman who experiences such complications will require a longer LOS in the post-natal unit to recover, which leads us to our next hypothesis:

(H3a) The direct effect of an increase in DU workload is an increase in post-birth LOS, caused by deterioration in the quality of execution.

In addition to the positive direct effect, we also expect DU workload to have a negative indirect effect. As argued in Hypotheses 1a and 1b, higher workload reduces the rate of interventions, particularly the administering of epidural analgesia. Women who receive such interventions often experience side effects, such as nausea, vomiting, urinary retention, and respiratory depression, and are also more likely to receive an emergency C-section (Anim-Somuah et al. 2011). Since these complications will prolong post-birth LOS, an increase in DU workload will lead to an indirect reduction in post-birth LOS.

(H3b) The indirect effect of an increase in DU workload is a decrease in post-birth LOS, caused by a reduction in the number of epidurals administered.

It is not possible to hypothesize whether the direct effect of Hypothesis 3a or the indirect effect of Hypothesis 3b dominates.

5. Data and Variable Description

Our primary data set contains information on all births that occurred in the hospital we study between January 1, 2007 and March 1, 2012. In total 18,128 births occurred in the DU during our observation period. For each patient we have information on (i) arrival and departure times and time stamps for any transfers between units, (ii) pregnancy-related diagnoses, classified according to the WHO's International Classification of Diseases ICD-10 (WHO 2011), and (iii) the procedures performed, classified according to the Classification of Interventions and Procedures OPSC-4.6 (HSC 2013), the UK equivalent of the American Medical Association's CPT coding system (AMA 2013). The ICD-10 data provides a detailed overview of any health-related conditions that the patient was diagnosed with prior to or during their stay. In addition to serving as indicators and contra-indicators for the choice of procedures, these factors allow us to account for much of the across-patient heterogeneity. The OPSC-4.6 codes provide information on a range of interventions, such as inductions and epidural analgesia, and the delivery method. On the staffing side, we have data from April 1, 2008 onwards detailing every shift worked by midwives, nurses, and maternity care assistants in the maternity hospital during this period. Therefore, the time window for our analysis is April 1, 2008 to February 1, 2012 (the last month of data is discarded to avoid problems

Table 1 Descriptive Statistics and Correlation Table

	Descriptive statistics				Correlation table					
Variable	Mean	SD	Min	Max	(2)	(3)	(4)	(5)	(6)	(7)
(1) Workload	1.18	0.37	0.22	3.28	-0.032***	-0.036***	-0.005	0.022*	-0.011	-0.006
(2) Intervention	0.43	–	0	1		0.855***	0.556***	0.031***	0.041***	0.109***
(3) Epidural analgesia	0.36	–	0	1			0.255***	-0.011	0.121***	0.122***
(4) Instrumental delivery	0.19	–	0	1				0.072***	-0.244***	0.036***
(5) Perineal tear	0.05	–	0	1					–	0.024**
(6) Emergency C-section	0.20	–	0	1						0.345***
(7) Post-birth LOS (hours)	44.0	41.3	0.4	578.7						

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

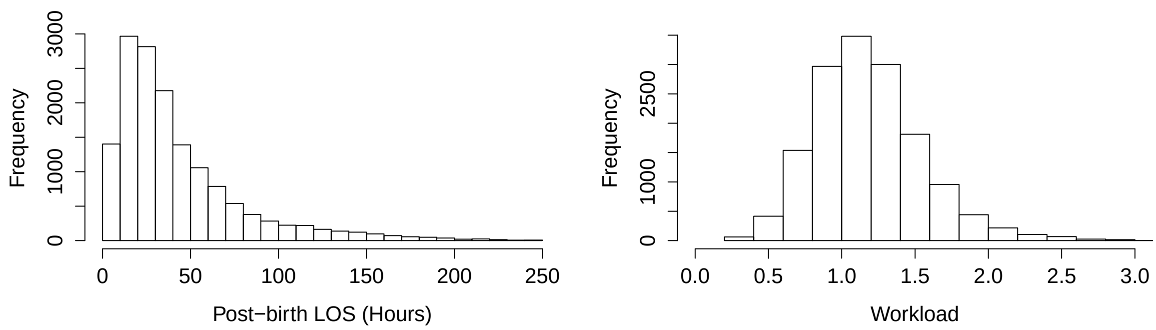
associated with, for example, delayed data entry). Together the data sources give us a detailed insight into the conditions surrounding each birth.

Our goal is to investigate the impact of workload on outcomes through its effect on both the choice of care pathway and the quality of execution along the chosen care pathway. To identify changes to the care pathway, we study natural vs. interventional ($INTERV$) births. We define an interventional birth as one in which epidural analgesia (EPI) and/or an instrumental delivery ($INSTR$) was used. As reported in Table 1, approximately 9 in 20 births was interventional, with 7 in 20 births involving an epidural and 4 in 20 requiring an instrumental delivery.

The first outcome variable of interest is the occurrence of perineal tearing ($TEAR$). Cases of tearing were identified using any of the following: ICD-10 codes O70.2-3 and O71.3-5 or OPSC-4 codes R32.1-2 and R32.8 with Z42.1. Since emergency C-section patients do not experience vaginal delivery or, subsequently, perineal tearing, we exclude them from the $TEAR$ analysis, leaving 12,080 observations. Approximately 5% of these remaining patients experienced third- or fourth-degree perineal tearing. The low incidence of tearing reflects the relatively safe nature of childbirth in developed countries; however, it also makes the estimation of our models more challenging. The second outcome, emergency C-section, is identified by the OPSC-4 codes R18.1-9 and was observed in approximately 20% of births. For the analysis pertaining to emergency C-sections we use the full sample of 15,117 births. In order to quantify our final outcome, post-birth LOS ($PBLOS$), we take the difference between time of birth and hospital discharge time, both of which are accurately reported in our data. This measure is heavily skewed, as shown in Figure 4 (left), with a median and mean of 32 and 44 hours, respectively. To reduce heteroskedasticity, for our main regressions we use its natural logarithm. The full sample of 15,117 non-elective C-section patients is used again for this investigation.

We also need to construct an accurate measure of workload. Other studies of hospital workload (e.g., KC and Terwiesch 2012, Kuntz et al. 2014) use occupancy, i.e., the number of patients present in the unit, to represent load. One problem with this measure is that it does not accurately account

Figure 4 Histogram of post-birth LOS (left) and workload (right).



for differences in staffing. Instead, these studies argue that staffing levels are stable over time and use day-of-the-week and shift-level dummy variables to capture any variation in the numbers and experience-mix of staff on duty. While these fixed-effect corrections are a good first approximation, especially for variations in physician staffing, they often fail to account for differences in nursing levels. Nursing levels exhibit a higher degree of variability than physician levels due to, for example, the chronic shortage of nurses, illness, holidays, maternity leave, restrictions due to labor laws and regulations, the availability of temporary agency staff, and requests for re-deployment to other areas of the hospital. Our detailed staffing data, which includes real-time information on how many midwives were present in the DU, allows us to overcome these limitations. For our study we use the patient count divided by the number of midwives on duty as our workload measure.

The workload that we assign to each patient is the time-weighted average load for the period two hours prior to birth. Two hours is chosen to capture both the majority of the time the patient spends in active labor, which is typically when the epidural decision is made, and the workload around the time of birth, which is the most critical when examining the direct effect on our chosen outcomes. It is also a short enough period to allow sufficient variation to enter into the workload variable.[4] If $N(t)$ is the number of patients in the DU at time t (including elective C-section patients, since they are also assigned a midwife) and $MW(t)$ is the number of midwives, the instantaneous workload at any time t can then be expressed as

$$LOAD(t) = \frac{N(t)}{MW(t)}.$$

[4] Instead of calculating the time-weighted average load over a fixed interval (two hours in this case), it might seem more natural to do so by averaging over the whole duration of labor. We do not do this, because it has the potential of introducing bias. In particular, as the time spent in labor is correlated with complications, the risk profile of a patient with a short labor will differ from that of a patient with a longer labor. Since averaging workload over a shorter time frame is more likely to generate more extreme levels of average workload, this would cause our dependent variables to be spuriously related to workload.

We calculate the time-weighted average load for a patient i who gives birth at time b_i using the following averaging formula

$$LOAD_i = \sum_{k \in L(s_i, b_i)} \frac{k}{b_i - s_i} \int_{s_i}^{b_i} \mathbb{1}[LOAD(t) = k]\, dt\,, \tag{1}$$

where s_i is the time two hours prior to birth, $L(s_i, b_i)$ is the set of all observed values of $LOAD(t)$ between $t = s_i$ and $t = b_i$ and $\mathbb{1}[\cdot]$ is the indicator function.

As reported in Table 1 the average workload is 1.18 patients per midwife, and its histogram is shown in Figure 4 (right). Since the ideal target is to have one midwife present per woman in labor (NAO 2013), this suggests that the DU is under-provisioned for 70% of deliveries. Furthermore, the positive correlation (coef. 0.114, which is significant at the 0.1% level and remains significant even after we control for time-of-day fixed effects) between occupancy and the number of midwives on duty (note that both occupancy and the number of midwives are, for the same reasons as described for workload, calculated as their time-weighted average in the two-hour period prior to birth) suggests that the hospital tries but fails to perfectly adjust the supply of resources to meet demand. This reinforces the importance of including information about staffing (and not just occupancy) in workload measures.

6. Econometric Models and Results: Interventions

We begin our empirical investigation by seeking to identify whether workload has an impact on the care pathway. We do so by estimating the likelihood of receiving an intervention at different levels of load.

6.1. Econometric Specification

To determine whether the prevalence of interventions changes with workload we build a latent-variable model (probit) for interventions with the time-weighted average load of (1) as the explanatory variable of interest, controlling for a wide range of maternal, contextual, medical, and other factors. A full list of controls can be found in Appendix A. This takes the form

$$INTERV_i^* = \alpha_0 + \mathbf{W}_i \boldsymbol{\alpha}_1 + LOAD_i \alpha_2 + \delta_i \tag{2}$$

$$INTERV_i = \mathbb{1}[INTERV_i^* > 0] \tag{3}$$

$$\delta_i \sim \mathcal{N}(0, 1)\,,$$

where $\mathbb{1}[\cdot]$ is an indicator function taking value one if the condition inside the brackets is satisfied and zero otherwise and the vector $\mathbf{W}_i$ contains the set of individual and environmental controls relevant for patient i. If higher workload has the effect of reducing the likelihood of receiving an intervention, as is hypothesized in **(H1a)**, then we should expect that $\alpha_2 < 0$.

Next, we unbundle the intervention variable and consider the workload effect on the rate of epidural analgesia and instrumental deliveries. In this section we model epidural analgesia and instrumental delivery as independent events. In §8.2 we show that this simplifying assumption does not have a material impact on our results. We model the impact of workload on epidural analgesia using the same set of exogenous covariates as for interventions, i.e.,

$$EPI_i^* = \beta_0 + \mathbf{W}_i \boldsymbol{\beta}_1 + LOAD_i \beta_2 + \omega_i \tag{4}$$

$$EPI_i = \mathbb{1}[EPI_i^* > 0] \tag{5}$$

$$\omega_i \sim \mathcal{N}(0,1)\,.$$

We proceed in the same way for instrumental deliveries but include epidural analgesia as an additional control. This is done to allow for the possibility that an epidural, which if administered would always be given prior to an instrumental delivery, may increase the risk of a patient requiring an instrumental delivery (Liu and Sia 2004). This takes the form

$$INSTR_i^* = \gamma_0 + \mathbf{W}_i \boldsymbol{\gamma}_1 + LOAD_i \gamma_2 + EPI_i \gamma_3 + \epsilon_i \tag{6}$$

$$INSTR_i = \mathbb{1}[INSTR_i^* > 0] \tag{7}$$

$$\epsilon_i \sim \mathcal{N}(0,1)\,.$$

For completeness we will consider both the model where γ_3 is restricted to 0 and that where it is a parameter for estimation. We are interested in whether workload has a greater effect on the rate of epidural analgesia than the incidence of instrumental delivery, due to the greater level of discretion, as hypothesized in **(H1b)**. If this is the case, then we would expect to find that the marginal effect of $LOAD$ on the epidural rate is greater than on the instrumental delivery rate.

The models described in this section were estimated using maximum likelihood estimation techniques (Greene 2002, p. 508-512), implemented in the `cmp` function in Stata 12.1 (Roodman 2011).

6.2. Results

Table 2 reports the estimated coefficients with robust standard errors. In discussing the findings we define a low-workload scenario as one that occurs at the 10th percentile of workload in the sample (0.752) and a high-workload scenario as one that occurs at the 90th percentile of workload (1.651). We compare outcomes under these two scenarios.

We find that there is a significant effect (coef. -0.097, p-value $= 0.008$) of an increase in workload on the likelihood of a patient receiving an intervention; this is in line with our hypothesis **(H1a)** that care pathways are modified as operational conditions change. Our model predicts that a patient who arrives in the DU when workload is high is 6.3% less likely (relative, 2.8% absolute) to receive an intervention than one who arrives when workload is low.

Table 2 Coefficient Estimates in Statistical Models for Interventions

	Probit			
	(1) Intervention	(2) Epidural	(3) Instr. Delivery	(4) Instr. Delivery
Workload	-0.097**	-0.079*	-0.057	-0.041
	(0.037)	(0.037)	(0.042)	(0.043)
Epidural	–	–	–	0.714***
	–	–	–	(0.031)
N	12080	12080	12080	12080
Log-Lik	-6746.28	-6853.96	-4790.23	-4524.16
Pseudo-R^2	0.181	0.106	0.281	0.321

Robust standard error in parentheses. Likelihood ratio (Pr $>\chi^2$) <0.0001 in all models.
$^{***}p<0.001$, $^{**}p<0.01$, $^{*}p<0.05$, $^{\dagger}p<0.10$

We break this down further to look at the workload effect on the individual processes captured within the intervention variable. In doing so we find evidence that supports **(H1b)**: The observed workload effect results primarily from a change in the rate of epidural analgesia, which is more discretionary and has a statistically significant coefficient; there is no statistically significant effect on the rate of instrumental delivery, which is a less discretionary intervention. If all patients arrived in the DU when workload was high, 31.9% could expect to receive an epidural. At low workload conditions, however, the expected epidural rate rises to 34.2%: 7.1% higher in relative terms.

7. Econometric Models and Results: Outcomes

Turning next to outcomes, we first identify any aggregate effect of workload and then divide this into its direct and indirect components.

7.1. Econometric Specification

To measure the direct effect we construct latent-variable models for perineal tears and emergency C-sections and a linear regression model for post-birth LOS, with workload as the main independent variable. To measure the indirect effect of workload through the choice of care pathway we perform a mediation analysis by adding the intervention as a mediator variable. Given that the sizes of the effect of workload on the constituent components of an intervention (i.e., epidural analgesia and instrumental delivery) are hypothesized to differ, epidural analgesia and instrumental delivery could also be treated as separate regressors. The model for perineal tearing takes the form

$$TEAR_i^* = \phi_0 + \mathbf{X}_i\boldsymbol{\phi}_1 + LOAD_i\phi_2 + EPI_i\phi_3 + INSTR_i\phi_4 + \mu_i \tag{8}$$

$$TEAR_i = \mathbb{1}[TEAR_i^* > 0] \tag{9}$$

$$\mu_i \sim \mathcal{N}(0,1)\,.$$

The model of the impact of workload on emergency C-sections is the same as for perineal tears except that in (8) we replace $\mathbf{X}_i$ with $\mathbf{Y}_i$ and the vectors of coefficients $\boldsymbol{\phi}$ with $\boldsymbol{\theta}$. The model for post-birth LOS takes the form:

$$PBLOS_i = \psi_0 + \mathbf{Z}_i\boldsymbol{\psi}_1 + LOAD_i\psi_2 + EPI_i\psi_3 + INSTR_i\psi_4 + \nu_i \qquad (10)$$
$$\nu_i \sim \mathcal{N}(0, \sigma_\nu^2)\,.$$

In these outcome models $\mathbb{1}[\cdot]$ is the indicator function and $\mathbf{X}_i$, $\mathbf{Y}_i$, and $\mathbf{Z}_i$ are the vectors of controls that we describe in Appendix A. In **(H2a)** and **(H2b)** the aggregate effects of workload on the rate of perineal tearing and emergency C-sections were assumed to be positive and negative, respectively. To find support for the former would require that $\phi_2 > 0$ in (8) when we restrict $\phi_3 = 0$. Similarly, to find support for the latter would require that $\theta_2 < 0$ when we restrict $\theta_3 = 0$. For post-birth LOS we do not speculate on the overall direction of the workload effect but note that it will be given by the coefficient ψ_2 in equation (10) when we restrict $\psi_3 = 0$. By restricting ϕ_3, θ_3, and ψ_3 to 0, the coefficients ϕ_2, θ_2, and ψ_2 estimate the aggregate effect of workload without parsing out the indirect impact through the epidural decision.[5]

We next discuss the decomposition into direct and indirect effects. This is simple to achieve in a linear model such as that for post-birth LOS. To see this we call the value of ψ_2 when we restrict $\psi_3 = 0$ the *total effect* and denote this by ψ_2^t. When we constrain ψ_3 to 0 we are effectively excluding the mediator variable; therefore, ψ_2^t captures both the direct and indirect effects of workload. When we include the mediator in the model, i.e., by removing the restriction $\psi_3 = 0$ in (10), we can estimate a new ψ_2, ψ_2^d. This is the unmediated component or the *direct effect*, i.e., the portion of the workload effect that does not occur through the intervention decision. The difference between the total effect and the direct effect is the mediated component, i.e., the *indirect effect* that occurs due to workload-driven changes in the intervention propensity. Therefore, if the direct effect of an increase in workload is an increase in the post-birth LOS, as hypothesized in **(H3a)**, we should find that $\psi_2^d > 0$ in (10). If the indirect component reduces the post-birth LOS, as per **(H3b)**, then we should find that $(\psi_2^t - \psi_2^d) < 0$ in (10).

These decomposition principles do not apply to non-linear binary probability models such as those used to model the occurrence of perineal tearing or an emergency C-section. This is because the regression coefficients and error variance are not identified separately. Instead, the estimated coefficients are scaled by a parameter that is a function of the standard deviation of the error. If the scaling parameter differs across models, then the decomposition of the total effect is not straightforward (MacKinnon and Dwyer 1993, Winship and Mare 1983). To get around this problem

[5] Note that it is not necessary to restrict ϕ_4, θ_4, or ψ_4 to 0 since there is no significant effect of workload on the rate of instrumental deliveries.

we use a method that holds the scale and the distribution of the errors constant (Breen et al. (2013)). This is implemented in the `khb` Stata module (Kohler and Karlson 2013). This methodology is used to test empirically for the primary workload effect hypothesized in **(H2a)** and **(H2b)**.

7.2. Accounting for Endogeneity

While it is possible to estimate the intervention model presented in §6.1 and the outcome models presented in the previous section independently, this would not be advisable due to the non-random assignment of patients to interventions. More specifically, there are some factors – which are observable to the care providers but may be unobservable to us the researchers – that make a patient more likely to experience a specific outcome and *also* more likely to receive some type of intervention. Examples include congenital anomalies in the baby, any previous history of tearing, and other complications not reported in our data but observable to the provider. As we have already shown that workload has no effect on the likelihood of a patient receiving an instrumental delivery we only need to correct for endogenous selection for epidural analgesia.

7.2.1. Endogeneity-corrected Models The endogeneity-corrected models that we consider are the simultaneous recursive bivariate probit (BivProbit) model for perineal tears and emergency C-sections and the Heckman treatment effects (HeckTreat) model for post-birth LOS (Maddala 1983, p. 123–129).

For tears and emergency C-sections, the main change when moving to the BivProbit model is in the structure of the errors, which will be jointly distributed according to a standard bivariate normal distribution with correlation coefficient ρ. Using tears as the example, the first stage will be as given in (5) and second stage as in (9), with error structure $(\omega_i, \mu_i) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho_{\omega\mu}\\ \rho_{\omega\mu} & 1\end{pmatrix}\right)$. This model is identified and can be estimated efficiently by ignoring the simultaneity using full information maximum likelihood (e.g., Greene 2002, p. 716).

The HeckTreat model for post-birth LOS is similar to BivProbit, with the first stage as in (5) and second stage as in (10) and with error structure $(\omega_i, \nu_i) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho_{\omega\nu}\\ \rho_{\omega\nu} & \sigma_\nu\end{pmatrix}\right)$. Note that HeckTreat differs from the standard Heckman sample selection model in two ways: First, the dummy variable indicating the treatment appears in the outcome equation; second, the outcome variable is observed regardless of whether or not the individual receives the treatment. Estimation is made using full information maximum likelihood (Fraser and Guo 2009).

The correlation between the error terms in the two models allows for the possibility that each equation is comprised of an unobserved individual heterogeneity component common to both. If there were no correlation (i.e., $\rho = 0$), then any unobserved effects would be unique to each equation, in which case it would be safe to estimate the model for perineal tears or emergency C-sections using the standard univariate probit model and for post-birth LOS using the OLS regression model. On the other hand, significant positive correlation would indicate that the unobserved components act on the outcome in the same direction in both equations, while negative correlation would indicate that they act in opposing directions.

While BivProbit and HeckTreat help us to account for potential endogeneity, they are restrictive as they assume that the effect of the exogenous variables on the outcome, as well as the unobservable variables (i.e., the error term), are the same regardless of which regime the patient is in (i.e., whether or not they received an epidural). However, often the fact that a patient is placed on one of these two pathways can imply a difference in the influence of the exogenous factors. To overcome this limitation we use an endogenous switching regression (SwitchReg) model. The SwitchReg model for the linear post-birth-LOS outcome can be expressed as (Winship and Mare 1983)

$$EPI_i = \mathbb{1}[EPI_i^* > 0], \tag{11}$$

$$PBLOS_{0i} = \psi_0^0 + \mathbf{Z}_i^0 \boldsymbol{\psi}_1^0 + LOAD_i \psi_2^0 + INSTR_i \psi_4^0 + \nu_i^0 \quad if \quad EPI_i = 0, \tag{12}$$

$$PBLOS_{1i} = \psi_0^1 + \mathbf{Z}_i^1 \boldsymbol{\psi}_1^1 + LOAD_i \psi_2^1 + INSTR_i \psi_4^1 + \nu_i^1 \quad if \quad EPI_i = 1, \tag{13}$$

$$(\omega_i, \nu_i^0, \nu_i^1) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_0\sigma_0 & \rho_1\sigma_1 \\ \rho_0\sigma_0 & \sigma_0^2 & x \\ \rho_1\sigma_1 & x & \sigma_1^2 \end{pmatrix} \right),$$

where EPI^* is as given in (4) and where $\sigma_0^2 = Var(\nu^0)$, $\sigma_1^2 = Var(\nu^1)$, $\rho_0\sigma_0 = \sigma_{\omega 0} = Cov(\omega, \nu^0)$ and $\rho_1\sigma_1 = \sigma_{\omega 1} = Cov(\omega, \nu^1)$. Note that the covariance between the errors in the two outcome equations, i.e., $x = Cov(\nu^0, \nu^1)$, is not identifiable, because only one of the two outcomes is ever observed. This is not a problem as estimates of the other parameters in the model do not depend on x. This model can be extended to the case where the outcome is binary.

The correlations ρ_0 and ρ_1 give the degree and direction of non-random selection of individuals into positions. If $\rho_0 < 0$, then there is positive selection of patients into the state in which they do not receive an epidural, meaning that post-birth LOS is higher for those patients who are selected to not receive an epidural than for the average individual in the population had random assignment of positions occurred. In other words, there are some unobservable characteristics of those patients that do not receive an epidural that make them more likely to stay longer in the hospital post-birth. $\rho_0 > 0$ indicates negative selection. The direction of the influence is reversed for ρ_1, with $\rho_1 > 0$ indicating positive selection and $\rho_1 < 0$, negative selection.

SwitchReg models can be estimated efficiently by maximum likelihood (Lee and Trost 1978). One drawback of the SwitchReg model is that we lose power since we are effectively estimating two different outcome equations. This makes it difficult to estimate this model when perineal tearing is the dependent variable due to the relatively rare occurrence: there are only 184 patients that both received an epidural and experienced a tear. Instead, for tears we estimate a model where we add the interaction term between workload and the epidural dummy in (8), i.e.

$$TEAR_i^* = \phi_0 + \mathbf{X}_i\boldsymbol{\phi}_1 + LOAD_i\phi_2 + EPI_i\phi_3 + INSTR_i\phi_4 + (LOAD_i \times EPI_i)\phi_5 + \mu_i \tag{14}$$

and estimate (9) and the BivProbit model with this new latent variable equation.

7.2.2. Exclusion Restriction In general, the simultaneous equation structure of models such as those defined in the previous section leads to identification problems. Overcoming them often necessitates the inclusion of at least one exogenous covariate that is present in the first-stage (selection) but not the second-stage (outcome) equation. This condition is often referred to as an exclusion restriction and the covariate as an instrumental variable (IV). In the BivProbit, HeckTreat, and SwitchReg models, however, this is not necessary as it is possible to identify the coefficients using the functional form of the recursive model (Wilde 2000, Maddala 1983, Mare and Winship 1988). While this may be the case, identification via non-linearity is generally not recommended and Monfardini and Radice (2008) show that (i) it may lead to problems in the empirical identification of the parameters and (ii) it is not robust to misspecification of the error distribution (i.e., deviation from the assumption of bivariate normality). As a result, we propose introducing an IV in the selection equation.

The IV used in our study is the average operating theater use other than for the focal patient in the period from six to two hours prior to the time of birth. An operating theater is taken to be in use in the hour prior to and 30 minutes post-birth of a baby delivered in theater. The times that we use are selected based on interviews with consultants and midwifery staff; the hour prior to birth accounts for the time that is required for the procedure to be performed, and the 30 minutes post-birth allows time for the patient to recover and the operating theater to be cleaned and prepared following the delivery. Operating theater use is expected to be influential in the intervention equation since, for example, an epidural can only be given when certain resources are available. Specifically, as discussed in §3, an epidural must be administered by an anesthesiologist at the recommendation an obstetric doctor. These resources become less available when operating theaters are busy, which may affect the likelihood of a patient receiving an epidural. Note that in calculating this measure we include both elective and emergency C-section patients since both types of patient affect the availability of specialist staff.

For operating theater use to be a valid IV it is also necessary that it is not correlated with the error in the outcome equation, i.e., that the IV has no effect on outcomes, either directly or indirectly through correlated omitted variables. The time lag between the measurement of the operating theater use (six to two hours before birth) and outcomes (which occur near to the time of

Table 3 Descriptive Statistics and Correlation Table for the Instrumental Variable

	Descriptive statistics				Correlation table						
Variable	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(8) Op. Tht. Use	0.28	0.32	0	2.06	0.096***	-0.003	-0.012	0.018*	0.016†	0.040***	0.023**

(1) Workload, (2) Intervention, (3) Epidural, (4) Instr. Delivery, (5) Tear, (6) Emerg. C-section, (7) Post-birth LOS
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

birth) makes it unlikely that it will have any direct impact on outcomes. As for the effect through omitted variables, we note that how busy operating theaters are will be primarily driven by time-related effects and the health characteristics of the other patients in the DU. We account for the former with time fixed-effects in our outcome equations. The latter is an omitted variable but not a correlated omitted variable since it cannot directly affect the outcome of the focal patient. As such, we believe that this proposed IV satisfies the necessary exclusion restriction. In §8.3 we strengthen this discussion with a formal test. Table 3 presents summary statistics of operating theater use ($Op.Tht.Use$), along with correlations with the variables presented earlier in Table 1.

7.3. Results

Tables 4, 5, and 6 report the estimated coefficients, standard errors, and model summary statistics of the outcome-related regressions. All controls have been omitted from the tables in order to concentrate on the three factors of interest: workload, interventions (which are potentially endogenous), and operating theater use (which serves as the IV). As in §6.2, in discussing the results we use the 10th (90th) percentile of workload to denote the low- (high-)workload scenario.

7.3.1. Perineal Tears When a baby is delivered by emergency C-section it is not possible for a perineal tear to occur and so the sample deployed to study perineal tears (Table 4) includes only non-C-section patients. The Probit (1) and BivProbit (1) models in Table 4 show the impact of workload and operating theater use on the epidural rate. The former ignores the problem of endogenous selection while the later explicitly corrects for it as explained in §7.2.1. We note that the instrumental variable in the selection equation is negative and highly significant (coef. -0.13, p-value $= 0.002$), as expected. Nevertheless, the results of the two models are nearly identical and the correlation coefficient ρ in the BivProbit is not significant at conventional levels (coef. 0.273, p-value $= 0.133$). This suggests that endogeneity is not important for perineal tears. For this reason we focus our discussion on the simpler Probit models using standard mediation analysis.

The Probit (2)–(5) models in Table 4 show the results of four different models for tears. We present all four to illustrate why it is necessary to carefully consider the care pathway when exploring the workload effect. In Probit (2) we include the workload variable only, while in Probit (3) we add the intervention regressor. In Probit (4) we split the intervention into epidural analgesia

Table 4 Coefficient Estimates in Statistical Models for Perineal Tears

	Probit					BivProbit	
	(1) Epidural	(2) Tear	(3) Tear	(4) Tear	(5) Tear	(1) Epidural	(2) Tear
Workload	-0.079*	0.108*	0.108*	0.101†	0.138*	-0.078*	0.084
	(0.037)	(0.057)	(0.057)	(0.057)	(0.066)	(0.037)	(0.058)
Workload×Epidural	–	–	–	–	-0.117	–	–
	–	–	–	–	(0.119)	–	–
Op. Tht. Use	-0.133**	–	–	–	–	-0.135**	–
	(0.043)	–	–	–	–	(0.043)	–
Intervention	–	–	-0.019	–	–	–	–
	–	–	(0.050)	–	–	–	–
- Epidural	–	–	–	-0.213***	-0.074	–	-0.653*
	–	–	–	(0.048)	(0.150)	–	(0.281)
- Instr. Delivery	–	–	–	0.372***	0.373***	–	0.363***
	–	–	–	(0.073)	(0.073)	–	(0.072)
N	12080	12080	12080	12080	12080	12080	
Log-Lik	-6853.96	-2229.59	-2229.52	-2207.82	-2207.32	-9061.35	
Pseudo-R^2	0.106	0.062	0.062	0.071	0.072	–	
ρ	–	–	–	–	–	0.272 (0.174)	

Robust standard error in parentheses; Likelihood ratio ($\Pr > \chi^2$) < 0.0001 in all models
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

and instrumental delivery. Finally, in Probit (5) we include an interaction between workload and epidural analgesia. We find that in all four models there is evidence of a positive direct effect of workload on the tear rate, in line with **(H2a)**. The Probit (2) and (3) models are very similar, initially suggesting that interventions do not mediate the workload effect. However, Probit (4) shows that the insignificance of the intervention variable recorded in Probit (3) is due to the opposing effects of administering an epidural and receiving an instrumental delivery on the tear rate. Considering these separately reveals that perineal tearing is less likely to occur to a patient that receives an epidural but is more likely for one who experiences an instrumental delivery. In Probit (5) the interaction is insignificant, indicating that there is no discernible difference in the workload effect on the tear rate between patients that did and did not receive an epidural.

To calculate the impact of workload on tears further we focus on the Probit (4) model. The model predicts that the total effect of moving from a low- to high-workload environment would be a 0.91% increase in the tear rate, or a relative increase of 20.1%. We also find that the average treatment effect (ATE) of epidural use is 1.94%.[6] In other words, had all patients received an epidural, the tear rate is predicted to have been 3.8%: 34.1% lower in relative terms than the projected 5.7% tear rate had none been given an epidural. Since workload affects the incidence of epidural analgesia and whether a patient receives an epidural affects the tear rate, we can conclude that there is an indirect effect of workload on tearing.

[6] ATEs are calculated under the assumption that all covariates except the binary variable (treatment) of interest remain the same but also that there is an exogenous factor that causes all observations to either receive or not receive the treatment.

Table 5 **Coefficient Estimates in Statistical Models for Emergency C-sections**

	Probit			BivProbit	
	(1) Epidural	(2) C-section	(3) C-section	(1) Epidural	(2) C-section
Workload	-0.068*	-0.035	-0.031	-0.069*	-0.008
	(0.033)	(0.040)	(0.040)	(0.033)	(0.039)
Op. Tht. Use.	-0.168***	–	–	-0.140***	–
	(0.039)	–	–	(0.039)	–
Epidural	–	–	0.132***	–	0.900***
	–	–	(0.031)	–	(0.104)
N	15117	15117	15117	15117	
Log-Lik	-8445.20	-5243.02	-5233.86	-13669.88	
Pseudo-R^2	0.145	0.309	0.310	–	
ρ	–	–	–	-0.461*** (0.059)	

Robust standard error in parentheses; Likelihood ratio (Pr $> \chi^2$) < 0.0001 in all models
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

We proceed to decompose the total effect of workload into the direct (due to quality of administered care) and indirect (via the epidural decision) effect as described in §7.1. We find that the contribution of the direct effect of moving from low- to high-workload on the tear rate is 0.87% and the indirect is 0.04%. Therefore, the effect of workload on the rate of perineal tearing is primarily (95% of the total) being driven by the direct effect, as hypothesized in **(H2a)**.

7.3.2. Emergency C-sections We next turn our attention to the results of the emergency C-sections models in Table 5. For these results we use the full sample.

As in the tear models in Table 4, the Probit (1) and BivProbit (1) models in Table 5 show the impact of workload and operating theater use on the rate of epidurals. Unlike for tears, the correlation coefficient ρ in this case is negative and highly significant (coef. -0.461, p-value < 0.001). This indicates that there exist unobservable variables that result in a patient who is less likely to receive an epidural being more likely to receive an emergency C-section and vice versa. This is not too surprising: if the unobservable factors are medical indicators of risk, then it seems reasonable that higher-risk patients would be less likely to receive an epidural – the epidural may exacerbate any existing complications (Anim-Somuah et al. 2011). For this reason, the more appropriate model for examining the impact of workload on emergency C-sections is the BivProbit.[7] Unlike for the tear models, instrumental deliveries are not included as a control in the outcome equation in this case as patients cannot receive both an instrumental delivery and an emergency C-section (i.e., they are mutually exclusive).

The BivProbit (2) finds that there is no direct effect of workload on the chance of a patient receiving an emergency C-section. On the other hand, the effect of receiving an epidural on the

[7] A SwitchReg model was also estimated, although is not reported since the findings are in line with those of the BivProbit model.

Table 6 Coefficient Estimates in Statistical Models for Post-birth LOS

	OLS				SwitchReg		
	(1) LOS	(2) LOS	(3) LOS	(4) LOS	(1) Epidural	(2) LOS-0	(3) LOS-1
Workload	-0.001	0.008	0.007	0.024	-0.067*	0.049*	-0.022
	(0.016)	(0.016)	(0.016)	(0.019)	(0.032)	(0.022)	(0.022)
Workload×Epidural	–	–	–	-0.047	–	–	–
	–	–	–	(0.031)	–	–	–
Op. Tht. Use	–	–	–	–	-0.138***	–	–
	–	–	–	–	(0.035)	–	–
Intervention	–	0.311***	–	–	–	–	–
	–	(0.012)	–	–	–	–	–
- Epidural	–	–	0.227***	0.283***	–	–	–
	–	–	(.013)	(0.039)	–	–	–
- Instr. Delivery	–	–	0.293***	0.294***	–	0.314***	0.248***
	–	–	(0.017)	(0.017)	–	(0.025)	(0.021)
N	15117	15117	15117	15117	15117	9678	5439
Prob > F	< 0.0001	< 0.0001	< 0.0001	< 0.0001	–	–	–
Adj. R^2	0.339	0.364	0.371	0.369	–	–	–
ρ	–	–	–	–	–	-0.674***	-0.046
	–	–	–	–	–	(0.033)	(0.033)

Robust standard error in parentheses; Likelihood ratio (Pr $> \chi^2$) < 0.0001 in all models
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

probability of needing an emergency C-section is highly significant (coef. -0.140, p-value < 0.001). Our model indicates that had none of the patients received an epidural, the predicted percentage of patients who would have required an emergency C-section would have been 13.9%, while had all of the patients received an epidural, this number instead increases to 33.8%: an absolute increase (or ATE) of 20.0% (144% relative).

Using the same decomposition as for tears, we can estimate the indirect effect of workload on the emergency C-section rate through changes in the provision of epidural analgesia. Overall there is a small but significant reduction in the emergency C-section rate (a relative decrease of 2%) at high-workload conditions as compared to low-workload, and since there is no direct effect, the effect is exclusively indirect through the change in the epidural rate. This is evidence in support of hypothesis **(H2b)**.

7.3.3. Post-birth LOS The final outcome we examine is post-birth LOS. As with the emergency C-section results, we work with the full sample. The two types of model we examine are OLS and SwitchReg (Tobit-5). The former ignores non-random assignment while the latter accounts for it. Model coefficients are presented in Table 6.

For our discussion we will focus on the SwitchReg model as there is evidence of non-random selection. This is indicated by the significant value of ρ_0 (coef. -0.674, p-value < 0.001). The negative sign indicates that post-birth LOS is higher for those patients that are selected to not receive an epidural than if random assignment had occurred. Based on our earlier discussion this is reasonable: since an epidural increases the chance of adverse outcomes, it is less likely to be given to a high-risk patient, and since these patients are higher risk, they also have a higher expected post-birth LOS. In contrast, there is no evidence of a selection effect for those patients that received an epidural, as indicated by the insignificant ρ_1.

Estimating the size of the direct and indirect effects of workload under the SwitchReg model is complex. For ease of exposition we defer the description to Appendix B. We find that there is a direct workload effect, but only for those patients that do not receive an epidural. When moving from low- to high-workload conditions there is a 4.4% increase in post-birth LOS for those patients that did not receive an epidural and no change otherwise. This finding supports hypothesis **(H3a)**. Note that this does not appear in the basic OLS models, even where we interact the workload variable with the epidural effect in OLS. This result illustrates the need to use the SwitchReg model, which not only controls for endogeneity but also allows for differential effects across different groups within the population.

In addition to the direct effect, there is a further indirect effect that occurs through the change in epidural rate. The average treatment effect (ATE) of epidural analgesia on post-birth LOS is 15.3 hours: a relative increase of 44.3%. As a result, the indirect effect (through the reduction in the epidural rate) of moving from low- to high-workload conditions is a 1.6% decrease in post-birth LOS for those patients that did not receive an epidural and no change otherwise. This finding is in line with **(H3b)**. Combining the direct and indirect effects of workload, the aggregate effect of moving from a low- to high-workload scenario is an increase of 2.8% for patients that did not receive an epidural and no change otherwise. When averaged across all patients the total effect is close to 0, as indicated by the OLS (2)-(4) models, where workload does not have a statistically significant aggregate effect on outcomes.

8. Robustness Checks

8.1. Alternative Workload Measures

The time-averaged workload employed throughout the empirical analysis was measured over the two-hour window prior to delivery. Since the choice of two hours is arbitrary, in this section we repeat the analysis using two alternative assumptions: 1.5 and 2.5 hours. We find no significant changes in the results (results not reported here).

8.2. Endogeneity of Instrumental Deliveries

In estimating the impact of workload on the likelihood of a patient receiving an epidural or instrumental delivery in §6.1 we assumed that the two events, epidural or instrumental delivery, are independent. In this section we relax this assumption.

Table 7 Coefficient Estimates in the Bivariate Probit Model for Instrumental Deliveries

	BivProbit	
	(1) Epidural	(2) Instr. Delivery
Workload	-0.081*	-0.061
	(0.037)	(0.040)
Op. Tht. Use	-0.143***	–
	(0.041)	–
Epidural	–	-0.280
	–	(0.194)
N	12080	
Log-Lik	-22376.62.00	
ρ	0.570*** (.105)	

Robust standard error in parentheses.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

To see why the independence assumption may be problematic note that this is a sequential decision process in which an epidural, if administered, would be given in advance of an instrumental delivery (Liu and Sia 2004). If there exist unobservable variables that drive both the decision to administer an epidural and the decision to choose an instrumental delivery, then the estimated effect of the epidural on the likelihood of a patient receiving an instrumental delivery, given by the epidural coefficient of Probit (4) in Table 2, would be biased. Furthermore, if workload happens to be correlated with the epidural variable, then the estimated coefficient of workload may also be affected. To overcome this problem we construct a BivProbit model, with the epidural decision in the first stage and instrumental delivery decision in the second, following the methodology of §7.1.

The results of this model are reported in Table 7. We find no change in the impact of workload. There is, however, evidence that the unobservable variables that increase the chance of a patient receiving an epidural also increase the likelihood that the patient will receive an instrumental delivery. This is indicated by the positive and significant ρ (coef. 0.570, p-value < 0.001). After accounting for this we find that contrary to the finding in Table 2, epidural analgesia does not increase a patient's chance of receiving an instrumental delivery. Overall the BivProbit model reaffirms that workload affects epidurals but not instrumental deliveries (and, furthermore, that epidurals do not affect instrumental deliveries directly). Therefore, it is sufficient to control for instrumental delivery in the outcome estimations, as we have done in §7.1.

8.3. Validity of Instrumental Variable

The IV used in §7.1 needs to be related to the outcome in the first-stage equation (after we control for observables) but unrelated to the error term (which captures unobservable factors) in the outcome equation. With weak instrumental variables, i.e. those that are only weakly correlated with the endogenous variables, estimates may be inconsistent, tests for the significance of coefficients

Table 8 Tests for Weakness of the Instrumental Variable

	Instr. Delivery	Tear	Emerg. C-section	Post-birth LOS
F-statistic	18.20	17.91	18.22	20.89
Cragg-Donald F	17.87	17.50	17.92	20.44

Critical value for F-statistic is 10; for Cragg-Donald F-statistic is 16.38

may lead to the wrong conclusions, and confidence intervals will be incorrect. Therefore, it is important to formally test for the validity of our instrument. Since the majority of tests are based on a linear IV regression model where the dependent variable in the outcome equation is continuous, in performing these tests we follow convention and relax the assumption of a binary outcome.

We perform two tests for weak instrumental variables. In the first we calculate the F-statistic from the OLS regression of the endogenous variable on the instrument, as per Staiger and Stock (1997). This is typically compared to the "rule of thumb" value of 10. Any value greater than 10 suggests that the instrument is not weak. A more formal approach proposed in Stock and Yogo (2005) is a test on the null hypothesis that the maximum relative bias (i.e., the furthest the coefficient estimate of the endogenous variable may be away from its true value) is at least 10%. To use this method we compare the Cragg-Donald F-statistic to the 5% critical values given in Table 1 of Stock and Yogo (2005). Results are presented in Table 8.

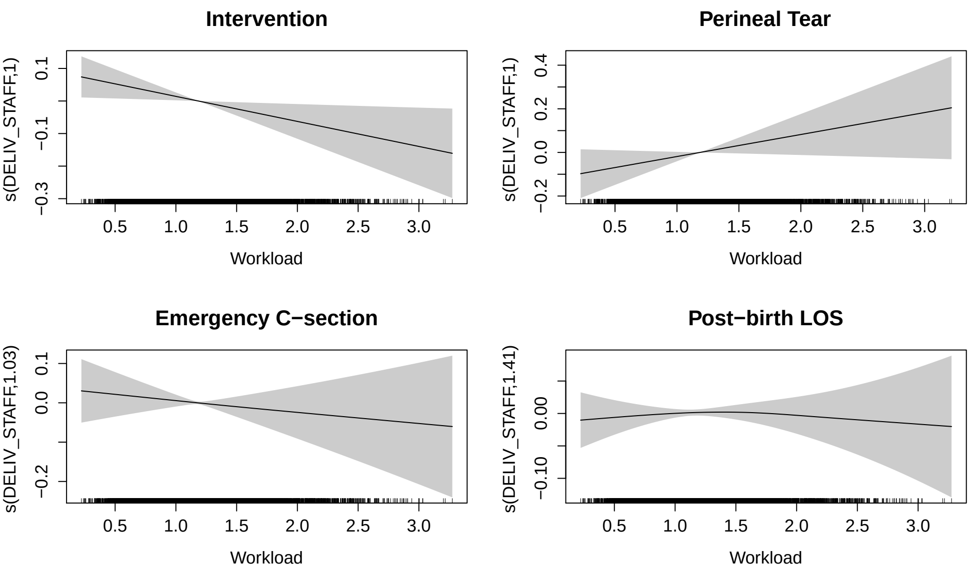
The critical value of the Cragg-Donald F-statistic is identical (16.38) across the models since the number of endogenous regressors is the same (one in each case). We find that the F-statistics are greater than the critical values using both methods, indicating that there is no evidence to suggest that our models suffer from the problem of weak instruments.

8.4. Non-linear Workload

Other work has shown that the workload effect may be non-linear (see e.g., Kuntz et al. 2014, Tan and Netessine 2012). Therefore, it is a useful exercise to check whether there is any indication of a non-linear workload effect in the DU setting.

We could use a number of methods to look for this, from dividing the workload variable into a number of binary variables with cuts made at arbitrary percentiles to fitting a piecewise-linear (segmented) regression model with knots (or breakpoints) as additional parameters for estimation. Since our dataset is relatively small, however, we eschew these methods in favour of an approach using generalized additive models (GAMs). These models are similar to the standard OLS or probit specification, except that the dependent variable is allowed to depend on some (unknown) smooth function of one or more of the predictors. This allows us to estimate a non-linear effect of workload on each of the outcome equations by applying a smoothing function to this variable during estimation. To avoid going into the technical details, we mention only that these models

Figure 5 Examining non-linear workload effects.



are estimated using a local scoring algorithm by iteratively fitting weighted additive models by backfitting using the Gauss–Seidel method (Wood 2011). This is implemented by the `gam` function in the `mgcv` package within the statistical software R version 3.0.0 (Wood 2013).

The plots in Figure 5 show the component smooth function from the fitted GAMs on the scale of the linear predictor, with two standard error bounds (approx. 95% Confidence Intervals) indicated by the shaded region. The estimated degrees of freedom (EDF), reported in the vertical-axis labels, approximately correspond to the highest order of a polynomial transformation of the workload variable that would need to be included in the standard OLS/Probit model in order to capture the non-linear relationship. The EDF of interventions, tears, and emergency C-sections are close to 1, suggesting that there is no deviation from linearity. Workload appears to have a small non-linear effect on post-birth LOS, although the confidence bounds around the component smooth function are wide. Since the deviation from linearity occurs far into the right tail (recall the 90th sample percentile is 1.651), however, it is unlikely to substantially affect the results.

8.5. Non-Maternal Outcomes

The nature of conducting a study on the DU means that there are actually two patients to consider: the mother and also the baby that is being delivered. Although we have focused on maternal outcomes, it is possible that there may also be an affect of workload on outcomes related to the baby. Since neonatal mortality is too rare an occurrence to be a useful outcome measure, discussions with obstetricians indicated two alternative outcome measures: (i) Apgar score at birth and (ii) neonatal intensive care unit (NICU) admission rate for babies born after 37 weeks of gestation.

The Apgar score is a commonly used measure for evaluating a newborn baby based on five criteria on a combined scale from 0 to 10, with higher values indicating a better health assessment (Papile 2001). Regarding NICU admissions, babies will only be admitted to a NICU if they need additional support in the early stages of their life. Term babies (i.e., those born after 37 weeks) should have developed fully and are therefore less likely to be admitted to a NICU. It is therefore of interest to consider whether the Apgar score or the propensity of NICU admissions differ based on workload levels. To investigate this we specified and estimated a series of models similar to those defined of §7.1. We did not find any link between workload, either direct or indirect, and outcomes for the baby (results not reported here).

9. Discussion

This study contributes to the ongoing debate on the effect of system state on system performance, with system load as the critical state variable. As in the existing literature, we use a detailed data set from a healthcare context, the delivery unit of a maternity department. In contrast to prior studies that considered the system itself as a "black box," in this study we open the box by considering the different effects of workload on discrete decisions about the type of service that a customer receives (e.g., the care pathway) and the potential deterioration in service quality (e.g., adverse medical outcomes) at elevated workload levels. We find that workload affects both and uncover a rich interaction between these two effects that provides valuable insight into the nature of interventions that might help safeguard service quality as workload varies.

First, our data provides strong evidence that care pathway decisions that have a significant degree of discretion, such as epidural analgesia, are most affected by workload. At elevated workload levels such decisions are affected by a "cutting-corners" effect (Oliva and Sterman 2001), while at low workload levels a "devil-makes-work-for-idle-hands" phenomenon results in over-treatment (Tan and Netessine 2012). Workload therefore exhibits a monotone effect on resource use at such decision points.

Second, we find that clinical outcomes, such as tears and emergency C-sections, and operational outcomes, such as post-birth LOS, are affected by workload either directly, through impaired care provision, or indirectly, as a consequence of changes at specific decision points along the care pathway. In our context, the first effect occurs in relation to perineal tearing. Continuous monitoring and timely interventions in response to developing maternal health during the final phase of labor are important to avoid tearing (Aasheim et al. 2011) and the unit's ability to provide this is reduced at high midwife workload. The data provides strong evidence of this effect, while the evidence that workload affects tears through the decision to administer an epidural is much less pronounced. By contrast, emergency C-section rates do not seem to be directly affected by midwife workload. The occurrence of an emergency C-section is predominantly driven by biological factors that are outside the control of midwives. However, this does not mean that workload does not affect emergency C-section propensity; the effect is indirect through the workload effect on the earlier decision to administer epidural analgesia, which is positively associated with emergency C-section rates.[8] This demonstrates that clinical outcomes can be affected by workload at decision points that are quite detached from the point at which outcomes are finally observed. The impact of workload on post-birth LOS is even more nuanced. While workload has no statistically significant aggregate effect, we find that higher workload levels lead to a positive direct effect, possibly due to impaired care provision, which leads to complications such as tears, and a negative indirect effect through the care pathway decision (i.e., workload reduces the propensity of receiving epidural analgesia, which is associated with longer post-birth hospital stays). This emphasizes the need to appropriately decompose the impact of workload in order to truly understand how it affects operational outcomes.

Our work has implications for managerial decisions such as capacity planning. More specifically, it allows management to predict how the incidence of interventions, such as epidural analgesia, and outcomes, such as tears, emergency C-sections, and post-birth LOS, would change under different staffing and/or demand scenarios. We illustrate this point by estimating how interventions and outcomes would changes for the DU we study if it was to move from the current staffing practice that achieves the desired one-to-one (or better) level of care 33% of the time to a staffing level that achieves this 70% or 95% of the time. The number of additional staff required to achieve this was recently investigated by the National Audit Office (NAO 2013) – see also (Green and Liu 2013)). As shown in Table 9, our results imply that increasing staffing levels to achieve 70% one-to-one care (Column 1) is associated with a relative change in the propensity of epidural analgesia by +2.85%, tears, by −6.93%, and emergency C-sections, by +1.04% at the mean workload, compared with current practices. In the case of 95% one-to-one care (Column 2) the resulting changes would be +4.58%, −10.90%, and +1.66% for epidural analgesia, tears, and emergency C-sections, respectively. The *average* post-birth LOS would not be materially affected by any of these changes as the direct and indirect effects cancel each other out. Furthermore, the difference in the incidence of interventions and outcomes is more pronounced for women delivering when the DU is very busy, e.g., when the DU operates at the 90th percentile of workload. Motivated by this observation we also examine a perfectly flexible staffing regime in which the DU adds staff only when workload exceeds the desired one-to-one level of care and up to a maximum of 14 staff members. As shown in Column 3, this purely hypothetical way of staffing has the potential to mitigate the adverse

[8] In the words of a senior midwife on this unit, epidural analgesia sets the patient on a "path to further intervention."

Table 9 Relative Change in Expected Outcomes under Alternative Staffing Scenarios. Effect Estimated at the 10th%ile, mean, and 90th%ile of Workload Scenarios

	Fixed 11 MW			Fixed 14 MW			Flexible MW		
	10^{th}%ile	Mean	90^{th}%ile	10^{th}%ile	Mean	90^{th}%ile	10^{th}%ile	Mean	90^{th}%ile
Workload Value	0.576	0.873	1.176	0.452	0.686	0.924	0.752	0.914	1.000
Pr(Workload>1)		0.287			0.043			0.043	
Avg. MW per Obs.		11			14			10.40	
Epidural Analgesia	1.57%	2.85%	4.46%	2.68%	4.58%	6.87%	0%	2.46%	6.14%
Perineal Tears	-4.05%	-6.93%	-10.20%	-6.80%	-10.90%	-15.26%	0%	-6.03%	-13.77%
Emerg. C-sections	0.58%	1.04%	1.59%	1.00%	1.66%	2.44%	0%	0.90%	2.18%
PB LOS (hours)	-0.32%	-0.58%	-0.89%	-0.54%	-0.92%	-1.35%	0%	-0.50%	-1.21%
- $\|EPI=0$	-0.56%	-0.99%	-1.52%	-0.94%	-1.58%	-2.31%	0%	-0.86%	-2.07%
- $\|EPI=1$	0.02%	0.03%	0.05%	0.03%	0.05%	0.08%	0%	0.03%	0.07%

impact of workload for those patients that experience a busy DU, without significantly increasing the average number of staff required. Implementing such flexible staffing procedures, however, requires substantial organizational changes.

From a purely financial perspective, it is sometimes argued that the costs associated with expanding capacity (e.g., increased staffing) may be partially offset by cost savings due to improvements in the quality of care (e.g., a lower incidence of complications and adverse outcomes). Our results can be used to estimate whether this is the case. In our setting, the cost of an unplanned C-section is estimated to be £3,042 versus £1,512 for a vaginal delivery, while an epidural birth costs £87 extra on average and tearing has an associated cost of £504 (NICE 2012). The full economic cost of an additional nurse is £66,365 per annum (Curtis 2012). Using these numbers, we find that moving to a fixed midwife level of 11 or 14 or to a flexible level as in Table 9, which would have added £398,000, £1,460,00, and £331,825, respectively to the staffing bill per annum, would also lead to an *increase* in direct patient costs of approximately £45,000, £71,600, and £38,400 p.a., respectively. This increase is mostly due to the increase in emergency C-sections.

We conclude this study by noting that the idea of decomposing the impact of workload into direct and indirect effects may be important in other service settings where servers have, at least to some extent, discretion over the type of service they provide. In order to understand and devise appropriately targeted interventions to cope with the impact of high workload, managers responsible for such services need to understand which parts of the process are affected by workload. The modeling methodology used in this paper provides an approach to this end.

Appendix

A. Controls

In Table 10 we list all of the exogenous regressors (controls) for the models presented in Tables 2–6. These can broadly be broken down into six categories: factors related to the mother, those specific to the pregnancy, those that relate to the baby, time controls, a subset of the medical conditions that may affect outcomes (chosen from the relevant medical literature), and other relevant clinical procedures. The number following the variables specified as categorical indicates the number of categories. We indicate the models in which they were included by either the direction of their effect, as indicated by the sign and significance of the estimated coefficient (+ for positive and significant, − for negative and significant, 0 for insignificant, all at the 5% level), and for categorical variables by Y if one or more of the levels was significant at the 5% level and 0 otherwise. N means the factor was not included as a control in the corresponding model. A useful exercise is to check that the direction of the reported effects in the models correspond with intuition and with medical literature (e.g., Bragg et al. 2010, Renfrew et al. 1998, Eason et al. 2000). For example, larger babies are associated with a higher likelihood of causing tearing, therefore, we should expect a positive coefficient for the "Baby Weight" variable in the "Tears" equation, which we find. Medical complications in general have been shown to lead to poorer outcomes (in terms of tears, emergency C-section rates, and required LOS), which again we see.

B. Effect Size Calculations for Post-birth LOS (§7.3.3)

In this Appendix we estimate the average treatment effect (ATE) of epidural analgesia on post-birth LOS and the direct and indirect effects of workload on post-birth LOS. To determine the ATE we calculate the expected post-birth LOS of each patient under two scenarios: the first assumes that the patient did not receive an epidural while the second assumes that she did. More specifically, for a patient that did not receive an epidural we first calculate the *actual* expected LOS (i.e., the expected LOS if they *had not* received an epidural given that in actuality they *did not*) and then the *counterfactual* expected LOS (i.e., the expected LOS if they *had* received an epidural given that in actuality they *did not*). We follow a similar procedure for patients that did receive an epidural. We then take the difference between the expected values under the two scenarios and average across all patients to estimate the ATE (Poirier and Ruud 1981, Mare and Winship 1988). Using the notation of equations (11)–(13), the actual expectations can be written as

$$\begin{aligned}\mathbb{E}\left[PBLOS_{0i}|EPI_i=0\right] &= \psi_0^0+\mathbf{Z}_i^0\boldsymbol{\psi}_1^0+LOAD_i\psi_2^0+INSTR_i\psi_4^0-\sigma_{\omega 0}\lambda_{1i}\,,\\ \mathbb{E}\left[PBLOS_{1i}|EPI_i=1\right] &= \psi_0^1+\mathbf{Z}_i^1\boldsymbol{\psi}_1^1+LOAD_i\psi_2^1+INSTR_i\psi_4^1+\sigma_{\omega 1}\lambda_{2i}\,,\end{aligned}$$

and the counterfactual expectations, as

$$\begin{aligned}\mathbb{E}\left[PBLOS_{1i}|EPI_i=0\right] &= \psi_0^0+\mathbf{Z}_i^0\boldsymbol{\psi}_1^0+LOAD_i\psi_2^0+INSTR_i\psi_4^0-\sigma_{\omega 1}\lambda_{1i}\,,\\ \mathbb{E}\left[PBLOS_{0i}|EPI_i=1\right] &= \psi_0^1+\mathbf{Z}_i^1\boldsymbol{\psi}_1^1+LOAD_i\psi_2^1+INSTR_i\psi_4^1+\sigma_{\omega 0}\lambda_{2i}\,,\end{aligned}$$

where $\sigma_{\omega 0}$ and $\sigma_{\omega 1}$ are the estimated covariances described in §7.2.1, $\lambda_{1i}=\frac{\phi(\mathbf{W}_i\boldsymbol{\beta})}{1-\Phi(\mathbf{W}_i\boldsymbol{\beta})}$, and $\lambda_{2i}=\frac{\phi(\mathbf{W}_i\boldsymbol{\beta})}{\Phi(\mathbf{W}_i\boldsymbol{\beta})}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf respectively, $\mathbf{W}_i$ are the exogenous regressors, and $\boldsymbol{\beta}$ are the fitted values from the epidural (selection) equation. These equations account for the dependence

Table 10 Table of Controls.

	Type	Interventions ($\mathbf{W}_i$)	Tears ($\mathbf{X}_i$)	C-sections ($\mathbf{Y}_i$)	LOS ($\mathbf{Z}_i$)
Maternal Characteristics					
- Age	Categorical (4)	Y	Y	Y	Y
- Num. Prev. Births	Categorical (4)	Y	Y	Y	Y
- Previous C-section	Binary	+	+	+	+
Pregnancy Characteristics					
- Gestation	Categorical (5)	Y	Y	Y	Y
- Multiple Birth	Binary	N	Y	+	+
- Medicine	Binary	0	0	+	Y
- History	Binary	+	0	0	+
- Ante-natal Problems	Binary	0	0	+	+
Baby Characteristics					
- Baby Weight	Continuous	0	+	0	−
Temporal					
- Year-Qtr	Categorical (16)	Y	Y	Y	Y
- Hour of Birth (4-hourly)	Categorical (6)	Y	Y	Y	Y
- Weekend	Binary	0	0	0	Y
Medical Complications					
- Breech	Binary	0	0	+	0
- Dystocia	Binary	+	+	+	+
- Pre-existing Diabetes	Binary	0	+	+	0
- Gestational Diabetes	Binary	0	−	0	+
- Pre-existing Hypertension	Binary	0	0	+	+
- Eclampsia	Binary	+	0	+	+
- PROM	Binary	+	0	0	+
- Placenta Praevia	Binary	N	N	+	0
Other Procedures					
- Induction	Binary	+	0	0	+
- Episiotomy	Binary	N	−	N	N
- Emerg. C-section	Binary	N	N	N	+
Other Operational					
- Post-birth Workload	Continuous	N	N	N	+

Medicine indicates that a patient was prescribed medicine during pregnancy for complications
History indicates that a patient has a history of experiencing complications during childbirth
Ante-natal Problems indicates that a patient experienced complications in the ante-natal period
PROM indicates that a patient experienced premature rupture of membranes

of post-birth LOS on observable regressors (captured by $\mathbf{Z}, LOAD, INSTR$) and by non-random selection (captured by λ_{1i} and λ_{2i}). Using this methodology and after retransforming back to the original scale by taking the exponential of the fitted values and rescaling by multiplying with the Duan smearing estimator (Duan 1983), the ATE of epidural analgesia on post-birth LOS is 15.3 hours, or a relative increase of 44.3%.

The conditional effect of workload can be written as

$$\frac{\partial \mathbb{E}\left[PBLOS_{0i}|EPI_i=0\right]}{\partial LOAD_i} = \psi_2^0 + \beta_2\sigma_{\omega 0}\lambda_{1i}(\mathbf{W}_i\boldsymbol{\beta} - \lambda_{1i}), \quad \frac{\partial \mathbb{E}\left[PBLOS_{1i}|EPI_i=1\right]}{\partial LOAD_i} = \psi_2^1 - \beta_2\sigma_{\omega 1}\lambda_{2i}(\mathbf{W}_i\boldsymbol{\beta} + \lambda_{2i}),$$

where β_2 is the estimated coefficient of workload in the epidural equation. The conditional direct effect of workload is given by the first term and the indirect effect by the second term. The term ψ_2^0 (ψ_2^1) is the coefficient of the workload variable in the post-birth LOS equation for $EPI_i = 0$ ($EPI_i = 1$) of the SwitchReg model. Thus, the direct workload effect on post-birth LOS conditional on not receiving an epidural is estimated to be $\hat{\psi}_2^0 = 0.049$ (p-value $= 0.028$) and conditional on receiving an epidural it does not significantly differ from 0 (p-value $= 0.324$). The conditional indirect effect is patient specific and so we report the average, i.e., we average $\beta_2\sigma_{\omega 0}\lambda_{1i}(\mathbf{W}_i\boldsymbol{\beta} - \lambda_{1i})$ across all patients who did not receive an epidural and $-\beta_2\sigma_{\omega 1}\lambda_{2i}(\mathbf{W}_i\boldsymbol{\beta} + \lambda_{2i})$ across all those who did. The indirect workload effect is estimated to be -0.0174 on average for a patient who did not receive an epidural and -0.001 for one who did. The small indirect effect for those patients that did not receive an epidural is due to the fact that ρ_1, the correlation between the error terms in the epidural equation and the post-birth LOS equation for those patients who did not receive an epidural, is estimated to be very close to 0 (coef. -0.046, p-value $= 0.161$).

The log-linear structure of the model for post-birth LOS means that the direct and indirect conditional workload effects on post-birth LOS can be interpreted as a percentage change. Therefore, to estimate the size of the effect when moving from low- to high-workload conditions we multiply the conditional direct and indirect workload effects by the difference between the 90th and 10th percentiles of workload (0.890).

References

Aasheim, V., A. Nilsen, M. Lukasse, L. Reinar. 2011. Perineal Techniques During the Second Stage of Labour for Reducing Perineal Trauma. *Cochrane Database of Systematic Reviews* (12) 1–47.

ACA. 2010. Patient Protection and Affordable Care Act. Tech. Rep. section 3403 as amended by section 10320, Public Law.

AHRQ. 2013. Patient Safety Indicators Technical Specifications. URL `http://www.qualityindicators.ahrq.gov/Modules/PSI_TechSpec.aspx`.

Albers, L., L. Migliaccio, E. Bedrick, D. Teaf, P. Peralta. 2007. Does Epidural Analgesia Affect the Rate of Spontaneous Obstetric Lacerations in Normal Births? *Journal of Midwifery & Women's Health* **52**(1) 31–36.

Alfirevic, Z., D. Devane, G. Gyte. 2013. Continuous Cardiotocography (CTG) as a Form of Electronic Fetal Monitoring (EFM) for Fetal Assessment During Labour. *Cochrane Database of Systematic Reviews* (5).

AMA. 2013. CPT - Current Procedural Terminology. URL `http://www.ama-assn.org/go/cpt`.

Anim-Somuah, M., R. Smyth, C. Howell. 2011. Epidural Versus Non-Epidural or No Analgesia in Labour. *Cochrane Database of Systematic Reviews* (4).

Batt, R., C. Terwiesch. 2012. Doctors Under Load: An Empirical Study of State-Dependent Service Times in Emergency Care. The Wharton School, Working paper.

Berry-Jaeker, J., A. Tucker. 2013. An Empirical Study of the Spillover Effects of Workload on Patient Length of Stay. HBS Working Paper.

Bragg, F., D. Cromwell, L. Edozien, I. Gurol-Urganci, T. Mahmood, A. Templeton, J. van der Meulen. 2010. Variation in Rates of Caesarean Section Among English NHS Trusts After Accounting for Maternal and Clinical Risk: Cross Sectional Study. *BMJ* **341**.

Breen, R., K. Karlson, A. Holm. 2013. Total, Direct, and Indirect Effects in Logit and Probit Models. *Sociological Methods & Research* **42**(2) 164–191.

Chan, C., G. Escobar, S.-H. Kim, M. Olivares. 2012. ICU Admission Control: An Empirical Study of Capacity Allocation and Its Implication on Patient Outcomes. Columbia Univeristy Working paper.

Cho, S.-H., S. Ketefian, V. Barkauskas, D. Smith. 2003. The Effects of Nurse Staffing on Adverse Events, Morbidity, Mortality, and Medical Costs. *Nursing Research* **52**(2) 71–79.

Curtis, L. 2012. Unit Costs of Health & Social Care 2012. Tech. rep., Personal Social Services Research Unit.

de Leeuw, J., P. Struijk, M. Vierhout, H. Wallenburg. 2001. Risk Factors for Third Degree Perineal Ruptures During Delivery. *British Journal of Obstetrics and Gynaecology* **108**(4) 383–387.

Deo, S., I. Gurvich. 2011. Centralized vs. Decentralized Ambulance Diversion: a Network Perspective. *Management Sci.* **57**(7) 1300–1319.

DH. 2013. Maternity Care Facilities. Tech. Rep. Health Building Note 09-02, Department of Health.

Donnelly, V., M. Fynes, D. Campbell, H. Johnson, R. O'Connell, C. O'Herlihy. 1988. Obstetric Events Leading to Anal Sphincter Damage. *Obstetrics & Gynecology* **92**(6) 955–961.

Duan, N. 1983. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association* **78**(383) 605–610.

Eason, E., M. Labrecque, G. Wells, P. Feldman. 2000. Preventing Perineal Trauma During Childbirth: a Systematic Review. *Obstetrics & Gynecology* **95**(3) 464–471.

Eskandar, O., D. Shet. 2009. Risk Factors for 3rd and 4th Degree Perineal Tear. *Journal of Obstetrics & Gynaecology* **29**(2) 119–122.

Fraser, M., S. Guo. 2009. Sample Selection and Related Models. *Propensity Score Analysis*. Sage Publications, Inc., 85–125.

Green, L., N. Liu. 2013. Capacity Planning for Hospital Obstetrics Units: Insights From a New York City Study. Columbia University Working Paper.

Green, L., S. Savin, N. Savva. 2013. "Nursevendor Problem": Personnel Staffing in the Presence of Endogenous Absenteeism. *Management Sci.* **59**(10) 2237–2256.

Greene, W. 2002. *Econometric Analysis*. 5th ed. Prentice Hall.

Griffiths, A., N. Hikary, A. Sizer. 2002. Induction to Delivery Time Interval in Patients with and Without Preeclampsia: a Retrospective Analysis. *Acta Obstetricia et Gynecologica Scandinavica* **81** 867–869.

Guerin-Calvert, M. 2011. Assessment of Cost Trends and Price Differences for U. S. Hospitals. Tech. Rep. Working Paper No. 294, Federal Trade Commission.

Handa, V., B. Danielsen, W. Gilbert. 2001. Obstetric Anal Sphincter Lacerations. *Obstetrics & Gynecology* **98**(2) 225–230.

Henderson, J., R. McCandlish, L. Kumiega, S. Petrou. 2001. Systematic Review of Economic Aspects of Alternative Modes of Delivery. *British Journal of Obstetrics and Gynaecology* **108**(2) 149–157.

Hopp, W., S. Iravani, G. Yuen. 2007. Operations Systems with Discretionary Task Completion. *Management Sci.* **53**(1) 61–77.

HSC. 2013. OPCS-4 Classification. URL `http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4`.

Hurst, J., S. Williams. 2012. Can NHS Hospitals Do More with Less? URL `http://www.nuffieldtrust.org.uk/publications/can-nhs-hospitals-do-more-less`.

Kane, R., T. Shamliyan, C. Mueller, S. Duval, T. Wilt. 2007. The Association of Registered Nurse Staffing Levels and Patient Outcomes: Systematic Review and Meta-Analysis. *Medical Care* **45**(12) 1195–1204.

KC, D., C. Terwiesch. 2009. Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations. *Management Sci.* **55**(9) 1486–1498.

KC, D., C. Terwiesch. 2012. An Econometric Analysis of Patient Flows in the Cardiac Intensive Care Unit. *Manufacturing Service Oper. Management* **14**(1) 50–65.

Kohler, U., K. Karlson. 2013. KHB: Stata Module to Decompose Total Effects Into Direct and Indirect via KHB-Method. Working paper.

Kuntz, L., R. Mennicken, S. Scholtes. 2014. Stress on the Ward: Evidence of Safety Tipping Points in Hospitals. *Management Science (Forthcoming)* .

Lee, L., R. Trost. 1978. Estimation of Some Limited Dependent Variable Models with Application to Housing Demand. *Journal of Econometrics* **8**(3) 357–382.

Lilley, R. 2003. *The Insider's Guide to the NHS: How It Works and Why It Sometimes Doesn't*. Radcliffe Publishing.

Liu, E., A. Sia. 2004. Rates of Caesarean Section and Instrumental Vaginal Delivery in Nulliparous Women After Low Concentration Epidural Infusions or Opioid Analgesia: Systematic Review. *BMJ* **328** 1410–1415.

MacKinnon, D., J. Dwyer. 1993. Estimating Mediated Effects in Prevention Studies. *Evaluation Review* **17**(2) 144–158.

Maddala, G. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York.

Mare, R., C. Winship. 1988. Endogenous Switching Regression Models for the Causes and Effects of Discrete Variables. J. Long, ed., *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research*. Sage Publications, Newbury Park, CA, 132–160.

MCWP. 2007. Making Normal Birth a Reality. URL `http://www.nct.org.uk/about-us/what-we-do/policy/normalbirth`.

Monfardini, C., R. Radice. 2008. Testing Exogeneity in the Bivariate Probit Model: A Monte Carlo Study. *Oxford Bull Econ & Stats* **70**(2) 271–282.

NAO. 2013. Maternity Services in England. Tech. rep., National Audit Office.

Nasr, A., K. Reichardt, K. Fitzgerald, M. Arumugusamy, P. Keeling, T. Walsh. 2004. Impact of Emergency Admissions on Elective Surgical Workload. *Irish Journal of Medical Science* **173**(3) 133–135.

Needleman, J., P. Buerhaus, S. Mattke, M. Stewart, K. Zelevinsky. 2002. Nurse-Staffing Levels and the Quality of Care in Hospitals. *N. Engl. J. Med.* **346**(22) 1715–1722.

Needleman, J., P. Buerhaus, V. Pankratz. 2011. Nurse Staffing and Inpatient Hospital Mortality. *N. Engl. J. Med.* **364**(11) 1037–1045.

NICE. 2012. Caesarean Section. Tech. Rep. Clinical guideline 132, National Institute for Health and Clinical Excellence.

Oliva, R., J. Sterman. 2001. Cutting Corners and Working Overtime: Quality Erosion in the Service Industry. *Management Sci.* **47**(7) 894–914.

Papile, L.-A. 2001. The Apgar Score in the 21st Century. *New England Journal of Medicine* **344**(7) 519–520.

Poirier, D. J., P. A. Ruud. 1981. On the Appropriateness of Endogenous Switching. *Journal of Econometrics* **16**(2) 249–256.

Powell, A., S. Savin, N. Savva. 2012. Physician Workload and Hospital Reimbursement: Overworked Physicians Generate Less Revenue per Patient. *Manufacturing Service Oper. Management* **14**(4) 512–528.

RCOG. 2011. Operative Vaginal Delivery. Tech. Rep. Green-top Guideline No. 26, Royal College of Obstetricians and Gynaecologists.

Renfrew, M., W. Hannah, L. Albers, E. Floyd. 1998. Practices That Minimize Trauma to the Genital Tract in Childbirth: A Systematic Review of the Literature. *Birth* **25**(3) 143–160.

Roodman, D. 2011. Estimating Fully Observed Recursive Mixed-Process Models with cmp. *Stata Journal* **11**(2) 159–206.

Staiger, D., J. Stock. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica* **65**(3) 557–586.

Stock, J., M. Yogo. 2005. Testing for Weak Instruments in Linear IV Regression. D. Andrews, J. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press, 80–108.

Tan, F., S. Netessine. 2012. When Does the Devil Make Work? An Empirical Study of the When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Server's Performance. INSEAD Working paper.

WHO. 2011. International Statistical Classification of Diseases and Related Health Problems. Tech. Rep. 10th Revision, World Health Organization.

Wilde, J. 2000. Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors. *Economics Letters* **69**(3) 309–312.

Williams, A., D. Tincello, S. White, E. Adams, Z. Alfirevic, D. Richmond. 2005. Risk Scoring System for Prediction of Obstetric Anal Sphincter Injury. *British Journal of Obstetrics and Gynaecology* **112**(8) 1066–1069.

Winship, C., R. Mare. 1983. Structural Equations and Path Analysis for Discrete Data. *The American Journal of Sociology* **89**(1) 54–110.

Wood, S. 2011. Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semi-parametric Generalized Linear Models. *Journal of the Royal Statistical Society (B)* **73**(1) 3–36.

Wood, S. 2013. mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation. URL `http://cran.r-project.org/package=mgcv`.

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.