Discussion Papers in Economics

No. 07/2017

Is it good to be too light?

Birth weight thresholds in hospital reimbursement systems

Simon Reif
University of Erlangen-Nürnberg

Sebastian Wichert
Ludwig-Maximilians-University Munich

Amelie Wuppermann
Ludwig-Maximilians-University Munich

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/research/iwf-discussion-paper-series/

Is it good to be too light?

Birth weight thresholds in hospital reimbursement systems

Simon Reif[*], Sebastian Wichert[†], Amelie Wuppermann[‡]

March 2017

Abstract

Birth weight manipulation is common in per-case hospital reimbursement systems, in which hospitals receive more money for otherwise equal newborns with birth weight just below compared to just above specific birth weight thresholds. As hospitals receive more money for cases with weight below the thresholds, having a (reported) weight below a threshold could benefit the newborn. Also, these reimbursement thresholds overlap with diagnostic thresholds that have been shown to affect the quantity and quality of care that newborns receive. Based on the universe of hospital births in Germany from the years 2005–2011, we investigate whether weight below reimbursement relevant thresholds triggers different quantity and quality of care. We find that this is not the case, suggesting that hospitals' financial incentives with respect to birth weight do not directly impact the care that newborns receive.

JEL Codes: I110, I180,

Keywords: neonatal care, DRG upcoding, quantity & quality of care

[*]FAU Erlangen-Nuremberg, Findelgasse 7, 90402 Nürnberg, Germany (simon.reif@fau.de)

[†]Ludwig-Maximilians-University Munich, Germany & MC-Health (sebastian.wichert@econ.lmu.de)

[‡]Ludwig-Maximilians-University Munich, Germany (amelie.wuppermann@econ.lmu.de)

We are grateful for helpful comments and suggestions by Tom Crossley, Michael Gerfin, Hendrik Jürges, Harald Tauchmann, and Joachim Winter. Participants of the *2nd German Health Econometrics Group Workshop* in Essen, the *Workshop on Early Care Interventions and their Effects on Children and Families* in Aarhus, the *SSPH Doctoral Workshop in Health Economics and Policy* in Lucerne, the *7th Annual Conference of the German Health Economics Association* in Bielefeld, the *11th European Conference on Health Economics* in Hamburg and the *ISER seminar in Colchester* provided valuable feedback. We also want to thank Melanie Scheller for assistance with data access. Funding through the International Doctoral Program "Evidence-Based Economics" of the Elite Network of Bavaria is gratefully acknowledged. Christina Schramm, Katrin Poschen and Katrin Ziegler provided valuable research assistance.

1 Introduction

Small changes in birth weight can have important financial implications for hospitals in many of the widespread prospective payment schemes (PPS) which reimburse hospitals with a fixed rate for the treatment of strictly defined diagnosis related groups (DRGs). More specifically, hospitals receive a higher reimbursement for newborns with birth weight just below certain thresholds than for newborns with weight above, leading to an incentive to under-report birth weight. The evidence is accruing that the introduction of birth weight thresholds has led to large under-reporting – so called upcoding – of birth weight (e.g. Jürges and Köberlein 2015; Shigeoka and Fushimi 2014).

At the same time, birth weight thresholds are used to diagnose newborns as having "extremely low" (weight $\leq$ 1000g), "very low" (weight $\leq$ 1500g), or "low" birth weight (weight $\leq$ 2500g) and appear in medical guidelines. Despite the fact that low birth weight is typically linked to worse health outcomes (see e.g. Hack et al. (2002), Hummer et al. (2014)), newborns with weight just below the 1500g diagnostic threshold have been found to have higher survival chances than newborns with weight just above (Almond et al. 2010; Bharadwaj et al. 2013; Breining et al. 2015). To the extent that reimbursement-relevant birth weight thresholds are identical to the diagnostic thresholds, the practice of upcoding newborns' weight may thus be to the benefit of the newborn. Furthermore, as the hospital receives more money for newborns with weight below the reimbursement relevant thresholds, they may also be able or willing to deliver additional care to these newborns.

In this paper, we investigate whether newborns benefit from having a reported birth weight below reimbursement relevant weight thresholds in the German DRG system. Based on an administrative hospital claims data set covering the universe of hospital births in Germany in the years 2005–2011, we compare the survival chances and the quantity of care that newborns with weight just below the relevant thresholds receive to those of newborns with weight just above the thresholds. We include all eight reimbursement relevant thresholds in the German DRG system. While some of these thresholds overlap with the diagnostic thresholds and/or are explicitly mentioned in medical guidelines, others are only relevant for reimbursement, allowing us to shed light on the importance of diagnostic thresholds, medical guidelines, and reimbursement for the care of newborns.

Different from the settings in the earlier literature on the effects of diagnostic thresholds (Almond et al. 2010; Bharadwaj et al. 2013; Breining et al. 2015), the fact that the diagnostic thresholds are relevant for reimbursement in our setting imposes a challenge to the empirical analysis: If the decision to upcode a newborn's weight below a threshold depends on the newborn's health status, the crucial assumption that newborns with reported weight above and below the threshold have ex ante similar health is not plausible, an issue that Barreca et al. (2011) already raised in the light of the earlier literature. As Jürges and Köberlein (2015) show, it is likely that birth weight manipulation in German hospitals is not random. To the contrary, hospitals primarily tend to upcode newborns for whom staff expects more care. These are relatively fragile newborns that still have non-negligible survival chances and will therefore receive a lot of treatment. We take three steps in our analysis to take this into account: First, we control for a large set of variables capturing a newborn's health at birth. We specifically choose variables such as sex and plurality of births that are not easy to manipulate and observable to the hospital staff who reports newborns' weight. These are variables that may influence reported birth weight. Second, we restrict our analysis to newborns who survive the first day of their life, thus excluding those very fragile cases for whom hospital staff may expect an early death, making upcoding of birth weight not worthwhile. In a third set of results, we further include hospital fixed effects in our estimations, taking into account possible differences in coding practices and treatment across hospitals.

Our findings show that without controls and not restricting to first day survivors newborns with weight below almost all of the eight thresholds stay in the hospital longer, receive more procedures, and have lower mortality (during the hospital stay). However, for all but the highest two thresholds (2000g and 2500g) these results become insignificant or even change sign, when we control for health at birth, exclude newborns who die on the first day of life, and include hospital fixed effects. We interpret this as evidence that neither reimbursement differences, nor the diagnostic threshold of 1500g or thresholds in medical guidelines trigger additional care or reduce mortality among newborns in Germany. While our results may indicate that newborns with weight just below 2000g and 2500g benefit from being below these thresholds, we believe that the differences rather reflect systematic upcoding related to factors that we cannot control perfectly. As further discussed in the final section of this paper, we therefore conclude that DRG-related upcoding in Germany does not affect the care upcoded newborns receive.

Our paper contributes to and brings together two strands of literature. The first focuses on the question whether hospitals upcode diagnoses or other health measures, such as birth weight, to generate higher payments in DRG reimbursement systems. Concerning the upcoding of diagnoses, Dafny (2005) shows that hospitals reacted to a recalibration of Medicare DRGs in 1988 by disproportionally shifting patients to diagnoses codes that became more lucrative. At the same time, she finds no changes in the treatment that patients receive nor in patient mortality. Silverman and Skinner (2004) focus on Medicare patients with respiratory disease and show that the share of patients coded to the highest paying DRG increases significantly over time, particularly so in the group of for-profit hospitals. Furthermore, there is evidence that also hospitals in the Italian region of Emilia-Romagna (Verzulli et al. 2016), Portugal (Barros and Braun 2016) and Norway (Januleviciute et al. 2016) upcode patients to the highest paying DRGs. Concerning upcoding of patient characteristics, Shigeoka and Fushimi (2014) show that hospitals in Japan have manipulated birth weight as a response to the introduction of a partial PPS in a way that increased hospital payments. Similarly, Jürges and Köberlein (2015) find that German hospitals reacted to the introduction of the DRG payment system in 2003 by systematically under-reporting newborns' weight.

The second strand of literature focuses on the effect of birth weight thresholds on the quantity and quality of care that newborns receive. Based on the census of U.S. births, Almond et al. (2010, 2011) find that newborns with weight just below the very low birth weight threshold at 1500g have higher survival chances than newborns with weight above. Based on hospital discharge data for five states they further find that birth weight below 1500g triggers additional care that also results in higher hospital charges. Their results are particularly concentrated among low quality hospitals, i.e., those hospitals that offer no or only low levels of neonatal intensive care. Similar effects of the very low birth weight thresholds have been found for newborns in Chile where medical guidelines explicitly recommend different treatment depending on a very low birth weight diagnosis (Bharadwaj et al. 2013). Breining et al. (2015) focus on Denmark where the treatment recommendations in medical guidelines only vary across the very low birth weight thresholds for newborns with at least 32 weeks of gestation. They find that indeed only for newborns born at 32 weeks of gestation or more, treatment depends on birth weight. For newborns born earlier in the pregnancy they find no treatment differences with birth weight, indicating that medical guidelines have an impact on the care that newborns receive.

Our paper brings these two strands of the literature together in investigating whether newborns benefit from upcoding of birth weight below thresholds that may themselves affect the care that newborns receive as they are diagnostic thresholds. Our analysis contributes to the first strand of the literature by investigating effects of upcoding on the care the patients receive. We add to the second strand by focusing not only on the very low birth weight threshold but also on other thresholds that impact diagnoses and appear in medical guidelines.

The rest of the paper is structured as follows: In section 2, we give an overview on the institutional background in Germany and discuss which different birth weight thresholds may be related to the quantity and quality of care that newborns receive. We introduce our data in section 3 and the empirical strategy in section 4. In section 5, we present our results, the sensitivity of which we explore in section 6. The paper closes with a discussion and conclusion in section 7.

2 Institutional Background

In this section we give a brief overview on the German DRG (G-DRG) system and then describe reasons why the treatment that newborns receive in German hospitals may vary around birth weight thresholds.

2.1 The G-DRG reimbursement system

Until the year 2003, German acute care hospitals were reimbursed through a multiple-source-system consisting of a hospital specific patient/day base-rate, a ward specific rate and case-based lump-sums. In order to increase efficiency and, in particular, reduce length of stay, the G-DRG system was introduced. Under the old and the new system, reimbursement was and is generally the same irrespective of individuals' insurance status (public or private). Based on the Australian DRG system, 664 DRGs defined by combinations of diagnoses, performed procedures, hours of ventilation, age and (for perinatal DRGs) birth weight were set up. To each DRG a case-weight was assigned which determined the final reimbursement. This case-weight multiplied by a base-rate gives the amount of money a hospital receives for treating the respective patient.

In a transition period from 2003 to 2010 hospital specific base-rates were used. These rates advanced towards state specific rates over the years.[1] Since 2010 most hospitals are reimbursed according to this state base-rate. From 2010 to 2014 these state specific base rates in turn narrowed down to a federal base rate interval (Salm and Wübker 2015).

Case-weights and group definitions vary from year to year. The German Institute for Hospital Reimbursement[2] decides on changes of case-weights and DRG definitions every year using information on actual costs of different treatments reported voluntarily by a group of hospitals. Since the introduction of the G-DRG system, the number of DRGs almost doubled to 1200 in 2015.[3]

[1] The reason for starting with hospital specific rates was that hospital cost structures were drastically different when the G-DRG system was introduced. Starting off directly with identical reimbursement in each state would have led to major financial struggles for some hospitals.

[2] Institut für das Entgeldsystem im Krankenhaus (InEK), financed by private and social health insurers as well as the German Hospital Association.

[3] For an analysis of these changes see Schreyögg et al. (2014).

The change over time in the number of DRGs in neonatal care, however, was not as drastic. When DRGs were first put into use in Germany in 2003, 38 different groups for the treatment of newborns were defined. Until 2015 this number has increased only slightly to 42 groups. The majority (37/42) of these groups depend on birth weight. The remaining 5 DRGs are for newborns that die within the first four days of their life and for specific cardiovascular conditions. The 37 birth-weight related DRGs depend on eight birth weight thresholds. Within each birth weight interval multiple DRGs cover different degrees of treatment complexity and complications. For our main analysis, we investigate differences across these thresholds.

In order to compare only similar cases left and right of the thresholds we also conduct analyses within "severity groups" across thresholds. To this end we group the DRGs into six severity groups that have the same level of complications and complexity based on the DRG catalog. Within each severity group birth weight should be the only difference across thresholds. Table A.1 (in the Appendix) shows which DRGs we assign to the different severity groups in the different birth weight intervals. Group 1 in Table A.1 comprises the least severe cases, group 6 the most severe ones. At very low birth weights there are only two different groups of severity. The first contains newborns with lower levels of complications and less severe health conditions. This group splits up into four groups for higher birth weight intervals (group 1-4). The second group contains more severe cases and splits up into two groups for higher birth weights (group 5/6). Starting from a birth weight of 1000g, three severity groups apply as the joint group 5 and 6 is split into two severity groups. At birth weights of 1500g and higher all 6 groups are present.

2.2 Treatment and birth weight

Broadly speaking, there are two reasons for differences in treatment of newborns across birth weight intervals. As explained in more detail below, the first is that DRGs are defined along birth weight intervals, leading to sharp increases in reimbursement when birth weight crosses from above to below thresholds and hence more funds to spend on costly procedures. The second reason is that birth weight thresholds are used in the diagnosis of newborns, define the level of specialization of the hospital that the newborn should be treated in, and appear in medical guidelines.

Thresholds and reimbursement

Figure 1 provides a systematic overview on changes in case-weights with birth weight within the six DRG severity groups defined above, i.e. holding everything else except for birth weight constant. Generally, the case-weight decreases at 600g, 750g, 875g, 1000g, 1250g, 1500g, 2000g, and 2500g. Reimbursement for newborns with weight just below the thresholds is thus generally higher than for newborns with weight on or above the thresholds. The sizes of the jumps in case-weights (and thus in reimbursement), however, vary across groups and within groups across thresholds. The 2500g threshold for example is not relevant for reimbursement for newborns in the least severe group (Group 1), while the 750g threshold only hardly matters for the cases with most severe conditions (Group 5/6).

– FIGURE 1 ABOUT HERE –

As an example for how small changes in birth weight can affect reimbursement let's consider a notional hospital in the German state of Hesse in the year 2010 which had a base-rate equal to the state base-rate of €2,968.56. A generic newborn in this hospital may have had a birth weight of 1004g and was assigned DRG P63Z[4] which in this year had a case-weight of 8.776. Total reimbursement to the hospital would have been €26,052.08. Had the weight instead been 999g, DRG P62D with a case-weight of 15.51 would have been assigned. This means the hospital would have received €46,042.37, a plus of almost €20,000 for a 5g change in birth weight.

As the financial incentives apply to the hospital, while treatment is typically decided on by the medical personnel, one might argue that these reimbursement differences should not play a large role for medical decision making. However, Jürges and Köberlein (2015) have shown that reimbursement differences by birth weight have led to manipulation of reported weight. As birth weight is reported by medical personnel, the individuals who decide on treatment are aware of the financial importance of reimbursement differences. If they are also informed about the costs of different procedures, they may also take these differences in reimbursement into account when deciding on which treatment newborns receive.

Thresholds and medical guidelines

In addition to reimbursement differences, birth weight determines the level of specialization of the hospital in which newborns should be treated. In Germany, four degrees of specialization are defined by the Gemeinsamer Bundesausschuss (2013)[5]: Level 4 indicates regular maternity clinics, Level 3 means that basic neonatal care can be provided, Level 2

[4] For example, because it was a girl, ventilation was in use for 48 hours, respiratory distress syndrome (ICD-10-GM: P28.5), systematic inflammatory response-syndrome (ICD-10-GM: R65.0), and some other infectious disease (ICD-10-GM: P37.9) were diagnosed and the girl was treated by monitoring cardiovascular levels (OPS: 8-930), was given a VR-infusion (OPS: 8-811.0), CAPA ventilation was performed (OPS: 8-711.00) and after this did not work out, endotracheal intubation (OPS: 8-701) came into use.

[5] The highest self administration unit in the German health sector which consists of 13 members representing health insurers, providers as well as the general public.

clinics have some specialization in neonatal care, and Level 1 clinics maintain neonatal intensive care units (NICUs). Pre-term births with weight above 1500g should be treated in clinics of at least Level 3, birth weight between 1250 and 1500g induces Level 2 care, while newborns below 1250g should be treated in Level 1 clinics.[6] [7] To the extent that better care is provided in more specialized hospitals, birth weight just below the thresholds that require care in more specialized hospitals, in particular below 1250g could be beneficial for newborns.

As results in the earlier literature suggest (Almond et al. 2010; Bharadwaj et al. 2013), further differences in treatment around birth weight thresholds could result as a consequence of thresholds in medical guidelines or as thresholds determine diagnoses, such as the international standard classifications of extremely low (below 1000g), very low (below 1500g), and low birth weight (below 2500g). The German Association of the Scientific Medical Societies (AMWF) publishes a multitude of medical guidelines for neonatal care. They range from guidelines concerning the care of very frail newborns to care for healthy newborns. Additionally, there are guidelines on after-hospital care for groups of newborns with specific conditions. Except for the 875g and 2000g thresholds, all reimbursement relevant thresholds are mentioned in at least one of the guidelines concerning neonatal care. The threshold that appears most often is 1500g, the diagnostic threshold for very low birth weight. Only few guidelines, however, give specific recommendations on care depending on birth weight. One example that does is the guideline on parenteral nutrition. Among other things, it recommends specific nutritional solutions for newborns with weight below 1500g (AWMF 2014). Most recommendations seem to be linked to gestational age rather than weight.

Although they do not contain specific treatment recommendations, several guidelines report risks of specific diseases or survival chances for birth weight intervals. For example, the risks of complications like necrotizing enterocolitis (11% for 401-750g, 9% for 751-1000g, 6% for 1001-1250g, and 4% for 1251-1500g) and patent ductus arteriosus (30% for pre-term birth with weight below 1500g, and 50-70% for newborns below 1000g) are reported for birth weight intervals. Furthermore, survival chances for esophageal atresia are reported separately for newborns with weight above or below 1500g, with higher chances

[6] Interestingly, the *German Association for Perinatal Medicine* uses different Level labels (1, 2a, 2b and 3) and slightly different thresholds which would admit less newborns to NICUs (Bauer et al. 2006). Nevertheless, the Gemeinsamer Bundesausschuss (2013) regulations are binding.

[7] Theoretically this structure could offer an incentive for hospitals to increase birth weight of newborns that would be too light for this type of clinic and would be transferred otherwise. Since case-weights are calculated for average newborns within a birth weight bracket, reimbursement would most likely not cover all the costs for newborns at the lower end of a birth weight DRG, making over-reporting of weight to keep newborns implausible (Jürges and Köberlein 2015).

for heavier newborns (AWMF 2012). Although risk classifications are no direct recommendations on care, they may result in treatment differences, to the extent that they influence physicians' awareness of risks for newborns in the different weight groups.

As some thresholds are only relevant for reimbursement (875g, 2000g), while others affect reimbursement as well as the type of hospital that a newborn should be treated in (1250g), or overlap with diagnostic thresholds and are mentioned in medical guidelines (600g, 750g, 1000g, 1500g, 2500g), we can shed light on the importance of the drivers of differences in medical care around the thresholds by comparing effects across thresholds.

3 Data

Our analyses are based on the universe of German hospital claims from the years 2005–2011. All German hospitals have to submit their DRG claims to the Institute for Hospital Reimbursement (InEK). InEK forwards parts of the data to the German Federal Statistical Office, which makes the data available to researchers.[8]

For each of the roughly 20 million hospital stays per year in Germany, the data contain basic demographic information on the patients, as well as detailed information on the hospital stay. In addition to the exact DRG used for reimbursement, the data contain information on the reason for the hospital stay, the length of the stay, procedures performed during the stay (the German version of ICPM codes, called OPS codes), diagnoses (ICD-10-GM) codes, and the reason of discharge (including regular discharge, death or referral to a different hospitals among others). For newborns, the data further contain information on birth weight.

In our analyses, we restrict the data to all births (cases with a perinatal DRG (those beginning with "P") and "birth" as cause of admission) with valid information on newborns' sex and a birth weight between 450 and 3000g.[9] After applying these selection criteria our sample amounts to a total of 985,885 cases.

Table 1 displays basic descriptive statistics for our data, averaged across all cases with birth weight 450-3000g and separately for the different birth weight intervals that are relevant for reimbursement (< 600g, 600-749g, 750-874g, 875-999g, 1000-1249g, 1250-1499g, 1500-1999g, 2000-2499g, 2500-3000g). It contains information on birth weight, on quality and quantity of care (length of the hospital stay, number of procedures, mortality), on the newborn's health at birth, as well as hospital related control variables.

[8] For further information on the data see www.forschungsdatenzentrum.de/en/database/drg.

[9] By selecting only cases with "birth" as cause of admission, we exclude cases that are re-admitted after a first discharge. Since patients cannot be linked across hospital stays, this is necessary to avoid double-counting newborns that leave the hospital and are re-admitted later.

– TABLE 1 ABOUT HERE –

On average, newborns have a birth weight of 2601g, have 2.6 reported procedures, and stay in the hospital for 8.5 days. Only 0.6% of newborns die during the hospital stay[10]. Birth weight increases mechanically across the birth weight intervals. At the same time quantity of care and mortality generally decrease. The 600g threshold constitutes an exception: Newborns with weight below 600g receive less care than newborns in the next interval. This may be explained by the fact that in the lighter group of newborns a larger share dies very early on (34% die on their first day of life, compared to 13% in the group with weight 600-749g), likely leading to shorter hospital stays and fewer procedures for the group with weight below 600g.

To measure health at birth we use variables that are visible at birth and known health risk factors for newborns. These are important variables for our analysis as they could influence the reporting of birth weight as well as the care that newborns receive in the hospital. In addition to sex, pre-maturity, and plurality of birth (which are mentioned explicitly as risk factors for newborns' survival in medical guidelines (AWMF 2007)), we focus on two conditions that usually become apparent during or right after delivery and are leading causes of death among pre-term births: asphyxia and infant respiratory distress syndrome (IRDS). Furthermore, we include several indicators for maternal health-behavior during pregnancy, namely whether the mother smoked, drank alcohol, or took illegal drugs.[11]

In our data, 44% of all newborns are boys[12], 19% of births are premature, i.e., born with at least 28, but less than 37 completed weeks of gestation (ICD-10-GM: P07.3). Moreover 1.2% are extremely premature, i.e., have a gestational age of less than 28 weeks (ICD-10-GM: P07.2). 11% of newborns are twins. 0.4% are higher order births.

Asphyxia is a form of oxygen deprivation that occurs during delivery. The ICD-10-GM coding system contains three different diagnosis codes for asphyxia, one for newborns with severe asphyxia and a 1-minute APGAR score[13] of 0-3 (ICD-10-GM: P.21.0), one for

[10] We derive the information on whether a newborn dies from the reason of discharge.

[11] Unfortunately, linking the data on newborns to their mothers – who constitute a separate hospital case – is impossible. Most procedures concerning delivery - except for those specifically undertaken for the newborn – are billed as the mother's case, not the newborn's case. Thus, the data contain almost no information on what happened during pregnancy or delivery. To the extent that this information determines the newborn's reported weight and later treatment, this is valuable information that we unfortunately do not have in the data. To account for this as best as we can, we proxy for mothers' behavior during pregnancy by relying on diagnoses that the newborn receives.

[12] This unusually low number may result from the high share of male still-births among low birth weight newborns. Still-births are not included in our data

[13] The APGAR score was developed in the 1950s to summarize a newborn's health. It measures **A**ppearance (skin color), **P**ulse (heart rate), **G**rimace (reflex irritability), **A**ctivity (muscle tone), and **R**espiration.

newborns with mild asphyxia and a 1-minute APGAR score of 4-7 (ICD-10-GM: P.21.1), as well as asphyxia without additional classification of severity (ICD-10-GM: P.21.2). Among all newborns in our data 2.1% had general asphyxia, 1.6% had mild asphyxia, and 0.5% severe asphyxia. IRDS (ICD-10-GM: P22.-) is a lung malfunction which is common among preterm births and begins shortly after birth. About 10.6% of all newborns in our data suffer from IRDS.

Only 1.2% of newborns in our data are (suspected to be) affected by maternal use of tobacco (ICD-10-GM: P04.2). This number is much lower than prevalences of smoking during pregnancy as reported in other sources. Kuntz and Lampert (2016), for example, report a prevalence of 12.1% for children born in Germany in the years 2003–2012 based on survey data. As only severe cases of smoking during pregnancy may be apparent to the hospital staff at the time of delivery, it is likely that only severe cases of smoking are flagged in our data. To the extent that newborns diagnosed with suspected damage due to maternal smoking are those for whom the medical personnel expects higher needs of care during the hospital stay and thus may also manipulate the weight, the indicator available to us flags the relevant cases. On average, there are less than 0.1% of newborns for whom alcohol use (ICD-10-GM: P04.3) and 0.2% for whom drug use during pregnancy (ICD-10-GM: P04.4) are reported. The shares are higher (0.1% for alcohol use and 0.3 to 0.5% for drug use) for lower birth weight intervals. Similar to maternal smoking, we believe that conditions related to maternal alcohol or drug use are coded for severe cases of use and thus flag relevant cases.

The third part of Table 1 shows three measures that capture information on the hospital as well as the medical personnel that delivered the treatment. The first variable is an indicator for whether a newborn was treated in an inpatient ward, defined as a department of the hospital where the staff (doctors, nurses, and midwives) are self-employed rather than employed by the hospital. Even if newborns are not treated in an inpatient ward, care can be partly delivered by self-employed doctors or midwives. Whether a newborn was treated by a self-employed (inpatient) physician or self-employed (inpatient) midwife is captured by the second and third variables. Information on the place of treatment and the doctors and midwives that delivered the treatment is important for our analysis as the incentives for birth-weight manipulation are slightly smaller in inpatient wards and as it is possible that treatment differs when delivered by inpatient doctors and midwives who generally also oversee outpatient treatment of the patient. Overall, 9.5% of newborns are treated exclusively in inpatient wards and by inpatient doctors; 7.1% received care from inpatient midwives.[14] Although inpatient wards and inpatient doctors and midwives are

[14] As midwives are mainly present in the hospital during delivery (and delivery is billed on the mother

allowed to treat all newborns irrespective of their birth weight, only few inpatient wards are specialized in treating low birth weight newborns. As a consequence, treatment in inpatient wards and by inpatient doctors is rare for newborns with birth weight below 1500g (less than 1% of cases). Even among newborns with weight between 1500g and 1999g only 1.5% are exclusively treated in inpatient wards, compared to 6.3% for those with weight between 2000g and 2499g, and 11.3% in the highest birth weight interval in our data (2500g-3000g).

– FIGURE 2 ABOUT HERE –

Figure 2 provides additional information on the distribution of birth weight in our data. It contains birth weight frequencies pooled across the years 2005–2011. The eight birth weight thresholds that determine DRGs are indicated by the vertical lines. Figure 2 shows statistically implausibly large increases in frequencies slightly below the thresholds, especially at 1250g, 1500g, 2000g, and 2500g. These increases are in line with the findings by Jürges and Köberlein (2015) and indicate that hospitals under-report birth weight.

4 Empirical strategy

We aim at investigating whether newborns benefit from having a (reported) birth weight below birth weight thresholds defined by the DRG system, medical guidelines or a combination thereof. We therefore document differences in quantity and quality of care comparing newborns with weight just below and just above the eight reimbursement relevant thresholds in Germany. To this end we start by estimating mean differences in 25g weight intervals above and below the thresholds. Quantity of care is measured by the length of the hospital stay and the number of procedures that newborns receive. Quality of care is measured by mortality (during the hospital stay). To exclude the possibility that differences across thresholds can be expected as birth weight is related to health and newborns with lower weight may thus simply need more care even if the difference in weight is not large, we also report results for four placebo thresholds that do not affect reimbursement and do not appear in medical guidelines.

Earlier studies on differences in birth weight thresholds have relied on the assumption that newborns left and right of the analyzed thresholds are similar in terms of health (e.g. Almond et al. 2010; Bharadwaj et al. 2013; Breining et al. 2015). The thresholds considered in these studies, however, were not relevant for reimbursement and thus did

not the newborn), fewer newborns are recorded as having been treated by inpatient midwives than are treated solely in inpatient wards or receive treatment by inpatient doctors.

not give hospitals a financial incentive to manipulate birth weight. As Figure 2 and Jürges and Köberlein (2015) (with another data source) show, birth weight manipulation is common in the German DRG system. The assumption that newborns left and right of the thresholds are similar in terms of required care and expected outcomes could therefore only hold if the manipulation of birth weight happened randomly, or at least independent of the newborns' health. Jürges and Köberlein (2015), to the contrary, suggest that manipulation is systematic: there seems to be more under-reporting of birth weight for newborns who are expected to need more care (those who are fragile but not likely to die very early on).

Raw differences in quantity and quality of care around the birth weight thresholds therefore do not constitute causal effects of having a weight below the threshold. To investigate whether at least part of observed difference is likely causal, we proceed by conducting parametric regressions, in which we subsequently add more control variables that account for differences in the likelihood of birth weight manipulation and may explain treatment differences. We implement the following parametric regression equation:

$$Y_i = \alpha + \beta \mathbb{1}(bw_i < T) + f(bw_i - T + 1) + \boldsymbol{\gamma} \boldsymbol{X_i} + u_i \ , \tag{1}$$

where Y_i is an outcome measure for the quality or quantity of care for newborn i, bw_i is the (possibly manipulated) birth weight, and $\mathbb{1}(bw_i < T)$ is an indicator for newborn i having a birth weight below the threshold T. The parameter of interest is β, which measures the difference in outcomes around the threshold. In our baseline regressions we include a (centered) second-order birth weight polynomial $f(bw_i - T + 1)$, fitted separately on both sides of the threshold and use all observations in a 100g bandwidth of the birth weight threshold. In robustness analyses, we vary the bandwidth as well as the order of the included polynomial. In an additional robustness analysis, we exclude newborns in 10g windows left and right of the thresholds – so called "donut regressions", as Barreca et al. (2011) have shown that the exclusion of newborns in small windows around the thresholds may have an influence on the results.

In our baseline analysis, we estimate four different specifications for each outcome. In the first, we estimate equation (1) using all observations with weight in the 100g bandwidth and without control variables $\boldsymbol{X_i}$. In the second, we include a vector of control variables, $\boldsymbol{X_i}$ to capture observable differences in the newborns' health at birth and hospital-level controls. If we observed all factors that trigger birth weight manipulation, $\widehat{\beta}$ estimated based on equation (1) including our control variables would measure the causal effect of having a weight below the different thresholds on the quantity and quality provided. As, however, it is unlikely that we can account for all factors that hospital staff observes at the time of birth, we conduct a third analysis in which we exclude all newborns who die on their first day of their life. All newborns who die during their first four days of life receive the same DRG irrespective of birth weight. There is thus no incentive to manipulate the weight for newborns for whom death can be expected at the time when birth weight is reported. Newborns who are expected to die are therefore likely over-represented just above the weight thresholds. This may explain lower mortality as well as higher treatment intensity below the thresholds. By excluding newborns who die during their first day of life, we aim at excluding this type of selective upcoding as a reason for differences in care around the thresholds.

In addition to selective upcoding of newborns, differences across hospitals are a possible driver of observed differences around thresholds: For example, it could be the case that well-managed hospitals understand the financial incentives stemming from the DRG system and thus manipulate birth weight, while at the same time they deliver good care. Newborns with (reported) weight below the thresholds would then over-proportionally be treated in hospitals with better care, which would result in average differences in outcomes across thresholds. To exclude this explanation for differences, we include hospital fixed effects in our fourth set of analyses. With the inclusion of hospital fixed effects, we make sure that treatment differences arise within hospitals, not across hospitals.

We further investigate drivers of observed differences in care across thresholds by conducting sub-group analyses splitting the sample into the different severity groups. To the extent that the severity groups succeed in holding health constant across thresholds, differences in received treatment and mortality across thresholds are likely driven by the difference in reimbursement or medical guidelines across thresholds.

To separate reimbursement effects and effects of diagnostic thresholds or medical guidelines, we compare effects across the different thresholds. There are two thresholds, 875g and 2000g that are neither diagnostic thresholds nor appear in any of the relevant medical guidelines. Differences across these thresholds are thus likely driven by reimbursement rather than by medical guidelines.

5 Results

– TABLE 2 ABOUT HERE –

Table 2 reports differences in means between groups of 25g intervals across the eight reimbursement-relevant birth weight thresholds (600g, 750g, 875g, 1000g, 1250g, 1500g, 2000g, 2500g), as well as for four thresholds that play no role in reimbursement or medical guidelines (700g, 1300g, 2200g, 2700g). In general, newborns with weight just below the relevant thresholds stay longer in the hospital and receive more procedures than their neighbors with weight on the other side of the thresholds. With the exception of an increase in mortality below 875g, the share of newborns who dies during the hospital stay and on their first day of life is smaller below than above the thresholds. At the 600g and 875g threshold hardly any of the differences are significantly different from zero. At the other thresholds most differences are significant.

At least part of the raw difference in length of stay and number of procedures, however, is likely explained by the differences in weight around the thresholds: At all placebo thresholds newborns below the threshold stay significantly longer in the hospital and receive more procedures. Controlling for birth weight as in equation (1) is thus crucial when analyzing differences in care around the thresholds.

Although differences in the quantity of care exist around placebo thresholds, there are no differences in mortality.[15] This contrasts with the significant decreases in mortality across the reimbursement relevant thresholds (except at 875g). As first day mortality varies even more strongly around the thresholds than overall mortality, a large part of the difference may be driven by selective upcoding.

The mean differences in health-related controls across the thresholds present additional evidence for selective upcoding. In particular at the highest threshold of 2500g, all variables except for extreme prematurity and maternal alcohol use indicate that newborns just below the threshold are in worse health than newborns above. The newborns just below are 3.1 percentage points more likely to be male, 3.3 percentage points more likely to be born prematurely, 3.1 percentage points more like to be twins and 0.1 percentage points more likely to be higher order births. They are more likely to have any type of asphyxia, IRDS or a suspected damage due to maternal smoking or drug use. Only few of these variables show significant differences around the placebo thresholds of 2200 or 2700g, and if the differences are significant they are much smaller in size. This suggests that hospital staff manipulates the weight of at risk newborns – possibly to cover the expected higher costs of treatment. At the other thresholds, the differences in health-related controls are not as drastic. However, they also point into the direction of selective upcoding in favor of the more fragile newborns, highlighting the possibility that selective upcoding may explain differences in treatment and mortality across thresholds.

– TABLE 3 ABOUT HERE –

[15] To the extent that lighter newborns have higher mortality risks ceteris paribus, the additional care that the lighter newborns receive may be effective in decreasing these additional risks.

Regression results to facilitate the interpretation of the raw differences in treatment and mortality around thresholds are displayed in Table 3. The first row for each outcome displays the estimated coefficient $\widehat{\beta}$ in equation (1) when only a second order polynomial in birth weight is included as a control. For the higher birthweight thresholds, this pattern can also be clearly seen by plotting average length of stay and performed procedures around the thresholds (Figures 3 and 4).

– FIGURE 3 ABOUT HERE –

The second row in Table 3 adds the controls for health at birth and hospital-level variables; the third row restricts the data to those newborns who survive their first day of life. The fourth row additionally adds hospital fixed effects to restrict the analysis to within hospital differences.

– FIGURE 4 ABOUT HERE –

In line with the comparison of mean differences across reimbursement-relevant and placebo thresholds in Table 2, adding controls for birth weight already reduces the number of significant differences in quantity of care. While in Table 2, 5 of 8 reimbursement-relevant thresholds showed significant differences in length of stay, and 6 of 8 in the number of procedures, only 4 of the 8 thresholds show significant differences when birth weight is controlled for in Table 3. Similarly, the differences at the placebo thresholds are rendered insignificant almost entirely when birth weight is controlled for. The inclusion of health-related and hospital-related controls in row (2) reduces the coefficients for differences in length of stay and number of procedures across the reimbursement-relevant thresholds by almost half on average. The coefficients for the placebo thresholds remain almost unchanged. When additionally restricting the analysis to first day survivors (row (3)), the coefficients are further reduced towards zero and only 3 out of 8 remain significant for length of stay and 4 of 8 for number of procedures. At least half of the difference in treatment across thresholds thus seems to reflect selective upcoding. When we focus at within hospital differences in row (4), only the differences across the highest weight thresholds (2000g and 2500g) remain positive and significantly different form zero. The differences around the other thresholds are not significant or sometimes even turn negative. Within hospitals, there thus seems to be very little difference in treatment of at risk newborns due to weight thresholds when observable health differences are controlled for. Only for heavier newborns (for whom treatment may not be as decisive as the risk of mortality is much lower) differences persist.

The share of newborns who die during the hospital stay is significantly lower below all but the 875g threshold when birth weight is included as control as can be seen in the first row of the mortality results presented in Table 3. The inclusion of health-related and hospital control variables shrinks some coefficients slightly towards zero but the differences at most thresholds remain statistically significantly different from zero. As row (3) of the results indicates, however, the observable differences in mortality are largely driven by first day mortality. When newborns who die on their first day of life are excluded from our data, only the differences at 750g and 2000g remain negative and statistically significantly different from zero. To the extent that medical personnel expects the early death of newborns who die during their first day of life and thus do not upcode their birth weight, these results suggest that selective upcoding of newborns drives almost all of the observed differences in mortality. These results remain almost unchanged when hospital fixed effects are included (row 4). Differences across hospitals thus do not seem to play a big role.

– TABLE 4 ABOUT HERE –

Table 4 displays the $\widehat{\beta}$s estimated for the different severity groups including polynomials in birth weight as well as all control variables and hospital fixed effects and excluding those newborns who die on their first day of life from the data. The results indicate that the differences in quantity of care observed for the higher thresholds (2000g and 2500g) in Table 3 are concentrated in severity groups 1 and 2, i.e., those newborns with the least complications and lowest level of severity. The differences within the higher severity groups are not statistically significant, which may, however, also reflect smaller sample sizes for the higher severity groups. Table 4 additionally shows significant differences in quantity of care for the severity groups around the 1000g threshold. At the same time, mortality is higher below the 1000g threshold in these severity groups. The reason for this is most likely that hospitals do get extra reimbursement for long ventilation hours if newborns are heavier than 1000g. Here, hospitals have an incentive *not* to upcode newborns that have a high survival probability with long expected ventilation hours.

Concerning the question whether reimbursement relevant thresholds alone trigger additional care for newborns with weight below the threshold, we look at the results for the 875g and 2000g thresholds that affect reimbursement but are not part of medical guidelines. For 875g, none of the differences are significantly different from zero – not even the raw differences in Table 2. For 2000g, some differences are significant and remain significant even in the specifications that take selective upcoding and differences across hospitals as well as (ex post) health severity into account in the subgroup analyses. However, there are no differences for the more severe cases in severity groups 4 to 6. While we thus cannot exclude that newborns with weight below the 2000g threshold receive some additional care due to the increase in reimbursement, this effect seems to be concentrated to newborns with the best health, suggesting that among those newborns who require care it is delivered independent of their weight. Furthermore, the results for the 1500g threshold – the reimbursement relevant threshold that is at the same time a diagnostic threshold and that appears most often in medical guidelines – indicates that adding the financial incentives to the possible medical indication to additionally treat newborns with weight below this threshold has likely no detrimental effect on newborns with weight above the threshold.

Overall, the results presented in this section show that the differences in treatment across birth weight thresholds are mainly driven by selective upcoding as are differences in mortality. Differences in treatment, however, remain across the diagnostic threshold for extremely low birth weight (1000g) as well as for relatively healthy and heavy newborns (severity groups 1 and 2 at 1500g, 2000g, and 2500g). The additional quantity of treatment does not seem to result in higher survival rates. It may - of course - have other health benefits for the newborns later on in life; an outcome that we cannot observe in our data.

6 Sensitivity Analyses

In this section we present results of three sets of sensitivity analyses. First, we explore the sensitivity of the results to a so-called donut-regression and thus to excluding from our data newborns with weight very close to the thresholds. Second, we explore whether the choice of the polynomial in birth weight affects the results, and third whether the choice of bandwidth has an impact.

– TABLE 5 ABOUT HERE –

All results presented in Table 5 are variations of the most comprehensive specification presented in Table 3 including controls for birth weight, health-related as well as hospital-related control variables, hospital fixed effects and excluding newborns who die on their first day of life.

The results of the donut regressions are presented in the first rows of results for each of the three outcomes (length of stay, number of procedures, and mortality). Compared to the main results in Table 3, excluding newborns in a 10g interval around each threshold induces only minor changes for the quantity of care but renders the mortality results insignificant and even turns around some of the signs. For length of stay, the previously significantly negative coefficients at the 600g and 1500g thresholds become insignificant when newborns with weight in a 10g interval at each side of the respective thresholds are excluded, while the positive coefficients at 2000g and 2500g even increase in value and stay significantly different from zero. The estimates for the number of procedures are hardly affected at all by the exclusion of the donuts in birth weight. In terms of mortality, the only coefficient that was significant in the main results at 2000g becomes insignificant when the donut is excluded. At the same time, differences at 1000g and 1500g become significantly positive. Overall, excluding the donut around the thresholds thus supports the conclusion that differences in treatment exists if at all only at the highest weight thresholds and that these differences do not affect mortality.

The second and third rows for each outcome in Table 5 report the results with different orders of the polynomial in birth weight. While we use a second order polynomial in the main specification displayed in Table 3, the results in row (2) of Table 5 are based on a linear fit and the results in row (3) on a third order polynomial. While there are slight changes in the estimates of the differences across thresholds, changing the order of the polynomial has in general only limited effects on our results. The results for the highest threshold (2500g) are particularly stable, indicating that – irrespective of the order of the polynomial included – newborns with weight below the threshold receive significantly more care than newborns above, while there are no differences in mortality. The results for the next highest threshold (2000g) are also very stable for mortality suggesting reduced mortality below the threshold. The differences in the quantity of care, however, become insignificant with a third order polynomial. For the lower thresholds, there are hardly any significant differences in any of the outcomes irrespective of the order of the polynomial included.

The fourth and fifth rows for each outcome in Table 5 report results when changing the bandwidth. The main results are based on a 100g bandwidth. Row (4) in Table 5 reports results for a 50g bandwidth, row (5) using the entire birth weight interval, e.g. 450-599g below 600g, and 600-749g above 600.[16] Again, there are only minor changes in the coefficient estimate of interest due to the changes in bandwidth. Similar to the other sensitivity analyses, the results are extremely robust for the highest threshold (2500g). But also at the lower thresholds, only minor changes occur.

Overall, the results presented in Table 5 highlight that our results are not driven by newborns that have weights very close to the thresholds, nor by the choice of the order of the polynomial in birth weight or by bandwidth choice. All sensitivity analyses indicate that – despite controlling for an extensive set of variables related to health and

[16] For the placebo thresholds, we use all observations in the interval between the reimbursement relevant thresholds in which the placebo threshold lies. E.g. for the threshold of 700g, we use 600–749g, and for 1300g, we use 1250–1499g.

the place as well as the people who deliver the care, hospital fixed effects and excluding newborns who die very early on – newborns below the highest weight threshold receive more care than their immediate neighbors with weight at or above the threshold. The results for differences in care around the next highest threshold (2000g) are somewhat more sensitive to the exact specification but also indicate overall that newborns below the threshold receive additional care. For these groups of newborns, mortality differences are also observed in most specifications. For all lower thresholds, hardly any of the differences are significantly different from zero – irrespective of the specification.

7 Discussion and Conclusion

In this paper, we investigate whether birth weight thresholds in hospital reimbursement systems affect the quantity and quality of care delivered to newborns. Using the universe of hospital births in Germany from the years 2005–2011, we document that newborns with weight below all but one reimbursement relevant threshold receive more care and have lower risk of dying during the hospital stay than their neighbors with weight at or above the respective thresholds. For all but the highest weight threshold (2500g) the differences in care around thresholds, however, do not remain significant when controlling for birth weight and differences in health around the thresholds that are observable to us as researchers and also to the medical personell that reports birth weight. The health differences therefore likely result from selective birth weight manipulation as medical staff in hospitals tends to under-report a newborn's weight if higher costs are expected. Overall, our results suggest that – if at all – only the heaviest newborns (in our sample with birth weight just below 2000g or 2500g) and among those the group with the least severe health conditions as measured by the assigned DRG stay longer in the hospital and receive more procedures because their weight is below the threshold.

These results lead to the questions what drives the differences in care around 2000g and 2500g and why there are no (robust) differences around the other thresholds. A likely explanation is that we do not account for selective upcoding well enough for the higher threshold. The differences in care for the lower thresholds mainly disappear when the analyses are restricted to first day survivors. As mortality and first day mortality decrease with weight, whether newborns are expected to survive or not is likely not the margin that determines upcoding for higher weight newborns. Instead, other measures such as gestational age may be more important. Jürges and Köberlein (2015) for example find that newborns in the 25g weight interval below the 2500g threshold have on average almost four fewer days of gestational age compared to newborns with weight of 2500-2525g. As in our data we only observe broad categories of gestational age, we cannot fully control for these differences. While for the lower weight groups this may not matter as differences in gestational age translate into early mortality (which we observe and control for), this is not necessarily the case for the high weight thresholds. To us, the most plausible explanation for the remaining differences in quantity of care around the higher weight thresholds is that we cannot fully control for selective upcoding.

Of course, the differences in care around the higher thresholds could also reflect actual differences, e.g. triggered by the additional reimbursement or the fact that newborns with weight below 2500g receive a diagnosis of low birth weight, which may possibly increase the attention of hospital staff. However, it is unclear why similar differences should not be observed for the lower weight thresholds, specifically as the latter also trigger diagnoses (e.g. 1500g and 1000g) and result in much larger differences in reimbursement for the hospital. Overall, we thus judge it to be more likely that neither reimbursement differences nor medical guidelines or diagnostic thresholds induce differences in neonatal care in German hospitals.

In light of the evidence from other countries concerning the benefits of having a birth weight just below 1500g and as reimbursement differences should if at all only add to the existing differences caused by the diagnostic threshold, these result may seem surprising. However, Almond et al. (2010) show that the differences in neonatal mortality and care are not present in all hospitals. Instead the effects are driven by low quality ones. Furthermore, medical guidelines in Germany give only very few recommendations that depend on birth weight, which may explain differences compared to other countries like Chile and Denmark where explicit recommendations with respect to birth weight exist (Bharadwaj et al. 2013; Breining et al. 2015). Last but not least, like in the other studies on upcoding (Dafny 2005; Verzulli et al. 2016), higher reimbursement only depends on the reporting of characteristics or diagnoses – in this case birth weight – not on the treatment that is delivered. Our findings of little changes in quantity and quality of care thus align with the earlier findings that additional resources acquired through upcoding do not profit the specific patients whose records are manipulated.

To conclude, we interpret our results in a way that financial incentives relating to birth weight in Germany do lead to birth weight manipulation but do not directly impact the care that specific newborns receive. This suggests that hospital staff is willing to manipulate records according to their employers' financial incentives but does not take the implications of these incentives (higher funds available for the specific case) into account when making critical medical decisions. This finding is in line with physicians taking treatment decisions in the interest of their patients.

Tables and Figures

Table 1: Summary statistics, by birth weight bracket, 2005-2011

	<600	600-749	750-874	875-999	1000-1249	1250-1499	1500-1999	2000-2499	2500-3000	All births
Birth weight (gram)	527.092	680.801	816.950	948.530	1149.752	1404.486	1808.097	2306.875	2808.359	2601.265
	(43.828)	(45.427)	(34.144)	(37.177)	(69.413)	(72.487)	(134.898)	(136.260)	(140.054)	(441.940)
Outcomes										
Length of stay (days)	53.706	69.951	67.946	61.468	50.428	39.012	23.776	9.821	4.662	8.452
	(62.826)	(55.507)	(43.185)	(34.718)	(27.785)	(21.412)	(16.439)	(9.564)	(4.196)	(14.196)
Procedures (#)	11.729	14.354	13.638	12.192	10.203	8.296	5.942	3.208	1.724	2.559
	(11.393)	(10.688)	(9.407)	(7.569)	(6.662)	(5.129)	(4.154)	(2.988)	(1.601)	(3.275)
Mortality	.487	.236	.126	.069	.042	.025	.012	.003	.001	.006
	(.500)	(.425)	(.332)	(.253)	(.201)	(.155)	(.111)	(.057)	(.028)	(.077)
First-day-mortality	.339	.125	.056	.031	.024	.014	.008	.002	.000	.003
	(.474)	(.331)	(.231)	(.174)	(.154)	(.118)	(.087)	(.044)	(.021)	(.059)
Health-related controls										
Male births	.474	.502	.521	.529	.518	.501	.484	.453	.432	.442
	(.499)	(.500)	(.500)	(.499)	(.500)	(.500)	(.500)	(.498)	(.495)	(.497)
Extreme prematurity	.590	.581	.495	.369	.146	.037	.010	.003	.001	.012
	(.492)	(.493)	(.500)	(.483)	(.353)	(.188)	(.102)	(.055)	(.025)	(.107)
Prematurity	.158	.198	.261	.364	.527	.611	.592	.350	.108	.188
	(.364)	(.399)	(.439)	(.481)	(.499)	(.488)	(.491)	(.477)	(.311)	(.391)
Twin birth	.172	.173	.186	.211	.229	.260	.303	.229	.058	.107
	(.378)	(.378)	(.389)	(.408)	(.420)	(.439)	(.459)	(.420)	(.235)	(.309)
Multiple birth	.022	.026	.026	.031	.043	.046	.026	.004	.000	.004
	(.145)	(.158)	(.158)	(.172)	(.203)	(.208)	(.159)	(.059)	(.012)	(.060)

continued on next page

Table 1: Summary statistics, by birth weight bracket, 2005-2011 (continued)

	<600	600-749	750-874	875-999	1000-1249	1250-1499	1500-1999	2000-2499	2500-3000	All births
Health-related controls (continued)										
Asphyxia	.127	.130	.125	.103	.096	.080	.056	.029	.013	.021
	(.333)	(.336)	(.331)	(.304)	(.295)	(.272)	(.229)	(.169)	(.112)	(.145)
Severe asphyxia	.065	.055	.051	.036	.029	.022	.013	.006	.003	.005
	(.247)	(.228)	(.220)	(.187)	(.168)	(.145)	(.115)	(.079)	(.050)	(.071)
Moderate asphyxia	.059	.074	.072	.065	.065	.057	.041	.022	.010	.016
	(.235)	(.262)	(.258)	(.246)	(.246)	(.233)	(.198)	(.147)	(.098)	(.124)
IRDS	.622	.784	.799	.790	.717	.611	.399	.160	.040	.106
	(.485)	(.412)	(.401)	(.408)	(.451)	(.487)	(.490)	(.367)	(.196)	(.308)
Maternal smoking	.013	.019	.023	.029	.027	.030	.037	.027	.006	.012
	(.112)	(.136)	(.150)	(.167)	(.161)	(.170)	(.189)	(.163)	(.076)	(.108)
Maternal alcohol use	XXX	.001	.001	.001	.001	.001	.001	.001	.000	.000
	XXX	(.036)	(.038)	(.034)	(.032)	(.037)	(.038)	(.028)	(.013)	(.019)
Maternal drug use	XXX	.002	.003	.003	.003	.005	.005	.004	.001	.002
	XXX	(.039)	(.056)	(.052)	(.054)	(.069)	(.074)	(.064)	(.038)	(.047)
Hospital-related controls										
In-patient ward	.007	.005	.006	.005	.004	.007	.015	.063	.113	.095
	(.085)	(.067)	(.079)	(.071)	(.065)	(.082)	(.121)	(.244)	(.316)	(.293)
In-patient midwife	.016	.014	.014	.011	.014	.014	.022	.051	.082	.071
	(.124)	(.118)	(.117)	(.103)	(.116)	(.118)	(.145)	(.219)	(.274)	(.256)
In-patient doctor	.007	.006	.007	.005	.005	.007	.015	.063	.113	.095
	(.081)	(.076)	(.083)	(.071)	(.072)	(.083)	(.122)	(.243)	(.316)	(.293)
Number of births	3059	4589	3461	5135	9698	15044	49708	167624	727567	985885

Notes: Standard deviations (in parenthesis) below means. Variable definitions: Procedures: Number of stated OPS codes (= German modification of ICPM); Mortality: Infant death during hospital stay; First-day-mortality: Infant death on first-day of live in hospital; Extreme prematurity: Birth with gestational age <28 weeks (ICD-10-GM: P07.2); Prematurity: Birth with gestational age >28 and < 37 weeks (ICD-10-GM: P07.3); Twin birth: (ICD-10-GM: Z38.3); Multiple birth: (ICD-10-GM: Z38.6); Severe asphyxia: Asphyxia with APGAR score below 4 (ICD-10-GM: P21.0); Moderate asphyxia: Asphyxia with APGAR score above 3 and below 8 (ICD-10-GM: P21.1); Asphyxia: Any asphyxia (ICD-10-GM: P21.x); IRDS: Infant respiratory distress syndrome (ICD-10-GM: P22.0-P22.9); Maternal smoking (ICD-10-GM: P04.2), alcohol use (ICD-10-GM: P04.3), or drug use (ICD-10-GM: P04.4): Newborn has been harmed by respective maternal behaviour; In-patient ward: Birth place is in-patient ward; In-patient midwife: Birth with external midwife; In-patient doctor: Birth with external surgeon. **Source:** Own calculations based on the DRG-Statistic.

Table 2: Differences above and below threshold, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Outcomes												
Length of stay (days)	3.187	11.364***	2.338	8.629***	2.099	3.135***	4.024***	1.602***	5.589*	2.582**	.758***	.165***
	(3.378)	(2.666)	(1.983)	(1.952)	(1.289)	(1.213)	(.282)	(.078)	(3.077)	(1.064)	(.184)	(.039)
Procedures (#)	.606	1.142**	.656	1.75***	.877***	.618**	1.042***	.737***	.982*	.058	.208***	.049***
	(.611)	(.579)	(.428)	(.444)	(.263)	(.3)	(.079)	(.028)	(.567)	(.269)	(.059)	(.015)
Mortality	-.022	-.088***	.03**	-.033**	-.015*	-.041***	-.007***	0	.008	-.009	0	0
	(.027)	(.022)	(.015)	(.016)	(.009)	(.01)	(.002)	(0)	(.022)	(.008)	(.001)	(0)
First-day-mortality	-.041*	-.065***	.011	-.05***	-.015**	-.043***	-.004**	-.001*	.014	-.001	0	0
	(.023)	(.016)	(.01)	(.015)	(.007)	(.01)	(.002)	(0)	(.016)	(.007)	(.001)	(0)
Health-related controls												
Male births	.004	-.039	-.01	.021	-.003	.029	.021**	.031***	.041	-.035	.007	.002
	(.028)	(.028)	(.026)	(.029)	(.022)	(.023)	(.01)	(.005)	(.026)	(.023)	(.009)	(.004)
Extreme Prematurity	.014	.032	.041	.085***	.01	-.006	0	0	.004	.012	0	0
	(.028)	(.028)	(.026)	(.025)	(.011)	(.007)	(.002)	(.001)	(.026)	(.012)	(.001)	(0)
Prematurity	.001	.021	-.003	.012	-.006	.024	.099***	.033***	-.006	-.043*	.013	.019***
	(.022)	(.023)	(.024)	(.029)	(.022)	(.023)	(.01)	(.005)	(.021)	(.023)	(.009)	(.003)
Twin births	-.004	-.015	-.047**	-.024	-.007	.003	.001	.031***	.019	.015	.006	.015***
	(.021)	(.022)	(.022)	(.024)	(.019)	(.02)	(.01)	(.004)	(.02)	(.02)	(.008)	(.002)
Multiple births	-.005	.008	-.011	-.022*	-.004	-.003	.002	.001***	-.011	.003	.003*	0
	(.008)	(.009)	(.008)	(.013)	(.01)	(.009)	(.002)	(0)	(.008)	(.01)	(.001)	(0)

continued on next page

Table 2: Differences above and below threshold, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Asphyxia	.033*	.008	-.006	-.004	.001	.018*	.002	.011***	.031*	-.03**	0	0
	(.02)	(.018)	(.016)	(.017)	(.013)	(.011)	(.005)	(.002)	(.018)	(.012)	(.003)	(.001)
Severe asphyxia	.019	-.013	.005	.003	0	.003	-.002	.002*	.01	-.011	.001	0
	(.014)	(.013)	(.011)	(.01)	(.007)	(.006)	(.002)	(.001)	(.012)	(.007)	(.001)	(0)
Moderate asphyxia	.011	.024*	-.012	-.003	.004	.018**	.003	.009***	.023	-.023**	-.001	0
	(.015)	(.014)	(.013)	(.014)	(.011)	(.009)	(.004)	(.001)	(.014)	(.01)	(.003)	(.001)
IRDS	.046*	.061***	.006	.097***	.024	.119***	.111***	.081***	-.013	.009	.01	.005**
	(.025)	(.023)	(.021)	(.027)	(.021)	(.023)	(.009)	(.003)	(.021)	(.022)	(.007)	(.002)
Maternal smoking	-.006	.005	-.001	.009	.003	.018***	-.001	.007***	-.005	-.002	.001	.002***
	(.008)	(.007)	(.009)	(.009)	(.007)	(.006)	(.004)	(.001)	(.008)	(.007)	(.003)	(.001)
Maternal alcohol use	0	-.004	.001	.002**	-.002	0	0	0	-.001	.002	-.001	0
	(0)	(.003)	(.002)	(.001)	(.002)	(.002)	(.001)	(0)	(.001)	(.002)	(0)	(0)
Maternal drug use	-.002	-.001	.004	0	-.004	0	.002	.003***	0	-.002	.002	0
	(.002)	(.002)	(.003)	(.003)	(.003)	(.003)	(.001)	(.001)	(.002)	(.003)	(.001)	(0)
Hospital-related controls												
In-patient ward	-.001	-.01*	-.006	-.003	-.006*	-.013**	-.024***	-.029***	.001	0	-.008**	-.007**
	(.005)	(.006)	(.004)	(.005)	(.004)	(.006)	(.004)	(.003)	(.003)	(.003)	(.004)	(.003)
In-patient midwife	-.005	-.009	-.009	-.007	-.006	-.001	-.014***	-.019***	.004	-.004	0	-.002
	(.006)	(.007)	(.007)	(.007)	(.006)	(.006)	(.004)	(.003)	(.006)	(.004)	(.003)	(.002)
In-patient doctor	.002	-.002	-.008	.002	.002	-.016**	-.022***	-.03***	.003	-.002	-.008**	-.005*
	(.004)	(.004)	(.005)	(.003)	(.003)	(.006)	(.004)	(.003)	(.004)	(.003)	(.004)	(.003)
N_L	587	1105	880	1881	1957	3881	5665	14432	635	731	5045	20825
N_R	666	446	629	347	693	538	3806	19384	844	1348	8777	37094

Notes: Difference between means below and above birth weight threshold T within a ±25 gram bandwidth ($\bar{X}_{below}$-$\bar{X}_{above}$). (Robust) Standard errors in parentheses below mean difference. N_L: Number of observations left/below threshold within 25 gram bandwidth. N_R: Number of observations right/above threshold within 25 gram bandwidth. *** p<0.01; ** p<0.05 and * p<0.1. For variable definitions see table 1. **Source:** Own calculations based on the DRG-Statistic.

Table 3: Regression estimates, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Length of stay (days)												
(1)	7.963	13.542***	2.055	11.249***	2.604	2.266	4.334***	1.733***	5.353	2.973*	.244	.039
	(5.141)	(3.9)	(2.986)	(2.757)	(1.893)	(1.745)	(.415)	(.126)	(4.441)	(1.574)	(.297)	(.062)
(2)	2.821	8.778**	1.117	7.584***	2.8	.677	2.753***	.911***	5.869	2.83*	.155	-.009
	(4.636)	(3.598)	(2.726)	(2.454)	(1.777)	(1.682)	(.393)	(.117)	(4.245)	(1.48)	(.276)	(.058)
(3)	-2.824	5.149	1.876	5.452**	1.722	-.899	2.707***	.898***	6.218	2.653*	.153	-.009
	(5.507)	(3.684)	(2.703)	(2.49)	(1.779)	(1.736)	(.394)	(.117)	(4.362)	(1.464)	(.276)	(.058)
(4)	-9.439*	4.206	-.078	-.457	-.883	-3.446*	1.451***	.834***	5.877	2.79**	-.053	.014
	(5.389)	(3.567)	(2.465)	(2.156)	(1.746)	(1.774)	(.367)	(.113)	(4.132)	(1.322)	(.258)	(.057)
Procedures (#)												
(1)	1.407	1.427*	.647	1.959***	1.049***	.246	1.122***	.859***	1.003	-.296	.164*	.011
	(.945)	(.78)	(.689)	(.55)	(.382)	(.455)	(.123)	(.045)	(.912)	(.394)	(.091)	(.024)
(2)	.234	.559	.486	1.16**	1.02***	-.225	.61***	.449***	1.01	-.171	.135*	-.014
	(.822)	(.708)	(.64)	(.487)	(.354)	(.436)	(.112)	(.039)	(.851)	(.368)	(.081)	(.021)
(3)	-.544	-.056	.566	.87*	.899**	-.487	.605***	.448***	.843	-.186	.132	-.015
	(1.006)	(.758)	(.656)	(.512)	(.363)	(.456)	(.112)	(.039)	(.92)	(.371)	(.081)	(.021)
(4)	-.596	-.221	.44	-.121	.508	-.569	.354***	.397***	1.145	-.083	.112	-.024
	(.964)	(.707)	(.591)	(.475)	(.339)	(.455)	(.105)	(.036)	(.82)	(.336)	(.075)	(.019)
Mortality												
(1)	-.086**	-.134***	.008	-.042*	-.032**	-.059***	-.011***	-.002**	-.038	-.016	0	0
	(.041)	(.031)	(.022)	(.023)	(.014)	(.015)	(.003)	(.001)	(.033)	(.012)	(.002)	(0)
(2)	-.06	-.108***	.009	-.036	-.03**	-.056***	-.01***	-.002**	-.046	-.016	0	0
	(.038)	(.03)	(.022)	(.023)	(.014)	(.015)	(.003)	(.001)	(.032)	(.012)	(.002)	(0)
(3)	.015	-.044*	-.002	.026**	-.004	0	-.005***	0	-.04	-.01	0	0
	(.038)	(.025)	(.018)	(.012)	(.008)	(.007)	(.002)	(.001)	(.028)	(.008)	(.001)	(0)
(4)	.024	-.031	0	.017	.002	-.001	-.005**	0	-.046	-.012	.001	0
	(.04)	(.026)	(.018)	(.014)	(.008)	(.007)	(.002)	(.001)	(.03)	(.008)	(.001)	(0)
Number of births												
(1)+(2)+(3) - N_L	2002	3326	3015	4506	5264	8728	16919	57475	2639	4717	23288	99190
N_R	2639	2581	3254	2470	3781	3938	15667	71505	2990	4892	30853	131754
(4) - N_L	1333	3017	2871	4366	5166	8642	16833	57397	2198	4626	23222	99129
N_R	2198	2424	3140	2388	3697	3867	15604	71449	2773	4803	30792	131682

Notes: Parametric regressions within ±100 gram intervals around birth weight thresholds. Coefficient of binary variable indicating observations below threshold. (Robust) Standard errors in parentheses. N_L: Number of observations left to/below the threshold . N_R: Number of observations right to/above the threshold. All regressions include a second order birth weight polynomial (fitted separately on each side of the threshold). Control variables are gender, (extreme) prematurity, multiples, asphyxia, IRDS, in-patient ward/midwife/doctor, maternal smoking/alcohol consumption/ drug use and years. [Specification] (1) includes only the binary birth weight threshold indicator variable . (2) additionally includes controls. (3) additionally restricts the sample to first-day-survivors. (4) additionally includes hospital-fixed effects. *** p<0.01; ** p<0.05 and * p<0.1. For variable definitions see table 1. **Source:**. Own calculations based on the DRG-Statistic.

Table 4: Regression estimates by Severity groups for first-day-survivors, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Length of stay (days)												
Severity group 1						8.398*	1.592***	.186***			.097	.045*
						(4.679)	(.43)	(.046)			(.16)	(.027)
Severity group 2						4.086***	1.287***	1.162***			-.012	.122
	-4.432	5.056	.323	12.425***	-.522	(1.196)	(.336)	(.174)	-2.688	.396	(.257)	(.141)
Severity group 3	(6.235)	(3.454)	(2.499)	(3.503)	(1.264)	-.713	1.125*	.642	(4.113)	(1.275)	.206	.724**
						(1.362)	(.618)	(.418)			(.611)	(.367)
Severity group 4						-8.507***	1.232	-.495			-.646	1.828
						(2.941)	(1.352)	(1.212)			(1.484)	(1.296)
Severity group 5				12.98**	-7.318**	-26.351	-1.261	.235		8.823***	-4.425	-4.647
	-12.335	3.665	-.964	(5.892)	(3.289)	(18.792)	(3.652)	(4.099)	7.472	(3)	(4.388)	(3.774)
Severity group 6	(9.098)	(7.322)	(6.782)	12.206**	-5.219	-1.316	2.23	3.072	(7.939)	5.08	-3.022	-8.268*
				(4.986)	(3.938)	(5.506)	(4.894)	(7.599)		(3.125)	(6.469)	(4.734)
Procedures (#)												
Severity group 1						3.001**	.514***	.06***			.019	.001
						(1.209)	(.093)	(.012)			(.033)	(.007)
Severity group 2						1.24***	.271***	.276***			.084	.027
	.239	.381	.171	2.077***	.417	(.236)	(.073)	(.049)	-.78	-.21	(.057)	(.048)
Severity group 3	(.748)	(.55)	(.437)	(.598)	(.271)	-.403	.133	.503***	(.571)	(.275)	.139	.093
						(.293)	(.153)	(.117)			(.151)	(.108)
Severity group 4						-1.509***	.229	-.123			.309	.41
						(.468)	(.328)	(.383)			(.335)	(.423)
Severity group 5				7.874***	-.25	-3.24	-.428	-1.672		.433	1.821	-.722
	-2.143	-.976	1.064	(1.265)	(.667)	(3.717)	(1.108)	(1.521)	2.678*	(.69)	(1.275)	(1.577)
Severity group 6	(1.78)	(1.49)	(1.808)	5.153***	.188	.277	-.327	.117	(1.554)	.996	.292	-5.389***
				(1.122)	(.708)	(1.676)	(1.37)	(1.888)		(.82)	(2.576)	(1.809)
Mortality												
Severity group 1						-.001	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						0	0	0			0	0
	.034	-.028	-.015	.011	.002	(.001)	(0)	(0)	-.018	-.003	(0)	(0)
Severity group 3	(.049)	(.028)	(.019)	(.022)	(.004)	-.002	0	0	(.035)	(.003)	0	0
						(.003)	(0)	(0)			(0)	(0)
Severity group 4						-.002	.004	-.023**			-.004	0
						(.003)	(.005)	(.01)			(.003)	(.003)
Severity group 5				.197***	.002	.008	0	.018		-.002	-.001	0
	.046	-.074	.009	(.042)	(.009)	(.015)	(0)	(.015)	-.029	(.023)	(.007)	(0)
Severity group 6	(.07)	(.055)	(.054)	.158***	-.007	.039*	-.043	.047	(.057)	-.022	-.063	.011
				(.037)	(.019)	(.021)	(.052)	(.04)		(.019)	(.044)	(.018)
Number of births												
1 - N_L						5400	1337	33587			6115	76359
N_R						34	3211	51886			11585	107440
2 - N_L						5400	6902	13677			9036	12098
N_R	785	2030	2111	3536	2061	1021	6305	10332	1354	2025	10613	13249
3 - N_L	1354	1734	2525	650	1844	5400	3829	4754	1930	2536	3674	5591
N_R						1046	2736	4897			3993	5651
4 - N_L						5400	1590	1560			1326	1342
N_R						618	1001	1155			1334	1321

continued on next page

Table 4: Regression estimates for first-day-survivors by Severity groups, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
5 - N_L				961	799	984	383	253		688	247	167
N_R	627	1154	833	359	525	212	203	142	947	656	220	143
6 - N_L	947	754	738	961	1114	1176	662	298	983	953	226	223
N_R				553	694	298	187	192		875	252	234

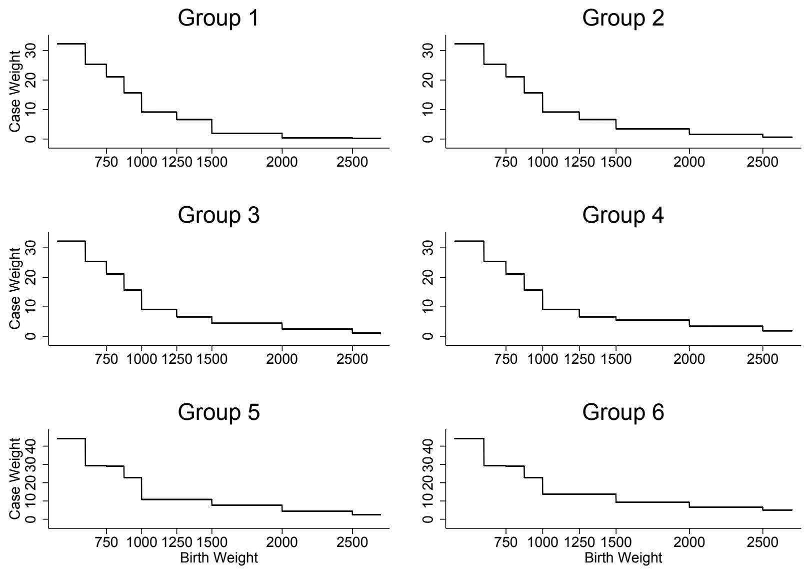
Notes: Parametric regressions within ±100 gram intervals around birth weight thresholds by Severity group for first-day-survivors. Coefficient of binary variable indicating observations below threshold. (Robust) Standard errors in parentheses. N_L: Number of observations left to/below the threshold . N_R: Number of observations right to/above the threshold. Control variables include a second order birth weight polynomial (fitted separately on each side of the threshold), hospital-fixed-effects and dummy variables for: gender, (extreme) prematurity, multiples, asphyxia, IRDS, in-patient ward/midwife/doctor, maternal smoking/alcohol consumption/ drug use and years. *** p<0.01; ** p<0.05 and * p<0.1. For variable definitions see table 1. **Source:**. Own calculations based on the DRG-Statistic.

Table 5: Regression estimates - Robustness, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Length of stay (days)												
(1)	-13.927	4.157	-2.075	2.877	-1.965	-1.722	2.941***	1.05***	5.905	3.982*	.328	-.044
	(9.022)	(5.9)	(3.987)	(3.458)	(2.247)	(1.873)	(.627)	(.155)	(7.966)	(2.209)	(.402)	(.089)
(2)	-4.953	.591	.26	-1.645	-.796	-.716	1.034***	.648***	5.858**	1.024	.072	.028
	(3.532)	(2.264)	(1.659)	(1.404)	(1.083)	(.993)	(.25)	(.069)	(2.767)	(.931)	(.164)	(.035)
(3)	-8.751	9.253*	2.089	-3.183	-.718	-4.214*	.541	.786***	6.894	.446	-.148	.12
	(7.83)	(4.752)	(3.351)	(2.78)	(2.233)	(2.347)	(.505)	(.167)	(5.26)	(1.953)	(.372)	(.091)
(4)	-5.3	6.527	4.274	-1.636	.697	-5.259*	1.048*	.797***	5.416	.859	-.436	.089
	(8.25)	(5.109)	(3.718)	(3.171)	(2.431)	(2.701)	(.556)	(.182)	(5.67)	(2.265)	(.39)	(.1)
(5)	-6.541	1.279	-.352	-1.488	-.386	-.202	.841***	.509***	6.33	.178	.006	.043
	(4.166)	(3.025)	(2.193)	(1.352)	(.891)	(.448)	(.168)	(.049)	(4.541)	(1.941)	(.164)	(.037)
Procedures (#)												
(1)	-2.305	1.007	-.387	-.339	.151	.409	.91***	.471***	.814	.272	.202*	.007
	(1.614)	(1.234)	(.776)	(.875)	(.543)	(.421)	(.161)	(.051)	(1.32)	(.551)	(.114)	(.03)
(2)	.084	-.146	.25	-.174	.164	.073	.31***	.281***	1.093**	-.16	.035	-.002
	(.653)	(.462)	(.364)	(.338)	(.23)	(.244)	(.069)	(.022)	(.538)	(.224)	(.048)	(.012)
(3)	.429	-.173	.734	-.128	.716	-1.04*	.02	.41***	1.765	-.43	.048	-.037
	(1.347)	(.924)	(.833)	(.532)	(.436)	(.621)	(.149)	(.053)	(1.166)	(.481)	(.109)	(.029)
(4)	.579	-.455	1.108	.12	.901*	-1.107	.218	.449***	.883	-.463	-.04	-.029
	(1.459)	(.967)	(.894)	(.593)	(.479)	(.718)	(.163)	(.058)	(1.296)	(.53)	(.118)	(.031)
(5)	-.364	-.361	.379	-.433	.201	.039	.239***	.208***	1.062	-.16	.045	-.004
	(.735)	(.61)	(.512)	(.322)	(.202)	(.109)	(.045)	(.015)	(.894)	(.46)	(.048)	(.013)
Mortality												
(1)	.045	-.039	-.02	.04*	-.01	.015**	-.001	0	0	.003	0	0
	(.07)	(.044)	(.029)	(.022)	(.016)	(.007)	(.003)	(.001)	(.051)	(.014)	(.002)	(0)
(2)	-.003	-.027	.007	.009	.006	-.002	-.004***	0	-.008	-.014**	-.001	0
	(.027)	(.017)	(.013)	(.009)	(.005)	(.004)	(.001)	(0)	(.019)	(.006)	(.001)	(0)
(3)	.033	-.056	.018	.004	.007	-.008	-.006**	.001	-.064	-.023**	0	0
	(.057)	(.035)	(.025)	(.018)	(.01)	(.01)	(.003)	(.001)	(.04)	(.01)	(.002)	(0)
(4)	.037	-.018	.025	-.005	.016	-.013	-.006*	.001	-.069	-.024**	-.001	0
	(.06)	(.037)	(.027)	(.021)	(.011)	(.011)	(.003)	(.001)	(.043)	(.011)	(.002)	(0)
(5)	.012	-.027	.005	.018**	-.001	0	-.002**	0	-.063*	-.016*	.001	0
	(.031)	(.023)	(.016)	(.009)	(.005)	(.002)	(.001)	(0)	(.033)	(.008)	(.001)	(0)

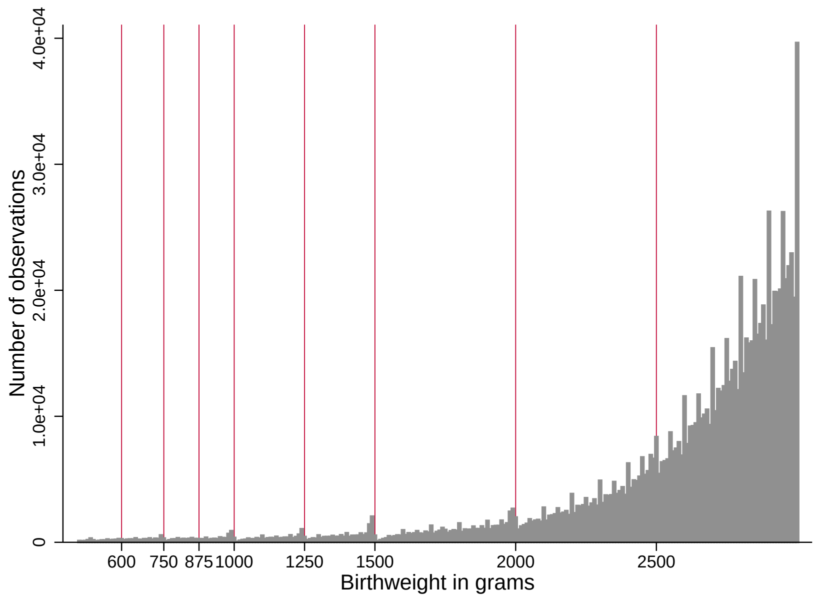
Notes: Parametric regressions within different intervals around birth weight thresholds always include control variables, hospital-fixed effects and are restricted to first-day-survivors. Coefficient of binary variable indicating observations below threshold. (Robust) Standard errors in parentheses. N_L: Number of observations left to/below the threshold . N_R: Number of observations right to/above the threshold. Control variables include gender, (extreme) prematurity, multiples, asphyxia, IRDS, in-patient ward/midwife/doctor, maternal smoking/alcohol consumption/ drug use and years. [Specification] (1) uses a second order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±100 gram bandwidth around the threshold, but excludes observations within a ±10 gram bandwidth around the threshold. (2) uses a first order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±100 gram bandwidth around the threshold. (3) uses a third order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±100 gram bandwidth around the threshold. (4) uses a second order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±50 gram bandwidth around the threshold. (5) uses a second order birth weight polynomial (fitted separately on each side of the threshold) and includes all observations between two thresholds. For variable definitions see table 1. **Source:**. Own calculations based on the DRG-Statistic.

Figure 1: Hospital Reimbursement by Birth Weight



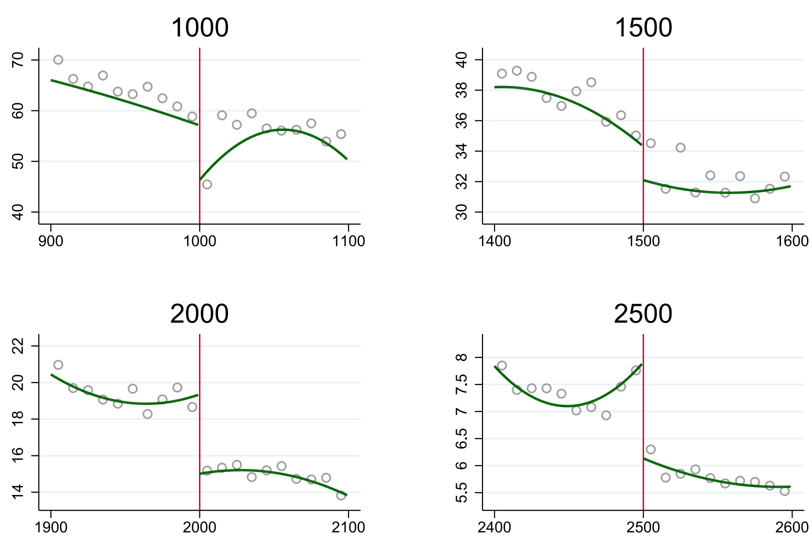
Notes: Development of case-weights for the six DRG-severity groups defined in tables A.1 for the year 2008. **Source:** DRG catalogue 2008

Figure 2: Low Birth Weight Frequencies (Years 2005-2011)



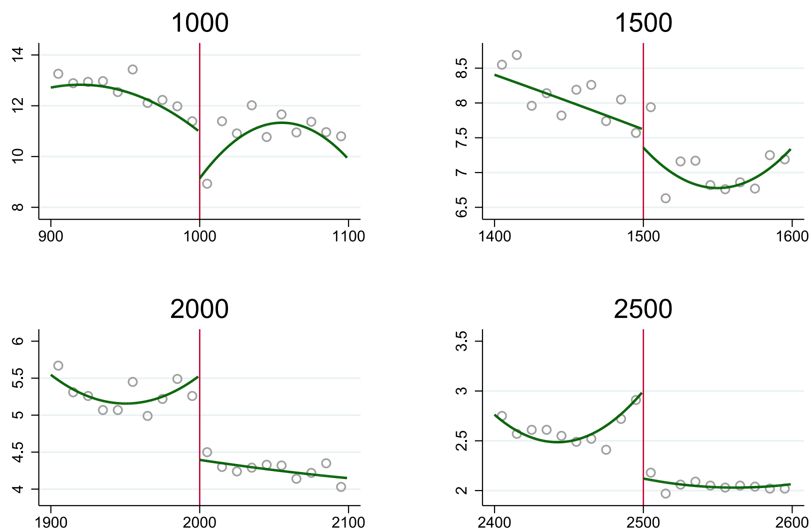
Notes: Distribution of birth weights below 3000 grams in the years 2005-2011. Red lines mark birth weight thresholds in the G-DRG system. All birth weight bins with less than three observations were omitted due to confidentiality. This does not change the look of the graph. **Source:** Own calculations based on the DRG-Statistic.

Figure 3: Length of stay (days) at highest thresholds (Years 2005-2011)



Notes: Circles show mean values in 10 gram intervals. Green lines are fitted values using all observations. Excludes newborns who died on their first day of life. **Source:** Own calculations based on the DRG-Statistic.

Figure 4: Number of procedures at highest thresholds (Years 2005-2011)



Notes: Circles show mean values in 10 gram intervals. Green lines are fitted values using all observations. Excludes newborns who died on their first day of life. **Source:** Own calculations based on the DRG-Statistic.

References

Almond, Douglas, Joseph J Doyle, Amanda E Kowalski and Heidi Williams (2010) "Estimating Marginal Returns To Medical Care: Evidence from at-risk newborns" In: *The Quarterly Journal of Economics* 125 (2), pp. 591–634.

— (2011) "The role of hospital heterogeneity in measuring marginal returns to medical care: A reply to Barreca, Guldi, Lindo, and Waddell" In: *The Quarterly Journal of Economics* 126 (4), pp. 2125–2131.

AWMF (2007) "S2k-Leitline: Frühgeborene an der Grenze der Lebensfähigkeit, AWMF Leitlinien-Register Nr. 024-019".

— (2012) "Leitlinine der Deutschen Gesellschaft für Kinderchirurgie: Kurzstreckige Ösophagusatresie, AWMF Leitlinien-Register Nr. 006-045".

— (2014) "S3-Guideline of the German Society for Nutritional Medicine (DGEM) in Cooperation with the GESKES, the AKE, the DGKJ and the GNPI, No. 073-023".

Barreca, Alan I, Melanie Guldi, Jason M Lindo and Glen R Waddell (2011) "Saving Babies? Revisiting the effect of very low birth weight classification" In: *The Quarterly Journal of Economics* 126 (4), pp. 2117–2123.

Barros, Pedro and Gisele Braun (2016) "Upcoding in a national health service: The evidence from Portugal" In: *Health Economics*.

Bauer, Karl et al. (2006) "Empfehlungen für die strukturellen Voraussetzungen der perinatologischen Versorgung in Deutschland" In: *Zeitschrift für Geburtshilfe und Neonatologie* 210 (01), pp. 19–24.

Bharadwaj, Prashant, Katrine V Løken and Christopher Neilson (2013) "Early life health interventions and academic achievement" In: *The American Economic Review* 103 (5), pp. 1862–1891.

Breining, Sanni, N. Meltem Daysal, Marianne Simonsen and Mircea Trandafir (2015) "Spillover Effects of Early-Life Medical Interventions" In: *IZA Working Paper Nr. 9086*.

Dafny, Leemore S (2005) "How do hospitals respond to price changes?" In: *The American Economic Review* 95 (5), pp. 1525–1547.

Gemeinsamer Bundesausschuss (2013) "Beschluss über eine Änderung der Vereinbarung über Maßnahmen zur Qualitätssicherung der Versorgung von Früh-und Neugeborenen vom 20 Juni 2013".

Hack, Maureen, Daniel J Flannery, Mark Schluchter, Lydia Cartar, Elaine Borawski and Nancy Klein (2002) “Outcomes in young adulthood for very-low-birth-weight infants” In: *New England Journal of Medicine* 346 (3), pp. 149–157.

Hummer, Michael, Thomas Lehner and Gerald Pruckner (2014) “Low birth weight and health expenditures from birth to late adolescence” In: *The European Journal of Health Economics* 15 (3), pp. 229–242.

Januleviciute, Jurgita, Jan Erik Askildsen, Oddvar Kaarboe, Luigi Siciliani and Matt Sutton (2016) “How do Hospitals Respond to Price Changes? Evidence from Norway” In: *Health Economics* 25 (5), pp. 620–636.

Jürges, Hendrik and Juliane Köberlein (2015) “What explains DRG upcoding in neonatology? The roles of financial incentives and infant health” In: *Journal of Health Economics* 43, pp. 13–26.

Kuntz, Benjamin and Thomas Lampert (2016) “Social Disparities in Maternal Smoking during Pregnancy” In: *Geburtshilfe und Frauenheilkunde* 76 (03), pp. 239–247.

Salm, Martin and Ansgar Wübker (2015) “Do Hospitals Respond to Increasing Prices by Supplying Fewer Services?” In: *Ruhr Economic Papers Nr. 567*.

Schreyögg, Jonas, Reinhard Busse, Matthias Bäuml, Jonas Krämer, Tilman Dette and Alexander Geissler (2014) “Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG - Endbericht”. Hamburg Center for Health Economics.

Shigeoka, Hitoshi and Kiyohide Fushimi (2014) “Supplier-induced demand for newborn treatment: Evidence from Japan” In: *Journal of Health Economics* 35, pp. 162–178.

Silverman, Elaine and Jonathan Skinner (2004) “Medicare upcoding and hospital ownership” In: *Journal of Health Economics* 23 (2), pp. 369–389.

Verzulli, Rossella, Gianluca Fiorentini, Matteo Lippi Bruni and Cristina Ugolini (2016) “Price Changes in Regulated Healthcare Markets: Do Public Hospitals Respond and How?” In: *Health Economics*.

Appendix

Table A.1: DRG-Severity Groups for Newborns in Germany (2005-2011)

2005

<table>
<tr><th></th><th><600</th><th>600 - 749</th><th>750 - 874</th><th>875 - 999</th><th>1000 - 1249</th><th>1250 - 1499</th><th>1500 - 1999</th><th>2000 - 2499</th><th>> 2499</th></tr>
<tr><td>Severity group 1</td><td rowspan="6">P61A</td><td rowspan="6">P61B</td><td rowspan="6">P62A</td><td rowspan="6">P62B</td><td rowspan="4">P63Z</td><td rowspan="4">P64Z</td><td>P65D</td><td>P66D</td><td>P67D</td></tr>
<tr><td>Severity group 2</td><td>P65C</td><td>P66C</td><td>P67C</td></tr>
<tr><td>Severity group 3</td><td>P65B</td><td>P66B</td><td>P67B</td></tr>
<tr><td>Severity group 4</td><td>P65A</td><td>P66A</td><td>P67A</td></tr>
<tr><td>Severity group 5</td><td colspan="2">P03D</td><td>P04C</td><td>P05C</td><td>P06C</td></tr>
<tr><td>Severity group 6</td><td colspan="2">P03B</td><td>P04B</td><td>P05B</td><td>P06B</td></tr>
</table>

2006-2010

<table>
<tr><th></th><th><600</th><th>600 - 749</th><th>750 - 874</th><th>875 - 999</th><th>1000 - 1249</th><th>1250 - 1499</th><th>1500 - 1999</th><th>2000 - 2499</th><th>> 2499</th></tr>
<tr><td>Severity group 1</td><td rowspan="4">P61B</td><td rowspan="4">P61D</td><td rowspan="4">P62B</td><td rowspan="4">P62D</td><td rowspan="4">P63Z</td><td rowspan="4">P64Z</td><td>P65D</td><td>P66D</td><td>P67D</td></tr>
<tr><td>Severity group 2</td><td>P65C</td><td>P66C</td><td>P67C</td></tr>
<tr><td>Severity group 3</td><td>P65B</td><td>P66B</td><td>P67B</td></tr>
<tr><td>Severity group 4</td><td>P65A</td><td>P66A</td><td>P67A</td></tr>
<tr><td>Severity group 5</td><td rowspan="2">P61A</td><td rowspan="2">P61C</td><td rowspan="2">P62A</td><td rowspan="2">P62C</td><td colspan="2">P03C</td><td>P04C</td><td>P05C</td><td>P06C</td></tr>
<tr><td>Severity group 6</td><td colspan="2">P03B</td><td>P04B</td><td>P05B</td><td>P06B</td></tr>
</table>

2011

<table>
<tr><th></th><th><600</th><th>600 - 749</th><th>750 - 874</th><th>875 - 999</th><th>1000 - 1249</th><th>1250 - 1499</th><th>1500 - 1999</th><th>2000 - 2499</th><th>> 2499</th></tr>
<tr><td>Severity group 1</td><td rowspan="4">P61B</td><td rowspan="4">P61D</td><td rowspan="4">P62B</td><td rowspan="4">P62C</td><td rowspan="4">P63Z</td><td rowspan="4">P64Z</td><td>P65D</td><td>P66D</td><td>P67D</td></tr>
<tr><td>Severity group 2</td><td>P65C</td><td>P66C</td><td>P67C</td></tr>
<tr><td>Severity group 3</td><td>P65B</td><td>P66B</td><td>P67B</td></tr>
<tr><td>Severity group 4</td><td>P65A</td><td>P66A</td><td>P67A</td></tr>
<tr><td>Severity group 5</td><td rowspan="2">P61A</td><td rowspan="2">P61C</td><td rowspan="2" colspan="2">P62A</td><td colspan="2">P03C</td><td>P04C</td><td>P05C</td><td>P06C</td></tr>
<tr><td>Severity group 6</td><td colspan="2">P03B</td><td>P04B</td><td>P05B</td><td>P06B</td></tr>
</table>

Notes: Severity groups and birth weight thresholds in gram. Within each severity group, birth weight is the only grouping criterion. This table is based on simulations, performed using the G-DRG Webgrouper: http://drg.uni-muenster.de/index.php?option=com_webgrouper&view=webgrouper&Itemid=26

Table A.2: Differences above and below threshold Severity groups, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Outcomes												
Length of stay (days)												
Severity group 1						23.647***	3.071***	.263***			.313***	.067***
						(2.96)	(.324)	(.033)			(.119)	(.019)
Severity group 2						6.991***	1.775***	1.7***			.515***	.285***
	-2.431	5.338**	3.098*	16.178***	1.631*	(.883)	(.247)	(.126)	-1.09	1.534*	(.189)	(.104)
Severity group 3	(3.838)	(2.492)	(1.684)	(2.893)	(.938)	1.595*	1.623***	1.455***	(2.76)	(.886)	.953**	1.12***
						(.933)	(.429)	(.307)			(.41)	(.271)
Severity group 4						-7.452***	2.305***	.52			.122	1.999**
						(2.228)	(.849)	(.781)			(.947)	(.976)
Severity group 5				11.136***	-1.901	-11.935	-3.819*	-2.74		4.492**	-1.825	-5.328***
	-4.618	7.211	5.525	(3.23)	(2.127)	(10.246)	(2.236)	(2.905)	9.657*	(1.945)	(2.215)	(1.821)
Severity group 6	(5.685)	(4.598)	(4.281)	14.056***	-1.721	3.246	1.282	-.784	(5.811)	3.583*	-3.681	-3.893
				(3.229)	(2.432)	(3.023)	(2.83)	(3.499)		(2.168)	(2.888)	(2.535)
Others	3.019	1.758	.817	-38.702***	5.906	-6.832*	.18	-.418	-.681	3.456	.362	-.377
	(2.169)	(2.775)	(.768)	(3.457)	(4.786)	(3.653)	(1.419)	(.923)	(2.1)	(4.189)	(1.137)	(.749)
Procedures (#)												
Severity group 1						5.233***	.671***	.082***			.026	.019***
						(.227)	(.069)	(.01)			(.028)	(.006)
Severity group 2						2.085***	.335***	.459***			.139***	.106***
	-.132	.477	.617*	3.702***	.438**	(.187)	(.057)	(.04)	-.1	-.148	(.046)	(.04)
Severity group 3	(.533)	(.445)	(.338)	(.467)	(.215)	-.25	.209*	.947***	(.447)	(.224)	.089	.209**
						(.222)	(.115)	(.091)			(.115)	(.088)
Severity group 4						-1.461***	.253	.318			-.01	.341
						(.374)	(.236)	(.243)			(.242)	(.287)
Severity group 5				5.917***	.134	-.986	-.712	-.14		-.019	.267	-.969
	-.636	-.546	1.569	(.682)	(.485)	(2.049)	(.872)	(.809)	1.693	(.476)	(.651)	(.787)
Severity group 6	(1.18)	(1.156)	(1.169)	4.159***	.723	.208	-.492	-.883	(1.105)	1.202**	-.545	-1.49**
				(.727)	(.536)	(1.417)	(.79)	(1.011)		(.574)	(1.094)	(.719)
Others	.598	1.413	.588	-5.452***	1.399	-1.772*	-.092	.223	-.356	-.053	.112	-.185
	(.58)	(1.303)	(.618)	(.904)	(.924)	(1.003)	(.4)	(.297)	(.708)	(1.028)	(.379)	(.267)
Mortality												
Severity group 1						.001*	0	0			0	0
						(.001)	(0)	(0)			(0)	(0)
Severity group 2						.001*	0	0			0	0
	.026	-.031	.005	.011	.001	(.001)	(0)	(0)	.015	-.003	(0)	(0)
Severity group 3	(.031)	(.019)	(.012)	(.015)	(.001)	.001*	0	0	(.023)	(.002)	0	0
						(.001)	(0)	(0)			(0)	(0)
Severity group 4						.001*	.002	-.019**			-.005	0
						(.001)	(.002)	(.008)			(.003)	(0)
Severity group 5				.116***	.003	.002	0	-.009		-.001	0	0
	.027	-.082**	.04	(.017)	(.003)	(.002)	(0)	(.027)	.017	(.013)	(0)	(0)
Severity group 6	(.043)	(.037)	(.035)	.116***	-.004	.023***	-.018	.021	(.036)	-.008	-.014	0
				(.017)	(.016)	(.007)	(.023)	(.015)		(.014)	(.029)	(0)
Others	.016	-.08	.115*	.07	-.076**	-.145***	-.036***	-.009	.014	-.022	.003	-.006
	(.041)	(.063)	(.068)	(.042)	(.036)	(.039)	(.013)	(.008)	(.057)	(.038)	(.012)	(.008)

continued on next page

Table A.2: Differences above and below threshold by Severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls												
Male births												
Severity group 1						-.051	-.034	.004			-.015	0
						(.145)	(.029)	(.007)			(.015)	(.005)
Severity group 2						.078*	.008	.021			.017	.004
	.055	-.063*	-.037	.065	0	(.045)	(.017)	(.013)	.014	-.043	(.014)	(.013)
Severity group 3	(.041)	(.036)	(.03)	(.061)	(.033)	-.003	0	.041**	(.034)	(.032)	-.003	.003
						(.044)	(.024)	(.019)			(.023)	(.02)
Severity group 4						-.032	.006	.062*			-.013	-.012
						(.058)	(.04)	(.036)			(.039)	(.041)
Severity group 5				.161**	-.107*	-.028	.117	-.071		.049	-.074	-.067
	.016	-.015	.001	(.073)	(.056)	(.093)	(.082)	(.09)	.019	(.062)	(.107)	(.108)
Severity group 6	(.048)	(.049)	(.053)	.045	.018	.074	.092	-.073	(.043)	.011	.039	-.104
				(.067)	(.052)	(.092)	(.082)	(.086)		(.053)	(.091)	(.098)
Others	-.048	.029	.071	-.021	.004	.036	-.028	-.023	.091	-.099*	.016	.012
	(.05)	(.064)	(.07)	(.051)	(.047)	(.05)	(.028)	(.024)	(.058)	(.053)	(.028)	(.024)
Extreme Prematurity												
Severity group 1						.012***	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						-.012	-.002	-.001			.002	.001
	.013	-.006	.053*	.245***	-.006	(.014)	(.002)	(.001)	-.005	-.006	(.002)	(.001)
Severity group 3	(.04)	(.036)	(.03)	(.03)	(.013)	-.01	.001	-.004	(.034)	(.012)	0	-.002*
						(.013)	(.002)	(.002)			(.002)	(.001)
Severity group 4						-.014	-.01	-.016**			.009	.004
						(.018)	(.008)	(.007)			(.007)	(.004)
Severity group 5				.154**	.025	-.005	.008	-.045		0	0	0
	-.026	.036	.044	(.06)	(.025)	(.034)	(.008)	(.032)	.003	(.034)	(0)	(0)
Severity group 6	(.046)	(.048)	(.053)	.146***	-.031	-.003	-.004	0	(.043)	.079**	-.003	0
				(.056)	(.032)	(.033)	(.024)	(0)		(.034)	(.027)	(0)
Others	-.008	.087	.053	-.041	.057**	.011	.008	.002	-.003	.009	-.009	-.002
	(.05)	(.064)	(.064)	(.047)	(.028)	(.018)	(.008)	(.008)	(.06)	(.034)	(.007)	(.006)
Prematurity												
Severity group 1						.394***	.157***	-.019***			-.005	.013***
						(.125)	(.03)	(.005)			(.014)	(.003)
Severity group 2						.031	.056***	-.03**			-.01	.028**
	-.037	.031	-.004	-.037	-.003	(.045)	(.017)	(.013)	-.03	-.018	(.014)	(.013)
Severity group 3	(.032)	(.031)	(.028)	(.061)	(.032)	.066	.032	-.066***	(.028)	(.032)	.006	.057***
						(.044)	(.023)	(.019)			(.023)	(.02)
Severity group 4						-.044	.028	.031			.065*	.017
						(.054)	(.038)	(.036)			(.038)	(.041)
Severity group 5				-.092	-.047	-.083	-.053	-.306***		-.006	.078	-.144
	.038	.006	.017	(.074)	(.056)	(.087)	(.075)	(.082)	.025	(.061)	(.102)	(.105)
Severity group 6	(.037)	(.039)	(.049)	-.119*	-.019	-.138*	.156*	-.061	(.037)	-.081	-.038	-.02
				(.067)	(.051)	(.078)	(.081)	(.086)		(.052)	(.089)	(.097)
Others	-.015	-.022	-.006	.027	.031	-.023	.052*	-.021	-.008	-.107**	.025	.024
	(.036)	(.047)	(.059)	(.047)	(.048)	(.05)	(.028)	(.023)	(.044)	(.052)	(.027)	(.023)

continued on next page

Table A.2: Differences above and below threshold by Severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Twin births												
Severity group 1						.026	-.005	.021***			.005	.014***
						(.125)	(.029)	(.005)			(.014)	(.003)
Severity group 2						-.022	.009	.146***			.008	.008
	.022	.007	-.033	-.018	-.035	(.042)	(.016)	(.009)	.025	-.038	(.013)	(.006)
Severity group 3	(.03)	(.027)	(.025)	(.051)	(.029)	.039	-.023	-.173***	(.025)	(.028)	.011	.035**
						(.038)	(.022)	(.017)			(.021)	(.017)
Severity group 4						.016	.016	.057**			-.026	.03
						(.051)	(.034)	(.024)			(.031)	(.023)
Severity group 5				-.012	.04	-.045	0	-.078		.172***	.035	-.192***
	.001	-.039	-.062	(.062)	(.049)	(.089)	(.073)	(.073)	.061*	(.059)	(.095)	(.055)
Severity group 6	(.037)	(.038)	(.045)	-.073	.04	-.099	.075	-.089	(.035)	.017	-.027	-.019
				(.061)	(.044)	(.088)	(.058)	(.063)		(.048)	(.063)	(.045)
Others	-.012	-.001	-.053	-.072*	-.021	-.022	.026	.009	-.058	.033	.011	.025**
	(.039)	(.052)	(.051)	(.038)	(.039)	(.041)	(.021)	(.014)	(.042)	(.043)	(.021)	(.012)
Multiple births												
Severity group 1						.047***	.004	.001**			.002	0
						(.004)	(.005)	(0)			(.001)	(0)
Severity group 2						-.002	-.006	0			.002	0
	-.009	.001	-.016	-.023	-.006	(.02)	(.005)	(.001)	-.021**	.007	(.003)	(0)
Severity group 3	(.012)	(.013)	(.01)	(.028)	(.015)	-.005	.006	.001	(.009)	(.015)	.006	0
						(.02)	(.005)	(.001)			(.005)	(.001)
Severity group 4						.021	-.003	.002			-.001	0
						(.019)	(.01)	(.002)			(.007)	(0)
Severity group 5				-.047	.012	-.091	.013	0		.024	-.032	.03
	-.013	.001	-.009	(.038)	(.026)	(.063)	(.025)	(0)	-.004	(.025)	(.022)	(.03)
Severity group 6	(.013)	(.016)	(.016)	-.063*	-.036	-.036	.009	.01	(.016)	.011	0	0
				(.037)	(.026)	(.045)	(.007)	(.01)		(.025)	(0)	(0)
Others	.005	.013	.012	-.007	.018	.006	.003	.002	-.004	-.034***	.003	0
	(.013)	(.009)	(.018)	(.017)	(.017)	(.015)	(.003)	(.002)	(.012)	(.011)	(.003)	(0)
Asphyxia												
Severity group 1						.066***	0	0**			0	0
						(.005)	(0)	(0)			(0)	(0)
Severity group 2						.041***	-.005	-.013**			.001	.004
	.032	-.005	-.013	.039	-.019	(.015)	(.006)	(.006)	.021	-.01	(.006)	(.007)
Severity group 3	(.029)	(.024)	(.019)	(.029)	(.017)	-.001	-.008	.013	(.021)	(.015)	-.008	.003
						(.022)	(.012)	(.01)			(.011)	(.01)
Severity group 4						-.025	-.01	.065***			-.025	.017
						(.033)	(.024)	(.023)			(.023)	(.025)
Severity group 5				.018	.015	.052	-.022	.034		-.063*	-.033	.03
	-.004	.016	-.016	(.04)	(.029)	(.036)	(.047)	(.036)	-.002	(.034)	(.043)	(.03)
Severity group 6	(.034)	(.032)	(.033)	.048	.034	.039	.023	.087*	(.03)	-.035	.004	-.018
				(.03)	(.033)	(.046)	(.052)	(.048)		(.034)	(.044)	(.075)
Others	.057*	.058	.019	-.061*	-.007	.025	-.036***	0	.076*	-.047*	-.001	-.031***
	(.033)	(.041)	(.045)	(.033)	(.032)	(.026)	(.014)	(.012)	(.042)	(.028)	(.013)	(.012)

continued on next page

Table A.2: Differences above and below threshold by Severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Severe asphyxia												
Severity group 1						.012***	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						.012***	.001	-.003			.001	.003
	.012	-.019	-.004	.016	-.002	(.002)	(.002)	(.002)	-.002	0	(.002)	(.002)
Severity group 3	(.018)	(.016)	(.012)	(.015)	(.006)	.012***	-.003	.001	(.012)	(.008)	.001	-.002
						(.002)	(.005)	(.005)			(.005)	(.005)
Severity group 4						-.014	-.012	.008			.002	.011
						(.018)	(.012)	(.014)			(.013)	(.014)
Severity group 5				.032***	-.011	.021***	-.03	.005		-.032**	.014	.03
	.015	.003	-.018	(.009)	(.019)	(.007)	(.028)	(.03)	.009	(.013)	(.034)	(.03)
Severity group 6	(.022)	(.02)	(.023)	.032***	.03**	.007	.024	.001	(.019)	-.008	.064*	.008
				(.009)	(.013)	(.033)	(.026)	(.024)		(.023)	(.036)	(.041)
Others	.054*	.024	.055	-.007	-.019	-.004	-.016*	-.01	.031	-.023	-.009	-.012
	(.03)	(.035)	(.036)	(.023)	(.024)	(.022)	(.009)	(.009)	(.035)	(.018)	(.008)	(.008)
Moderate asphyxia												
Severity group 1						.052***	0	0*			0	0
						(.004)	(0)	(0)			(0)	(0)
Severity group 2						.028*	-.004	-.01*			-.001	.002
	.017	.018	-.01	.021	-.012	(.015)	(.006)	(.005)	.024	-.012	(.005)	(.006)
Severity group 3	(.024)	(.019)	(.015)	(.025)	(.016)	-.007	-.006	.011	(.018)	(.013)	-.01	.002
						(.021)	(.011)	(.008)			(.01)	(.009)
Severity group 4						-.012	-.004	.05***			-.028	.006
						(.028)	(.021)	(.017)			(.019)	(.021)
Severity group 5				-.012	.026	.03	-.001	.028		-.036	-.048*	0
	-.02	.022	.002	(.039)	(.022)	(.035)	(.038)	(.02)	-.007	(.03)	(.027)	(0)
Severity group 6	(.028)	(.025)	(.026)	.019	.002	.062***	-.001	.096**	(.025)	-.035	-.048**	-.027
				(.029)	(.03)	(.011)	(.047)	(.044)		(.026)	(.023)	(.066)
Others	-.002	.034	-.036	-.042*	.017	.026*	-.02*	.006	.045*	-.031	.013	-.02**
	(.016)	(.023)	(.029)	(.024)	(.022)	(.014)	(.01)	(.008)	(.027)	(.021)	(.01)	(.009)
IRDS												
Severity group 1						.521***	.011	.003***			0	.001
						(.01)	(.008)	(.001)			(.003)	(.001)
Severity group 2						.295***	.021	.076***			-.015	.004
	.01	-.02	.02	.274***	-.009	(.039)	(.013)	(.01)	-.037*	.021	(.011)	(.01)
Severity group 3	(.024)	(.023)	(.02)	(.06)	(.032)	-.035	.056**	.169***	(.022)	(.032)	.021	.03
						(.044)	(.024)	(.018)			(.023)	(.019)
Severity group 4						-.076	.044	.11***			-.038	-.017
						(.057)	(.039)	(.036)			(.039)	(.041)
Severity group 5				.118*	-.046	.054	.158**	.071		.016	.198**	-.131
	.006	.056*	.018	(.064)	(.042)	(.08)	(.075)	(.094)	-.001	(.048)	(.088)	(.111)
Severity group 6	(.027)	(.032)	(.036)	-.105***	-.02	.015	.272***	.221***	(.025)	.024	.109	-.112
				(.032)	(.03)	(.062)	(.079)	(.084)		(.032)	(.087)	(.097)
Others	.037	.113*	.025	-.277***	.107**	-.063	.007	.028	-.003	-.016	.007	.023
	(.048)	(.063)	(.066)	(.049)	(.048)	(.046)	(.022)	(.019)	(.059)	(.053)	(.022)	(.019)

continued on next page

Table A.2: Differences above and below threshold by Severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Maternal smoking												
Severity group 1						.043***	.026**	.005***			-.004	.001**
						(.004)	(.012)	(.001)			(.005)	(.001)
Severity group 2						.011	-.01	.003			.006	.005
	.007	-.004	.006	.023	.013	(.016)	(.007)	(.004)	-.008	-.003	(.006)	(.004)
Severity group 3	(.012)	(.011)	(.011)	(.015)	(.012)	.028**	-.003	.002	(.011)	(.011)	-.006	.006
						(.011)	(.01)	(.008)			(.009)	(.007)
Severity group 4						.017	-.001	-.006			-.007	.001
						(.019)	(.018)	(.014)			(.016)	(.015)
Severity group 5				.022***	.01	.026***	-.019	0		-.016*	0	-.019
	-.013	.002	-.016	(.008)	(.013)	(.008)	(.019)	(0)	-.005	(.009)	(0)	(.019)
Severity group 6	(.012)	(.012)	(.021)	-.009	-.022	.014***	.014*	-.037	(.014)	.015	.021	.011
				(.023)	(.018)	(.005)	(.008)	(.036)		(.019)	(.021)	(.028)
Others	-.017**	.02*	.008	-.023	.005	.004	-.008	.004	0	-.008	-.001	.002
	(.008)	(.011)	(.008)	(.015)	(.003)	(.003)	(.006)	(.004)	(0)	(.011)	(.005)	(.004)
Maternal alcohol use												
Severity group 1						.002**	0	0			0	0
						(.001)	(0)	(0)			(0)	(0)
Severity group 2						.002**	-.001	0			0	0
	0	-.002	-.002	.003**	0	(.001)	(.001)	(.001)	-.002	.005	(.001)	(0)
Severity group 3	(0)	(.004)	(.002)	(.001)	(0)	.002**	0	.001	(.002)	(.004)	-.002	.003*
						(.001)	(.002)	(.002)			(.001)	(.002)
Severity group 4						-.011	.001	0			-.005	-.002
						(.013)	(.005)	(0)			(.003)	(.002)
Severity group 5				0	.003	.002	0	-.023		-.005	0	0
	0	.003	.009	(0)	(.003)	(.002)	(0)	(.023)	-.004	(.005)	(0)	(0)
Severity group 6	(0)	(.003)	(.006)	0	-.008	.002	0	0	(.004)	0	0	0
				(0)	(.008)	(.002)	(0)	(0)		(0)	(0)	(0)
Others	0	-.01	0	0	-.007	.002	0	-.002	0	0	-.001	.001
	(0)	(.01)	(0)	(0)	(.007)	(.002)	(0)	(.002)	(0)	(0)	(.001)	(.002)
Maternal drug use												
Severity group 1						.007***	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						.007***	-.001	.003			.003	-.002
	-.003	-.004	.006	.003**	-.002	(.002)	(.002)	(.002)	.001	0	(.002)	(.002)
Severity group 3	(.003)	(.004)	(.004)	(.002)	(.003)	.007***	0	.008	(.003)	(.004)	.001	.006
						(.002)	(.004)	(.005)			(.005)	(.004)
Severity group 4						-.032	.018**	.004			.004	.021*
						(.022)	(.008)	(.011)			(.009)	(.012)
Severity group 5				-.014	0	.002	.016	0		0	0	0
	0	0	.005	(.019)	(0)	(.002)	(.011)	(0)	0	(0)	(0)	(0)
Severity group 6	(0)	(0)	(.005)	.005	-.012	.008**	-.022	.01	(0)	-.011*	0	-.027
				(.004)	(.012)	(.004)	(.022)	(.01)		(.007)	(0)	(.019)
Others	0	.007	0	0	-.004	0	0	.003	0	.004	-.001	0
	(0)	(.007)	(0)	(0)	(.007)	(0)	(.003)	(.003)	(0)	(.008)	(.005)	(.003)

continued on next page

Table A.2: Differences above and below threshold by Severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Hospital-related controls												
In-patient ward												
Severity group 1						-.167	-.024	-.012***			-.007	-.005
						(.108)	(.018)	(.004)			(.01)	(.003)
Severity group 2						0	-.005**	-.024***			.002	-.012**
	.003	0	0	0	0	(0)	(.002)	(.005)	0	0	(.003)	(.006)
Severity group 3	(.003)	(0)	(0)	(0)	(0)	0	.001	-.01***	(0)	(0)	-.002	-.013***
						(0)	(.001)	(.004)			(.002)	(.005)
Severity group 4						0	0	0			-.002	.004
						(0)	(0)	(0)			(.002)	(.004)
Severity group 5				0	0	0	0	0		0	0	0
	.005	0	0	(0)	(0)	(0)	(0)	(0)	0	(0)	(0)	(0)
Severity group 6	(.005)	(0)	(0)	0	0	0	0	0	(0)	0	0	0
				(0)	(0)	(0)	(0)	(0)		(0)	(0)	(0)
Others	-.004	-.033	-.044	.022	-.028*	-.016	-.023	-.037**	.007	.003	-.026	-.01
	(.014)	(.027)	(.028)	(.017)	(.017)	(.023)	(.016)	(.015)	(.017)	(.015)	(.017)	(.018)
In-patient midwife												
Severity group 1						-.079	-.031**	-.009**			.005	-.002
						(.08)	(.012)	(.004)			(.008)	(.003)
Severity group 2						.004***	-.01***	-.015***			0	-.003
	0	.001	-.001	.001	-.001	(.001)	(.003)	(.005)	0	-.002	(.004)	(.006)
Severity group 3	(.005)	(.001)	(.003)	(.001)	(.005)	.004***	.006**	-.012**	(0)	(.004)	-.003	-.014**
						(.001)	(.003)	(.005)			(.004)	(.005)
Severity group 4						.004***	-.001	-.007			.009	.012
						(.001)	(.007)	(.006)			(.007)	(.011)
Severity group 5				.003	0	0	0	0		0	0	0
	.005	.003	.005	(.003)	(0)	(0)	(0)	(0)	0	(0)	(0)	(0)
Severity group 6	(.005)	(.003)	(.005)	.003	.005	.004	0	.021	(0)	0	0	0
				(.003)	(.003)	(.003)	(0)	(.015)		(0)	(0)	(0)
Others	-.009	-.019	-.072	.014	-.029	.032	.022	-.029*	.021	-.012	0	.019
	(.018)	(.033)	(.045)	(.019)	(.025)	(.024)	(.018)	(.016)	(.029)	(.021)	(.019)	(.018)
In-patient doctor												
Severity group 1						-.249**	-.034*	-.015***			-.004	-.003
						(.125)	(.018)	(.004)			(.01)	(.003)
Severity group 2						-.007	-.004	-.024***			.001	-.008
	0	0	-.007	.001	.001	(.008)	(.003)	(.005)	.006	-.003	(.004)	(.006)
Severity group 3	(.005)	(0)	(.005)	(.001)	(.001)	.001	-.002	-.013***	(.004)	(.002)	0	-.014***
						(.001)	(.003)	(.004)			(.003)	(.005)
Severity group 4						.001	.006*	-.001			-.007*	.002
						(.001)	(.003)	(.004)			(.004)	(.005)
Severity group 5				.005	.003	0	0	.014		0	0	0
	.005	.005	-.007	(.004)	(.003)	(0)	(0)	(.014)	-.011	(0)	(0)	(0)
Severity group 6	(.005)	(.004)	(.014)	-.01	.01**	.006*	0	.01	(.008)	-.004	-.012	-.014
				(.016)	(.005)	(.004)	(0)	(.01)		(.004)	(.012)	(.014)
Others	.007	-.01	-.028	.029***	-.004	-.024	-.004	-.037**	.025*	.003	-.025	-.001
	(.009)	(.02)	(.022)	(.011)	(.011)	(.023)	(.015)	(.015)	(.014)	(.015)	(.015)	(.017)

continued on next page

Table A.2: Differences above and below threshold by Severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Number of births												
1 - N_L						2497	404	7988			1509	16474
N_R						12	778	13960			2984	29625
2 - N_L						2497	2384	3575			1929	2354
N_R	294	745	641	1483	813	124	1439	2790	361	365	3130	4061
3 - N_L	304	264	472	70	328	2497	1327	1402	528	636	756	1041
N_R						135	670	1312			1201	1718
4 - N_L						2497	536	463			257	227
N_R						77	225	317			433	406
5 - N_L				370	302	423	125	71		99	33	33
N_R	218	373	221	52	98	30	53	44	250	186	63	52
6 - N_L	225	142	147	370	419	484	216	97	275	132	47	41
N_R				65	118	31	45	52		262	84	74
Others - N_L	170	150	127	243	423	477	673	836	121	135	514	655
N_R	234	101	84	160	149	129	596	909	170	264	882	1158

Notes: Difference between means below and above birth weight threshold T within a ±25 gram bandwidth in each Severity group ($\bar{X}_{below}$-$\bar{X}_{above}$). (Robust)Standard errors in parentheses below mean difference . N_L: Number of observations left/below threshold within 25 gram bandwidth. N_R: Number of observations right/above threshold within 25 gram bandwidth. *** p<0.01; ** p<0.05 and * p<0.1. For variable definition see table 1 **Source:** Own calculations based on the DRG-Statistic.