Detecting dishonest behavior in medical decision-making: Evidence from a lab experiment on DRG upcoding in neonatology

Heike Hennig-Schmidt, Bonn University, Germany
Hendrik Jürges, Wuppertal University, Germany
Daniel Wiesen, Cologne University, Germany

Objectives

Prospective payments systems (diagnosis-related groups, DRG) to reimburse hospital inpatient care became very popular two decades ago (Ellis and McGuire, 1993). DRGs, however, may incentivize dishonest provider behavior, like inappropriately early discharge of patients, withholding of (necessary) diagnostics and therapies, cream-skimming of low-cost patients or coding of patients into groups with higher reimbursement, so-called DRG upcoding. DRG systems provide a strong financial incentive for dishonest behavior (e.g. Silvermann and Skinner 2004).

Our experimental study investigates the latter phenomenon by focusing on DRG-upcoding behavior in neonatology, i.e., underreporting birth weights of early-born infants. In Germany, DRG rates are based on birth weights classified along eight thresholds. Differences of a few grams can induce more than 15,000 Euro additional payments for neonatal care units. Thus, the financial incentives to manipulate birth weights documented in the birth record by a few grams can be strong for the neonatal care unit. Using data from official German birth statistics, Jürges and Köberlein (2015) found a significant increase in the reported number of birth weights below and a significant decrease above the thresholds relevant for reimbursement. A plausible explanation for such a shift in the distribution of birth weights of early-borns is upcoding.

At the same time, false documentation of birth weights is practically non-verifiable ex post because infants generally lose about 10% of their initial weight in the first days after birth. DRG upcoding in neonatology leads to considerable welfare losses. Jürges und Köberlein (2015) found that since the introduction of DRGs in 2003, hospitals have gained additional reimbursement in excess of 100 million Euro.

DRG upcoding is interesting also for other reasons. In particular, it is not done by some revenue-maximizing managers in the hospital administration, who usually have no contact with patients. In contrast, it is doctors or nurses who report too low birth weights. It is, therefore, important to better understand behavior and underlying motivations of decision-makers in a setting that allows for dishonest behavior. In this respect, it is also crucial to analyze the role of detection and audit to reduce fraud (see e.g. Becker et al. 2005, Lindeboom et al. 2015).

Methods

To contribute to providing the missing insights we apply the method of a controlled incentivized laboratory decision experiment. We developed a new experimental design that mimics the main aspects of medical-decision making in a neonatology care-unit setting. In particular, our design allows us to analyze honest and dishonest reporting of birth weights by means of individual decision data. Participants (medical and non-medical students) in a double-blind setting enter given birth weights into a computerized birth report (N = 98). Following standard experimental economic methodology, participants in the experiment are incentivized with real monetary payoffs.
The lab experiment is framed in a medical decision-making context. Participants decide in the role of employees in a neonatology care unit. Actual weights of newborns are given to subjects on their screens. Next, the participant has to record the weight in the newborn’s medical records. The reported birth weight determines the DRG-payment. Participants’ decisions are incentivized by a lump-sum payment and a DRG-based component. The latter depends on the individually reported birth weights. Participants are informed about DRG-rates, treatment costs, and their profits before entering birth weights. We vary the probability of detecting fraud by running conditions with and without a 10 percent chance of discovery. If dishonest behavior is detected, participants lose the variable payoff component. Participants also answered a post-experimental questionnaire that enables us to investigate whether personal characteristics correlate with honest or dishonest behavior. Finally, we ask participants for their decision motives.

Results

We find strong evidence for dishonesty (DRG upcoding) in our behavioral data. In the No-Detection condition where detection of upcoding is not possible, a rather large fraction of participants show fully dishonest behavior: 71% of participants are “full” liars, 27% are “partial” liars, entering 1 to 4 wrong birth weights, and only 2% are fully honest. This finding is in stark contrast to e.g. Fischbacher and Föllmi (2013) who found only 22% full liars, 39% partial liars, but 39% fully honest participants in their experiment (see also Abeler et al. 2016 for a survey on experiments using the Fischbacher-Föllmy approach).

When a 10% detection probability is introduced, full-lying behavior vanishes. Participants are partially dishonest only which replicates the findings of Jürges und Köberlein (2015) that in German hospitals only 'partial' upcoding seems to be practiced. The probability of audits reduced the proportion of dishonest answers by 27 percentage points. However, this number underestates the true effect of the audit. If we distinguish further between birth weights where gainful misreporting can or cannot be detected, we find 88% dishonest answers if misreporting cannot be detected but only 41% when dishonesty can be detected. In other words, introducing a 10% detection probability has reduced dishonesty by half. Moreover, the degree of dishonesty varies with personal characteristics (gender, personal values like Neuroticism). Findings differ between medical and non-medical students.

Conclusion

Our behavioral data provide strong evidence for DRG upcoding. We do not only observe partially dishonest behavior but, surprisingly, fully dishonest behavior to an even larger extent. Very few subjects always truthfully report the birth weights. We also collected subjects' stated motivations for their decisions in a post-experimental questionnaire. Apart from the pure profit-maximizing motive, subjects use the framed context of the experiment as a justification for misreporting the birth weights (e.g., the early-borns do not suffer as their medical treatment depends on the actual not the reported birth weight or making losses has a negative impact on the hospital’s neonatal care unit). Further stated motives are that other participants are not affected, the artificial laboratory situation (in real life 'of course' weights would have been reported honestly), the difference between actual and reported birth weight should not be too large, and the lack of a controlling entity in the No-Detection condition such that misreporting cannot be detected.

Our study also contributes to the experimental literature analyzing dishonest behavior. The rather high share of fully dishonest behavior in the No-Detection condition is in contrast to the literature that mainly reports partial dishonesty (e.g. DePaulo et al., 1996; Dreber and
Johannesson, 2008; Mazar et al., 2008; Erat and Gneezy 2012; Conrads et al. 2013; Fischbacher and Föllmi-Heusi, 2013; Jiang, 2013; Abeler et al. 2014, 2016; Conrads et al. 2014; see also Gächter and Schulz 2016), even though in different settings. Our novel design provides an experimental benchmark condition for inducing a high percentage of participants to display dishonest behavior. This is notable as our design allows us to track whether given and reported birth weights coincide. This feature of our experiment differs from most of the above studies that are only able to test behavior against normal distributions. To the best of our knowledge, we are one of the two studies that analyze dishonest behavior with individual decision data (see Abeler et al. 2016 for a related experiment in a different context).

Literature