Title

Impacts of Hospital Wait Time on Health and Labor Supply

Authors

Godoey, Anna
Haaland, Venke Furre
Huitfeldt, Ingrid
et al.

Publication Date

2019-11-01

IRLE
Institute for Research on Labor and Employment

IRLE WORKING PAPER
#111-19
November 2019

Impacts of Hospital Wait Time on Health and Labor Supply

Anna Godøy, Venke Furre Haaland, Ingrid Huitfeldt, and Mark Votruba

Cite as: Anna Godøy, Venke Furre Haaland, Ingrid Huitfeldt, and Mark Votruba. (2019). "Impacts of Hospital Wait Time on Health and Labor Supply". IRLE Working Paper No. 105-17.
https://irle.berkeley.edu/files/2019/11/Impacts-of-Hospital-Wait-Time-on-Health-and-Labor-Supply.pdf

Impacts of Hospital Wait Time on Health and Labor Supply*

Anna Godøy[†] Venke Furre Haaland[‡] Ingrid Huitfeldt[§] Mark Votruba[¶]

November 12, 2019

Abstract

We estimate the effects of wait time for orthopedic surgery on health and labor market outcomes of Norwegian workers. Our identification strategy exploits variation in wait times for surgery generated by the idiosyncratic variation in system congestion at the time of referral. While we find no significant evidence of lasting health effects, longer wait times have persistent negative effects on subsequent labor supply. For every 10 days spent waiting for surgery, we estimate health-related workplace absences increase 8.7 days over the five years following referral, and the likelihood of permanent disability insurance increases by 0.4 percentage point. Cost benefit calculations point to sizable fiscal savings from shorter wait times.

JEL codes: I120, J320.

Keywords: Wait time, queues, hospital treatment, health outcomes, labor market attachment, sickness absence.

*This research has received support from the Research Council of Norway (grants #214338 and #227117); the Western Norway Regional Health Authority; and the research department at Stavanger University Hospital. Data made available by Statistics Norway and the Norwegian Patient Register have been essential. We are grateful to Martin E. Andresen, David Bishai, Sverre Kittelsen, Edwin Leuven, Mari Rege, Kjetil Telle, and participants at the AEA/ASSA meeting (Philadelphia, 2018), NBER Summer Institute (2017) and EEA-ESEM (Geneva, 2016) for valuable comments.

†University of California, Berkeley, and Statistics Norway, anna.godoy@berkeley.edu

‡University of Stavanger, Norway and Statistics Norway. venke.f.haaland@uis.no

§Statistics Norway and the Frisch Centre, Oslo, Norway. ingrid.huitfeldt@gmail.com

¶Economics Department, Weatherhead School of Management, Case Western Reserve University and Center for Healthcare Research and Policy at MetroHealth Medical Center. mark.votruba@case.edu

1 Introduction

Queues are a ubiquitous feature of universal health care systems, and an issue of persistent public concern. Universal systems employ queues to handle excess patient demand under existing capacity constraints (Martin and Smith, 1999), leaving a constant backlog of patients awaiting care. This backlog fluctuates over time with the (irregular) flow of new patients and the magnitude of that flow relative to the system's throughput capacity, often resulting in significant wait times for non-emergency surgery and medical procedures that can vary widely over time and across systems. For example, the average wait time for a hip replacement in 2014 was 91 days in the UK and 152 days in Norway.[1] In principle, wait times could be reduced by expanding a system's delivery capacity, which highlights the inherent tension between the goal of cost containment and the goal of delivering timely care. Policymakers inevitably have to resolve this tension, but currently operate with limited information about the costs associated with longer waits. Our paper seeks to better inform such decisions by providing evidence on the labor supply, benefit and health care utilization effects of longer wait times on adult Norwegian workers referred for orthopedic procedures. In so doing, our paper adds to a young but growing literature on the effects of waiting for non-emergency treatment, albeit one that has focused almost entirely on the health implications of waiting.

The potential costs associated with longer wait times are multi-faceted. At a minimum, waiting imposes welfare costs on patients seeking treatment by extending the period of time the patient remains debilitated. While waiting, patients are often unable to work and frequently utilize sickness leave benefits, with short-term consequences for productivity and government finances. Longer wait times could also have implications that extend beyond a patient's treatment and (usual) recovery period. If lengthy wait times reduce the efficacy of treatment, as might be the case if a patient's health deteriorates while waiting for treatment, longer waits could have long-term health consequences for affected patients (Malmivaara et al., 1995),

[1] Figures obtained from *OECD.Stat* at https://data.oecd.org/health.htm Health Care Utilisation/Waiting times (date retrieved 10/28/2016).

reducing their future productivity and increasing their future utilization of sickness-related benefits and healthcare services.[2]

Importantly, long run effects on labor supply and benefit utilization outcomes are possible even in the absence of permanent health effects for at least two reasons. First, longer times spent unable to work could contribute to human capital depreciation, including loss of network and lower productivity (Rees, 1966; Mincer, 1974; Mincer and Ofek, 1982; Becker, 1991; Calvó-Armengol and Jackson, 2004). Being rendered a "less valuable" or "less connected" worker might reduce the future utility cost of taking temporary work absences or leaving employment. Second, individual preferences for work and workplace absenteeism could conceivably be affected if a person is forced to experience a longer period in a work-disabled state. Drawing on theories of social identity (e.g. Sowell, 1975, 1981; Hofstede and Bond, 1988; Barke et al., 1997; Sowell, 2005; Chiswick, 1983; Murray, 1984) and endogenous preference formation (Bowles, 1998), a patient's self-image is potentially altered by experiencing an extended period of work incapacitation. If longer wait times increase the likelihood of a patient self-identifying as "work debilitated" or "disabled", this could increase that individual's propensity for future sickness-related work absences.[3]

Identifying a causal effect of patient wait times on labor and health outcomes is challenging, as wait times are presumably affected by patient characteristics that we cannot observe but might independently affect the outcomes of interest. In the Norwegian healthcare context, more serious cases are given priority over less serious ones, which leads to healthier patients generally having longer wait times than sicker patients. As a result, standard regression estimates would be expected to be biased towards findings of better health outcomes and lower workplace absenteeism among

[2] An extensive literature examines the impact of health on labor market outcomes (see, e.g. Stephens Jr and Toohey, 2018, for a review).

[3] Sociological theories on role, stigma and labeling (Parsons, 1951; Goffman, 1963) suggest that interaction with the health care system and receiving a diagnosis can contribute to labor force detachment. Parsons (1951) argued that transitioning from roles like "healthy" or "employed" to roles like "sick" or "disabled" is associated with new rights and new obligations. Sick individuals are expected to seek and comply with the advice of the health care system, but in return they are freed from culpability for their illness and exempted from everyday social roles like the obligation to provide for oneself and ones family through employment. Having the illness certified by the medical profession, by being attributed an official diagnosis or being eligible for health-related welfare, may ease the transition from a role of worker to the role of sick person.

patients with longer waits. On the other hand, patients with more resources might be more skilled at navigating the health care system, enabling some degree of queue jumping even within a public system. This channel could introduce a negative bias in the relationship between wait time and later outcomes.

We address these endogeneity concerns by employing an instrumental variable (IV) approach that exploits the idiosyncratic variation in system congestion facing different patients based on the time when they enter a particular queue for treatment. Specifically, we instrument for patient i's wait time with the average wait time of other patients queuing for the same procedure at the same hospital around the same time as patient i, while also controlling for general time and hospital factors. This empirical approach is enabled by rich administrative data covering the entire population of Norway, matched with unique individual patient data comprising all visits to general practitioners (GPs) and to publicly-funded specialists and hospitals.

The crucial identifying assumption for our IV approach is that the patients who enter a queue when wait times are long are not systematically different from patients entering the same queue when wait times are short. As Martin and Smith (1999) have argued, wait times could operate as a rationing device that causes some people to forego care or opt for a private alternative when the queues for publicly-financed care are long. If so, differential selection of patients away from "long queues" could lead to a potential violation of our identifying assumption.[4] While we cannot fully rule out such concerns, since we cannot observe patients who opt for private care or forego treatment altogether, our rich data allow us to carefully investigate the plausibility of our identifying assumption by exploring the correlation between our instrument, congestion, and a battery of observable individual characteristics such as age, education, income, prior labor market attachment and health care history. Importantly, we find no evidence that patients referred during periods of long ex-

[4]Empirically, there is mixed evidence on the effectiveness of queues as a rationing device in health care. Martin and Smith (1999) find that demand for treatment is relatively inelastic with respect to wait times, while Martin and Smith (2003) find demand elasticities for elective surgery between negative .1 and .2 (-0.07 for orthopedics). Finally, Sivey (2017) studies emergency department waiting times, and estimates that the waiting time elasticity of demand for low-urgency patients is approximately -0.25. This setting differs from ours, however, in that patients are physically waiting in the emergency room (as opposed to waiting at home for elective treatment).

pected wait time are different from patients referred for treatment in periods of short expected wait time. This finding suggests any bias arising from differential selection away from long queues is likely to be small.

Our paper draws on data from orthopedic surgical procedures. Orthopedics is an interesting context for exploring wait time effects for at least two reasons. First, musculoskeletal conditions are the leading causes of health-related work absence, constituting about 40% of all sick leave spells in Norway (Brage et al., 2013). Thus, wait time effects in the context of orthopedic surgeries could have labor supply and fiscal implications of particular importance to policymakers. Second, because orthopedic conditions are rarely life-threatening and the efficacy of orthopedic surgeries is not believed to greatly depend on wait time, policymakers and hospital administrators may feel less compelled to ensure prompt service to orthopedic patients.[5] As a result, individual wait times for (non-emergency) orthopedic patients are driven to a substantial degree by the backlog of patients in the queue when a new patient is referred for treatment.

Evidence of the causal relationship between wait times and the medical efficacy of orthopedic surgeries is rather thin, with most studies suffering from low power and/or questionable identification strategies (multiple regression models estimated on observational data). Medical research on the effects of waiting for knee and hip surgery has mostly focused on whether health and functional status decline as a patient waits for treatment as opposed to whether waiting contributes to poorer post-surgical outcomes. As Hoogeboom et al. (2009) document in their systematic review, little support has been found for such effects.[6] Employing a large observational sample drawn from the British National Health Service (NHS) and controlling for a rich set of covariates, Nikolova et al. (2016) estimated a significant negative association between wait times for hip and knee replacement surgery and post-surgical health indicators at 6 months; however, the magnitude of the estimated effects was

[5]Of the 10 surgical procedures for which the OECD tracks patient wait times, two are orthopedic procedures – knee and hip replacement surgeries (see OECD 2013).

[6]One study designated as "high quality" by Hoogeboom et al. (2009); Kapstad et al. (2007) found evidence of a small but statistically significant reduction in self-reported functional status at the time of surgery for patients who waited longer for knee replacement surgery.

very small. Hamilton and Bramley-Harker (1999) exploit the decrease in surgical wait times occurring as a result of NHS reforms and find no evidence that the post-operative health outcomes of hip fracture patients were substantially affected by the reductions in wait time. Evidence from randomized clinical trials conducted in Finland also found no evidence that longer wait times for total knee replacement or total hip replacement led to poorer health status at surgery (Hirvonen et al., 2007, 2009) nor any evidence of health differences 3 and 12 months after surgery (Tuominen et al., 2009, 2010). Notably, the average wait times in these studies were roughly half those in our Norwegian sample. If wait time effects are convex, we might anticipate larger negative health effects in our setting than these studies suggest.

Our data lack measures of self-reported health or physical functioning, but broadly support the notion that the long-term health effects of longer wait times are probably small, though (in light of the point estimates) more likely to be negative than positive. Our IV estimates of wait time effects on general practitioner (GP) visits and hospital stays over the five year period following referral are positive but generally small and statistically nonsignificant. Results pertaining to the probability of resurgery do not indicate that longer waits undermine the efficacy of treatment. Mortality rates were also unaffected, but are a poor proxy for health outcomes in this context.

In contrast, we find significant evidence that longer waits contribute to substantial increases in health-related work absences. Over the five years following referral, an additional 10 days spent waiting for treatment increases health-related work absences by an estimated 8.7 days. While some of this is due to extended sick leave while a patient awaits surgery, long waits induce higher levels of health-related absence extending into the fifth post-referral year, well after the recovery period for the vast majority of our subjects. An additional 10 days of waiting also increases the probability of a patient entering the permanent disability program by 0.4 percentage point by the end of year 5. A substantial fraction of the increase in health-related absences in year 5 can be attributed to this increase in disability participation.

We also uncover substantial heterogeneity in the impact of wait time on health-related absence and disability benefit receipt. Our aggregate findings appear largely driven by workers who were already on sick leave at the time of referral, who comprise just under 25 percent of our sample. Among these workers, 10 additional days of wait time is estimated to increase health-related absence by 27.2 days, with disability participation rates 1.3 percentage points higher by the end of year 5. These findings are potentially consistent with theories of habit formation and endogenous preferences, as we find no evidence of larger health utilization effects among these workers. We also find significantly larger effects on the labor outcomes of less educated workers, though again without any indication of larger health care utilization effects in this group.

Our findings also point to substantial fiscal costs arising from longer wait times. Using data on sickness-related benefit transfers, we estimate that an additional 10 days of wait time leads to an increase in transfers totaling around NOK 6,400 (or USD 740) over the five years following referral.[7] Back-of-the-envelope calculations suggest opportunities for substantial fiscal savings from efforts to reduce wait times under plausible assumptions about the costs incurred by such as effort.

Although an extensive literature examines the impact of health on labor market outcomes (see, e.g. Stephens Jr and Toohey, 2018, for a review) we are not aware of any papers that specifically address the relationship between hospital wait times and labor market outcomes.[8] The closest work to ours is Aakvik et al. (2015) who analyze the effect on sickness absence of being exposed to a reform in Norway that aimed at reducing wait time. They do not, however, explicitly estimate effects of wait time, but rather identify a reform effect.[9] Moreover, their sample includes only people who are on sick leave before admission to the hospital, and define wait time as days from the first day of the absence spell until treatment. Our approach exploits the exact date of referral to the hospital, and we can therefore additionally include people who

[7]This figure is discounted to the date of referral employing a 3% annual discount rate.

[8]Andrén and Granlund (2014) potentially qualifies, though the explicit goal of the paper is to evaluate the robustness of other parameters in a labor supply ("return-to-work") model when wait time is also controlled for, with no effort to address the endogeneity of wait times. Perhaps as a consequence, the authors finds surgical patients with longer waits had faster returns to work.

[9]The reform, 'Faster Return to Work', is discussed in Section 2.

are not on sick leave on referral date. Furthermore, we augment our analyses with measures of healthcare utilization (including resurgery), which allows us to evaluate whether the observed wait time effects on labor market outcomes are explained by poorer health outcomes, or more likely to be the result of behavioral factors (endogenous preferences) or human capital depreciation. To our knowledge, only one study attempts to investigate a causal relationship between prolonged sickness absence and work force detachment. Hultin et al. (2012) utilize Swedish Public Health Survey data and regress long term sick leave on future disability participation. The study demonstrates that even when controlling for a rich set of self-reported health measures there is a large and significant association between long term sick leave and future disability participation. This is consistent with the hypothesis that long term sickness absence fosters future labor market detachment, but could reflect unobserved differences (including differences in preferences) that are not captured by self-reported measures of health.

The rest of the paper is organized as follows: Section 2 gives an overview of the relevant institutions. Data is presented in section 3, and section 4 lays out our empirical strategy. Results are presented in section 5 and section 6 provides a conclusion.

2 Institutional Setting

Hospitals. Somatic specialist health care in Norway is funded primarily through taxes and transfers from the national government. Access to hospital services is either via emergency admissions or through referrals from general practitioners acting as gatekeepers, who are responsible for all initial assessment, examinations and treatment of patients. Patients who are referred to hospital services are typically assigned a hospital on the basis of their home address, but are free to choose the hospital at which they want to receive treatment. In practice, however, choice is often limited due to vast geographic distances, and 80% end up receiving care at their local hospital (Godøy and Huitfeldt, 2018).

Patients pay a very low or zero price for using hospital services.[10] In addition to explicit rationing by gatekeepers, utilization is rationed by wait times, aiming at prioritizing patients according to their medical need for health care. After an individual has been referred for specialist health treatment, the patient is assigned either a priority status or a non-priority status. Patients with priority status receive an assigned 'time limit' denoting the time by which the patient should receive treatment. The time limit is assigned by health professionals based on the patient's medical condition and the expected efficacy of the treatment and, since 2007, on his or her labor market attachment. This last criterion was the consequence of a 'Faster Return' reform (FRW), the purpose of which was to decrease the wait time for those who were on sick leave while waiting for treatment, promoting a faster return to work. The reform allocated the hospitals additional resources to provide individuals on sick leave with fast treatment, while, theoretically, not affecting the wait time of other patients without FRW status.

Health-related benefits: sickness absence and disability insurance. Employees usually receive sick pay equivalent to their regular salary from the first day of sickness absence. Expenses during the first 16 days are covered by the employer, while the Norwegian Labour and Welfare Service (NAV) takes over the responsibility on the 17th day of sick leave. The wage replacement ratio for sick pay is 100% and benefits can be maintained for up to 12 months.[11] After 12 months of continuous absence, patients are no longer eligible for sick pay. Persons who are still unable to work after one year of sickness may apply for temporary or permanent disability benefits. Disability insurance benefits amount to 66% of the applicant's wage.[12] All health-related benefits must be certified by a physician. While the exact rules regarding

[10] Patients' health care expenses are mainly subsidized by national insurance schemes. Some services, such as outpatient visits and visits to primary care physicians are subject to small co-payment rates. In 2015, the out-of-pocket payment for an outpatient procedure was NOK 320 (USD 40). However, once a patient's yearly total out-of-pocket health care expenditures exceed about NOK 2,100 (USD 260) all further expenses within that calendar year are reimbursed.

[11] Benefits are capped at higher earnings; in 2015, the benefit cap was approximately NOK 540,000 or around USD 68,000. However, all public sector workers and many private sector workers are covered by employer-provided top-up insurance.

[12] DI benefits are calculated based on the three best years among the 5 latest years before sickness. Benefits are capped at about NOK 540,000 or around USD 68,000.

temporary disability insurance have changed over time, during the sample period temporary disability benefits could normally be claimed for up to four continuous years.

Appendix figure A1 illustrates a stylized timeline of the different health-related benefits for a person commencing sick leave with full eligibility who continuously claims benefits. This timeline, while highly stylized, illustrates how a single absence spell may span several different types of benefits, as patients exhaust eligibility for each specific benefit. This could potentially complicate our empirical analysis. For example, longer wait times could increase the likelihood that patients exhaust their sick pay benefits in the first year after referral. This could show up in the data as a negative correlation between wait times and sick pay in year 2. However, this effect should not be interpreted as a causal reduction in sickness absence, as it would be arising mechanically from the eligibility cutoffs in the sick pay rules. To address such complications, our preferred empirical models will instead study health-related absence as a whole, without distinguishing between the types of benefit payments, as well as permanent disability benefit at year 5.

3 Data and Descriptives

3.1 Data Sources

The empirical analysis is based on data that combine several administrative registers obtained from Statistics Norway and the Norwegian Directorate of Health. A unique personal identifier is provided for every Norwegian resident at birth or upon immigration, enabling us to match the wait list records with administrative data on the entire resident population of Norway. Data provided by Statistics Norway contain birth and death dates, sex, district and municipality of residence, country of origin, education, occupation, annual earnings and health-related benefits. Our preferred measure of earnings comprises labor income only, excluding any social insurance benefits. Information on sickness absence and disability benefit receipt comes from social security registers that contain complete records for all individuals. As employers are responsible for the initial period of sickness-related absence, administrative social security data only identify sick leave spells lasting at least 17 days.

The Norwegian Patient Register contains complete patient level observations for all somatic public hospitals and private hospitals contracting with regional health authorities in Norway since 2008. Records include hospital identifiers, patient identifiers, main and secondary diagnoses (ICD10), surgical/medical procedures (NCSP/NCMP),[13] DRG cost weight,[14] exact time, date and place of admissions, discharges and, since 2010, the date at which the hospital received the referral. In addition, all publicly funded visits to primary care or specialists have been recorded electronically since 2006 in the Control and Payment of Health Reimbursement (KUHR) database. These data include patient identifier, date of visit, diagnosis, reimbursement code and size of patient deductible.

3.2 Sample

The starting point of the sample is all individuals referred for orthopedic surgery in 2010 or 2011. This includes all planned admissions with non-missing date of referral. We identify orthopedic procedures as surgical procedures based on the recorded Classification of Surgical Procedures (NCSP) codes, using the first two digits of the NCSP codes to identify 5 distinct procedures.[15] We exclude observations with wait times of longer than two years from the sample employed to construct the instrument, as these are likely to represent erroneous records.[16] This yields a referrals sample of 69,257 individuals. This is the sample used to construct our instrument.

[13]Surgical procedures are coded according to the NOMESCO Classification of Surgical Procedures (NCSP). Medical procedures are classified according to NCMP - Norwegian classification of medical procedures.

[14]Each patient discharged from a somatic hospital is assigned a DRG group that uniquely determines the reimbursement rate. Patients within the same DRG group are theoretically homogeneous with respect to both medical criteria and financial costs of treatment. Main diagnosis, comorbidities, medical and surgical procedures, age, and resource consumption, are crucial components when allocating patients to a particular group.

[15]See Appendix A for NCSP codes included.

[16]Note that while patients who wait more than two years are removed from the sample before constructing the instrument, they are retained in the estimation sample, as the probability of waiting more than 2 years may be endogenous to congestion.

The main estimation sample is a subset of the queue sample. We retain only patients with a likely attachment to the labor market, excluding individuals younger than 25 and older than 60 who earned less than twice the substantial gainful activity level in the year before referral (about NOK 180,000 in 2017).[17] We exclude patients who, two years before referral, were either receiving long term disability benefits or were absent from work for more than half of that year.[18]

For each patient referred for surgery, we construct a measure of observed wait time as the number of days spent waiting from the referral date to the first observed treatment date. To reiterate, these observed wait times likely reflect a number of factors, including patient health status, as well as idiosyncratic fluctuations in capacity and congestion. For each observation in the estimation sample, we use a subset of the queue sample to construct an instrumental variable, which we refer to as "congestion", as the average wait time of patients referred to the same hospital and same procedure in a set time window immediately preceding the focal worker's referral date. Our baseline specification calculates congestion using patients referred in the preceding 30 day window. In the results section, we explore the robustness of our findings to varying the choice of window length.[19] In other words, this is calculated using the full sample of referrals, without conditioning on labor force attachment or age. The number of patients fluctuates over time: in the estimation sample, we exclude any hospital-procedure groups where the number of referrals in

[17]The substantial gainful activity level ('basic amount') corresponds to NOK 93,634 (USD 12,000) in 2017. The 'basic amount' is used by the Norwegian Social Insurance Scheme to determine eligibility for and the magnitude of benefits like old age pension, disability pension, and unemployment compensation. The 'basic amount' is adjusted annually by the Norwegian Storting (parliament) to account for inflation and general wage growth. Following previous studies (Havnes and Mogstad, 2011a,b), we define employment (part-time or full-time) as earnings above twice the 'basic amount'.

[18]Put differently, we exclude patients who were already partly out of the labor force, including DI recipients even if they meet the earnings threshold and patients who were absent due to health reasons for more days than they were actually working. Note that this means we potentially include a small number of patients who commence receiving DI in the calendar year prior to referral date; the inclusion of these individuals may dilute the estimated effects on labor market outcomes if these patients are likely to leave the labor force regardless of their assigned wait times for surgery. On the other hand, restricting the sample allows us to better assess the validity of our empirical approach by testing for problematic pre-trends in DI enrollment, that is, whether patients who encounter more congestion have higher rates of DI entry in the year before referral. Results are qualitatively robust to dropping these restrictions.

[19]Note that we do not include patients who are referred on the same date or later; if hospitals assign wait times in the order referrals are received, their wait times are potentially endogenous to focal worker wait times.

a 30-day window ever dips below 3.[20] [21] This leaves us with a sample of 26,410 individuals in the main estimation sample. There are 27 hospitals in our sample, with data on 5 distinct orthopedic procedures (see Table A1 for description and volume of included procedures). In total, this amounts to 104 groups of hospitals-by-procedures, as not all procedures are performed at all hospitals.

The sample is merged to data on individual observable characteristics - demographics and education - as well as health and labor market outcomes covering the first five years after referral. Table 1 presents summary statistics of the sample. We include the following health outcomes, which are summed over the five years following referral date: (i) number of visits to the general practitioner (GP); (ii) number of days in hospital (including the surgery day); (iii) hospital utilization in NOK, calculated by summing the DRG weights; (iv) resurgery, defined as the number of visits to the hospital within the same diagnostic group for which the patient is waiting; (v) days at the hospital for emergency admissions; and (vi) mortality, measured as death within five years of referral. The resurgery variable may be of special importance as an indicator of whether treatment efficacy declines with longer wait times, as it is arguably more likely to capture variation in utilization that is directly related to the original reason for referral.[22] Labor market outcomes are: (i) total health-related absence from work over the five years following referral date, including sick leave and longer-term disability benefits, and the following variables measured in the 5th year after referral: (ii) an indicator variable for receiving disability benefits (DI); (iii) labor earnings (excluding any benefits and transfers from government);[23] (iv) an indicator variable for having positive earnings; (v) labor

[20] Estimates are robust to alternative choices of window lengths and thresholds for the minimum number of referrals used to construct the instrument; see Figure 5 and Table 7.

[21] We additionally exclude patients for whom a reliable instrument could not be constructed. This means that patients referred in January 2010 are excluded from the sample, as the instrument, which is constructed using a thirty day window immediately preceding referral date, is not well defined for this group.

[22] To construct the health care utilization measure, we apply the nationally set DRG-specific weights for all hospital stays (see `https://helsedirektoratet.no/finansieringsordninger/innsatsstyrt-finansiering-isf-og-drg-systemet/innsatsstyrt-finansiering-isf`). For visits to GPs or specialists outside of the hospital, we sum overall fee-for-service reimbursement rates using the prices set nationally, following 'Fastlegetariffen': `http://normaltariffen.legeforeningen.no/pdf/Fastlegetariff_2016.pdf`.

[23] In contrast to health-related absence and DI, which can be measured by exact dates, labor earnings is measured by calendar year. In this case, year 0 refer to the calendar year of treatment.

earnings given positive earnings; and (vi) total benefits transfers over the five years. We apply the inverse hyperbolic sine transformation to the earnings measure – this approximates the natural logarithm and allows us to retain the zeros. In the same way as the logarithmic transformation, estimates can be approximately interpreted as (semi-)elasticities (Bellemare and Wichman, 2018).

Table 1. Descriptive Statistics

	Mean	SD
Age	46.3	(9.40)
Female	0.47	
Foreign-born	0.09	
Partner	0.56	
Education		
- Primary	0.31	
- High school graduates	0.37	
- College	0.31	
Manual job	0.25	
Office job	0.42	
Sick leave on referral date	0.24	
Wait time	190.3	(184.3)
Congestion	176.6	(52.3)
Hospitals	27	
Queues	104	
Observations	26,410	

Notes: Descriptive statistics of estimation sample. Wait time and congestion are measured in days. Manual job is a dummy for occupation codes starting with 6 (skilled agricultural, forestry and fishery workers), 7 (craft and related trades workers), 8 (plant and machine operators and assemblers) or 9 (elementary occupations), while office job is a dummy for occupation codes starting with 1 (managers), 2 (professionals), 3 (technicians and associate professionals) or 4 (clerical support workers). Primary education is a dummy for education codes (NUS) $<$ 4 or 9; high school graduates has NUS codes 4-5, while college educated patients have NUS codes 6-8. Queues give the number of hospital by procedure groups.

On average, patients experience substantial wait times between referral and surgery: From Table 1 we see that the mean wait time is 190 days with a standard deviation of 184 days. The distribution of this variable is depicted in Figure A2. The sample is fairly representative with respect to gender (slightly more men than women) and education: the share of patients with primary education, high school graduation and college education is about one third for all groups. About one fourth of the patients are on sick leave on referral date. Figure 1 illustrates average absence rates and hospital days relative to referral. Note that absence rates grow from year -2 to year -1; this likely reflects a combination of absence due to orthopedic conditions as well as mean reversion stemming from the fact that we exclude patients with high absence rates in year -2. Both variables exhibit a spike in the year of referral. As the majority of patients wait less than a full year, this also captures any hospital stays and work absences directly related to surgery and recovery. In years 2-5, hospital use and absence rates both fall, though absence rates in particular appear to stabilize at a slightly higher level compared to pre surgery.

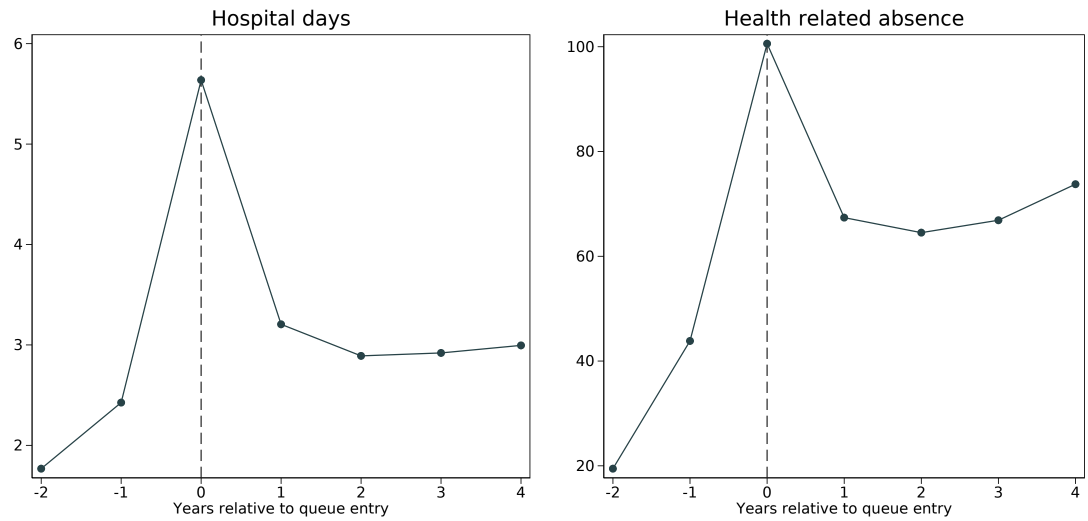


Figure 1. Hospital days and health-related absence, before and after referral.
Notes: Figure plots average hospital days and health-related work absence days for patients in the estimation sample. Time measured relative to referral: year 0 (vertical line) is the year starting with and including the day of referral.

4 Identifying the effects of wait time for hospital treatment

Waiting for hospital treatment may affect health outcomes, as well as the incidence and duration of sickness leaves. Identifying a causal effect of wait time for hospital treatment on health outcomes and labor market attachment is challenging, as wait time is presumably correlated with unobservable individual characteristics, such as health and propensity to work, which affect both health outcomes and labor supply.

This means, that a regression of sick leave duration or health on wait time provides an unbiased estimate of the causal effect of wait time only under the assumption that variation in wait time is (conditionally) uncorrelated with unobservable determinants of the outcome.

There are several reasons why the exogeneity assumption is unlikely to hold. First, patients with the greatest need are given priority in the allocation of treatment slots. As a result, healthier patients typically have longer wait times than patients with a more urgent need for medical care. While the prioritization mechanism ensures that healthy people are subject to longer wait times, healthy people are also less likely to have long absence spells, possibly biasing our estimate of wait time. Moreover, after the Faster Return to Work reform was passed in 2007, hospitals are allowed to give priority to patients who are on sick leave or at high risk of entering sick leave. This scheme could also lead to an association between short wait times and a high incidence of absence from work.

Finally, observed wait time may to some extent be determined by individual behaviors that are correlated with health outcomes. For example, patients with a better knowledge of the health care system may be able to queue-jump. If these individuals are more likely to have a fast recovery and lower sick leave duration independent of wait time, estimates could be biased towards finding negative effects from longer wait times.

To summarize, OLS estimates are likely contaminated by omitted variable bias, though the direction of that bias is not clear. To address concerns of omitted variable bias and endogeneity, we therefore instrument for patient wait time with a constructed measure for the congestion facing each patient.

4.1 Instrument: Congestion - average wait time

In our empirical strategy we exploit variation in wait times that arises because the degree of system congestion fluctuates over time. As a result, otherwise similar patients have different expected wait times depending on the date they enter the queue for a particular treatment type at a particular hospital. To exploit this source of quasi-random variation, we construct a measure of the "congestion" facing each patient, defined as the average wait time of patients queued for the same procedure at the same hospital in a window of time just preceding the focal patient's entry onto the queue. This congestion measure is then used to instrument for patient wait time in a traditional instrument variable estimation framework.

Our measure of congestion is constructed using a sample of all patients who undergo non-emergency orthopedic procedures at Norwegian hospitals over the relevant period of time. In our baseline specification, the congestion facing patient i is calculated as the average observed wait times of all other patients treated at the same hospital in the same procedure group whose referral dates fall within the thirty-day window immediately preceding the focal patient's referral date.

In order for the identification strategy to be valid, the independence assumption must hold, meaning congestion should be as good as random within hospital-by-procedure groupings. That is, it should be uncorrelated with patients' observed and unobserved pre-referral characteristics. If this assumption holds, reduced form models linking individual outcomes to the instrument will estimate causal effects of congestion.

Institutional factors suggest that this assumption is likely to hold in our setting. Referral to specialist health care is based on a medical evaluation, leaving little scope for patients to strategically time referrals to periods when wait times are shorter. Moreover, as there are no direct costs for being on the wait list, there is no incentive for patients or primary care providers to delay referral once the decision has been made that a surgical procedure is the best treatment choice. As the instrument is constructed using only the wait times of other patients, congestion is not determined by i's own underlying health, priority status, or previous labor market attachment. While hospitals with long wait times may be different from hospitals with shorter wait times, our regression model controls for time-invariant hospital characteristics by including hospital-by-procedure fixed effects. Year-by-month fixed effects are also included in our primary specification to control for seasonality and general time effects.

However, the independence assumption may be violated if some patients respond to long average wait times by seeking treatment at private hospitals operating outside the public health care system. While a large majority of orthopedic procedures are performed in public hospitals or private hospitals contracting with the government, there is a small and growing market for privately funded hospitals that perform certain surgical operations. The costs of these procedures are not reimbursed by the government, but are paid for by the patients themselves or through individual or employer-sponsored private health insurance. Thus high income patients might opt out of public health care when wait times are long, resulting in a negative correlation between socioeconomic status and congestion.

Similarly, patients with less serious ailments may exit the queue if they spontaneously get better during the period after referral, before surgery. Relatively healthy patients who are randomly assigned long wait times may thus be more likely to exit the queue, potentially leading to a negative correlation between health and observed wait times even if wait times were randomly assigned (given the restriction that we only observe wait times for patients who eventually undergo surgery).

Whether or not high socioeconomic status patients choose private health care options when wait times are long cannot be tested directly, as privately funded procedures are not included in the patient register data. Moreover, we lack data on referrals that do not result in surgery. However, the dataset does include a large set of observable characteristics that are correlated with health and labor market outcomes, including age, education and previous earnings, as well as proxies for pre-referral health status such as visits to GP and hospital, and time spent on sick leave in the years prior to referral.

Table 2 shows estimates from OLS regressions of wait time (column 1) and congestion (column 2) on a vector of patient-level covariates capturing predetermined demographic, work and health-related characteristics. These models also control for listing time and fixed effects for hospital-by-procedure group. The first column documents that these characteristics are strongly predictive of patients' wait time. Recalling the discussion on threats to identification, our fears that individual wait times are correlated with unobserved determinants of health appear to be justified. In particular, patients who are not Norwegian-born and those with higher education tend to experience lower wait times, consistent with a scenario in which better knowledge of the health care system facilitates some degree of "queue jumping". Being on sick leave is also associated with significantly shorter wait times, consistent with the health care system giving priority to patients with more serious health problems. Though the significant associations between background variables and individual wait time are interesting per se, they pose no threat to our identification strategy unless the same characteristics are also associated with the instrument.

Importantly, these same characteristics are generally not correlated with our congestion instrument. Only one covariate is found to be a significant predictor of congestion; being married or living with a domestic partner predicts a slightly higher value for the instrument. The size of this relationship is, however, economically marginal and amounts to only 0.5% of the instrument mean. Given the number of covariates being tested, the risk of obtaining one marginally significant variable by pure chance is high, so we do not find this result particularly troubling. As the bottom of Table 2 shows, we find strong evidence that predetermined patient characteristics are jointly correlated with actual patient wait time (p-value <0.001) but not jointly correlated with the congestion instrument (p-value of 0.575). We finally note that while the lack of correlation between observable characteristics and our instrument is reassuring for our identification strategy, we should be concerned that other unobserved differences bias our results. However, following existing literature it is natural to assume that the selection on observables is informative about the selection on unobservables (see, e.g. Altonji et al., 2005). Hence, we interpret the results as a strong argument in favor of the independence assumption.

To further examine the exogeneity of our congestion instrument, we use the characteristics in table 2 to calculate a composite measure of predicted health and labor market outcomes. Specifically, we estimate the following regression:

$$y_i^{0-4} = x_i\beta + \varepsilon_i \tag{1}$$

Table 2. Instrument Validity

	(1) Wait time		(2) Congestion	
Age	-0.035	(0.170)	-0.027	(0.020)
Female	5.934*	(3.134)	-0.257	(0.343)
Foreign-born	20.935***	(5.229)	0.219	(0.491)
Partner	-2.599	(2.218)	0.893**	(0.381)
Education (ref.: college)				
- Primary education	6.601**	(3.144)	0.518	(0.519)
- High school graduates	3.068	(3.606)	-0.259	(0.429)
Office job	3.615	(2.834)	0.120	(0.419)
IHS* earnings t-2	-0.426	(1.434)	-0.077	(0.244)
IHS* earnings t-1	-1.464*	(0.782)	0.110	(0.122)
Sick leave on referral date	-45.247***	(3.952)	0.012	(0.429)
Permanent DI t-1	0.311	(0.365)	0.052	(0.042)
Health-related absence t-2	0.024	(0.033)	0.004	(0.005)
Health-related absence t-1	0.073***	(0.020)	-0.001	(0.003)
GP visits t-2	0.584**	(0.277)	-0.025	(0.040)
GP visits t-1	-0.050	(0.227)	0.041	(0.038)
Hospital days t-2	0.075	(0.282)	-0.040	(0.045)
Hospital days t-1	-0.920***	(0.238)	0.020	(0.038)
Observations	26,410		26,410	
Dep. mean	190.26		176.55	
Joint F-statistics [p-value]	14.94	[0.000]	0.88	[0.594]

Notes: Table shows estimates of wait time (column 1) and congestion (column 2) on observable patient characteristics measured prior to referral. Age, sex, nationality, partner, education and occupation are measured one year prior to referral. Earnings, absence, GP visits and hospital days are measured both one (t-1) and two (t-2) years prior to referral. We use the inverse hyperbolic sine (IHS) of earnings. Both models include fixed effects for year by referral month and hospital by procedure. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

where x_i is a vector of predetermined individual characteristics: age (dummy coded) and all the variables of table 2. This model is estimated on the full sample of workers, then, using the estimated $\hat{\beta}$, we construct predicted values for each outcome, $\hat{y}_i^{0-4}$.

We can think of $\hat{y}_i^{0-4}$ as a proxy for underlying health status - it reflects known predictors of health and absence such as age, gender and previous absence rates. These are characteristics that are at least partially observable by health care providers. If sicker patients are given priority, we should expect a negative relationship between wait times and predicted absence rates, and a positive relationship between wait times and predicted healthcare utilization. However, if congestion is independent of individual characteristics, we should find no correlation between $\hat{y}_i^{0-4}$ and the congestion instrument. Table 3 presents bivariate regressions of wait time and congestion on each of these predicted outcomes. The predicted outcomes are strongly correlated with individual wait times: patients with higher predicted health care utilization, absence rates, and DI entry likelihood tend to experience shorter wait times, while patients with higher predicted future earnings and employment wait longer on average. These correlations are consistent with a prioritization scheme where more needy patients are assigned shorter wait times. Meanwhile, there are no significant correlations between these predicted outcomes and our congestion measure.

To analyze the relationship between actual wait time and absence propensity, we calculate the ventiles of the distribution of actual wait time for each procedure-hospital group. This yields a rank from 1 to 20 indicating the relative wait time conditional on procedure and hospital. Next, for each of these bins, we calculate the average predicted hospital days, absence rates and DI receipt over the 5 years following referral.[24]

The results of this exercise are shown in Figure 2. As before, the panels on the left show a negative association between predicted hospital days, absence rates and DI receipt and individual wait times. Meanwhile, as indicated by the panels on the right, there is no such association between predicted outcomes and the instrument. This lack of correlation further supports our assertion that the instrument is conditionally random.

4.2 Instrumental variable model

Our empirical model can be described by the following two-equation system:

[24] In this calculation, we pool all hospitals and procedures, as by construction, each bin will have (approximately) the same composition of hospital-by-procedure groups.

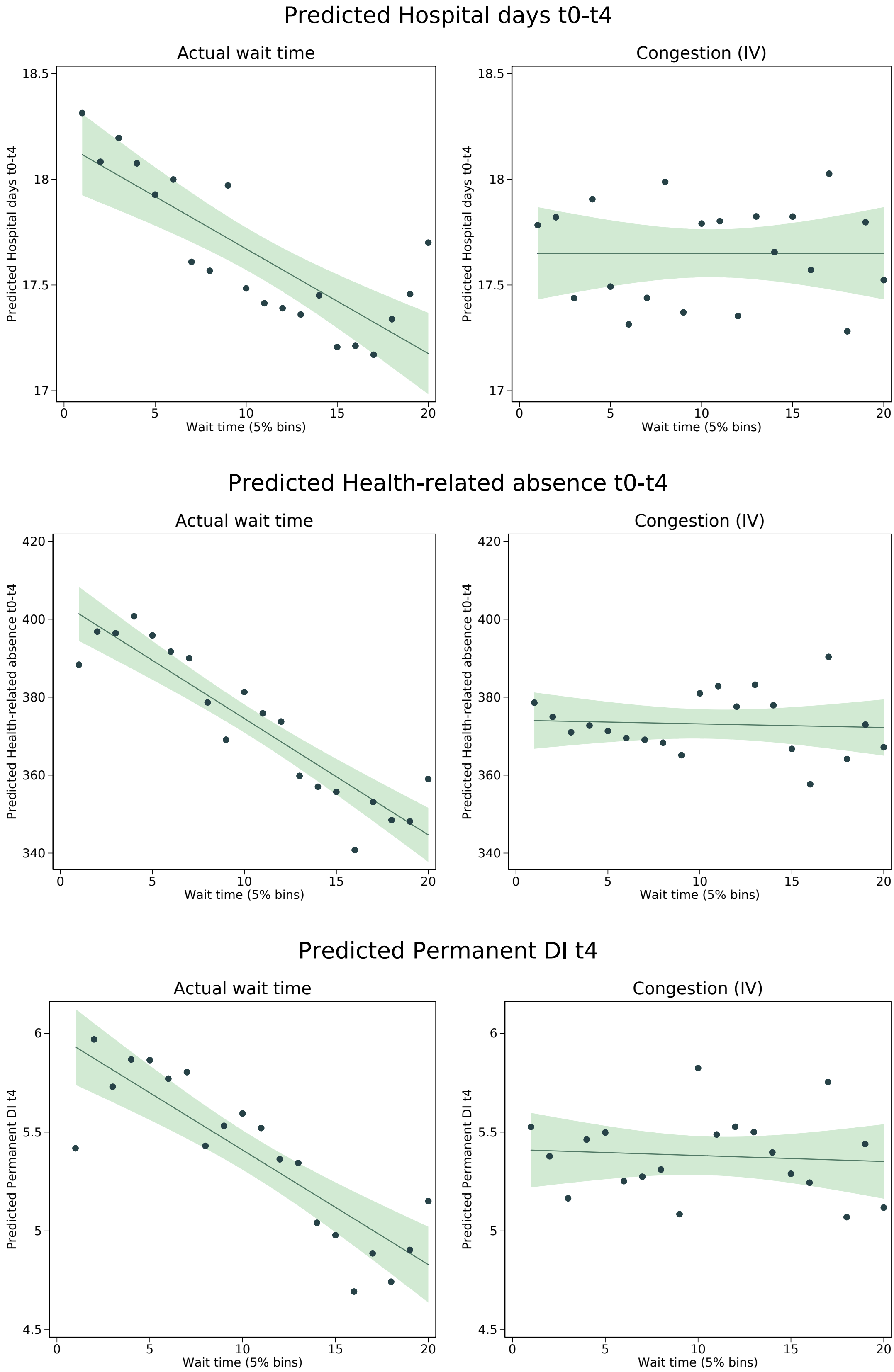


Figure 2. Predicted absence rate, by actual wait time and congestion.

Note: Figure plots average rates of predicted hospital days, absence days and permanent DI receipt (calculated using the covariates in Table 2) against the ventiles of the distribution of actual wait times (left) and congestion (right), calculated separately by hospital-procedure group.

Table 3. Bivariate regressions of wait time and congestion on predicted outcomes

Predicted outcomes	(1) Wait time		(2) Congestion	
Health-related absence t0-t4	-0.024***	(0.004)	0.000	(0.000)
Permanent DI t4	-0.518***	(0.141)	0.015	(0.016)
IHS earnings t4	2.763**	(1.320)	0.003	(0.150)
Positive earnings t4	39.405***	(13.875)	0.198	(1.612)
IHS earnings t4 given >0	1.155	(2.691)	-0.030	(0.305)
GP visits t0-t4	-0.008	(0.053)	0.005	(0.007)
Hospital days t0-t4	-0.391**	(0.159)	0.009	(0.018)
Hospital utilization t0-t4	-0.083***	(0.027)	0.002	(0.003)
Resurgery t0-t4	-90.530***	(15.486)	0.393	(2.113)
Emergency admission t0-t4	-4.482*	(2.324)	0.160	(0.277)
Mortality t4	-0.455***	(0.153)	-0.012	(0.018)
Observations	26,410		26,410	
Dep. mean	190.26		176.55	

Notes: The table shows estimates from bivariate regressions of wait time and congestion on various predicted outcomes. Predicted outcomes are calculated using the covariates in Table 2. We use the inverse hyperbolic sine (IHS) of earnings. t0 refers to the first 365 days starting with the date of referral; t4 is the 5th year (day 365*4 to day 365*5) relative to referral, while t0-t4 is the full period from referral until and including the fifth year. All regressions include fixed effects for year-by-month and for hospital-by-procedure. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

$$WT_{iht} = \alpha Congestion_{ht(i)} + \lambda_h + \theta_t + \varepsilon_{iht}, \tag{2}$$

$$Y_{iht+s} = \delta WT_i + \pi_h + \tau_t + \nu_{iht}, \tag{3}$$

where $Congestion_{ht(i)}$ denotes the instrument, i.e. average wait days at hospital-by-procedure group h in the 30-day window preceding patient i's referral date at year-by-month t). The parameters λ_h and π_h are hospital-by-procedure fixed effects, while θ_t and τ_t are year-by-month fixed effects. These control for any time invariant differences across hospitals and/or procedure groups in the quality of care or health of patients. In the second stage equation 3, Y_{iht+s} is a dependent variable of interest that is measured for patient i at some point $t + s$ after entering the queue (e.g.

health-related absence five years after the referral date).

The coefficient of interest, δ, represents the effect of wait time for hospital treatment on the outcome variable. While the independence assumption is sufficient for a causal interpretation of reduced form estimates of effects of the instrument (congestion) on wait time, additional assumptions are required for our IV model to produce a causal effect of δ.

In addition to the assumption of instrument independence, several other conditions must be met for 2SLS to produce estimates of δ that reflect the causal effect of wait time. Of critical importance, the instrument must be relevant; that is, our proxy for system congestion at the patient's time of listing should be predictive of the actual time that patients wait for treatment. As discussed in the introduction, our paper's focus on orthopedic surgery implies that this assumption is likely to hold. Orthopedic conditions are rarely life threatening, leaving hospitals with considerable discretion in delaying surgery when excess demand is high. Regardless, the relevance assumption can be tested directly by examining the first stage estimation results.

Second, the instrument must affect the outcome only through its effect on individual wait time. This exclusion restriction would be violated if, say, health outcomes were worsened through lower quality caused by congestion in the hospital unit. The exclusion restriction cannot be tested directly, but we can examine whether there are signs of correlation between congestion and the volume of orthopedic procedures. Moreover, patients who are admitted for immediate surgery (emergency admissions) may provide a useful control group, as they are treated by the same medical teams without being subject to a waiting period. If the exclusion restriction holds, then congestion should have no effect on outcomes for this group. We will return to this test in the robustness section.

Interpreting the magnitude of our 2SLS estimates is complicated if (i) wait time effects are heterogeneous across different patients in our sample, and (ii) the effects of congestion on wait times are heterogeneous.[25] For instance, if our sample con-

[25]Interpreting IV estimates when treatment effects are heterogeneous has been a matter of substantial

sists of two types of patients, some whose wait times are affected (in some constant amount) by congestion and others whose wait times are unaffected by congestion, 2SLS estimates of δ reflect a weighted average of the treatment effects pertaining to the former group, with greater weight placed on those patients facing greater deviations in congestion at listing. The literature generally uses the term "local average treatment effect" (LATE) to convey this interpretation. However, as Heckman et al. (2006) demonstrate, the LATE interpretation of IV estimates is potentially undermined when *essential heterogeneity* is present – that is, when wait time effects vary, and the responsiveness of individual wait times to congestion covaries with the size of the wait time effects. In this case, the usual LATE interpretation of δ is only maintained if congestion exerts monotonic effects on wait times.[26] The monotonicity assumption would be violated if there exists some subset of our patients for whom lower (higher) levels of congestion predict longer (shorter) wait times. While we consider this unlikely, we cannot fully rule out this possibility. For instance, monotonicity could conceivably be violated if some subset of patients, when faced with greater congestion, engage more successful efforts to "jump the queue." Alternatively, workers who are less eager to return to work might be more inclined to request a delay if wait times are short.

As Fiorini and Stevens (2014) discuss, the dual assumptions of independence and monotonicity have a number of testable implications: the estimated first stages should be positive across subgroups in our data; wait times should be monotonically increasing in the value of the instrument; and the distribution of wait times for patients with low congestion values should stochastically dominate the distribution of wait times for patients with congestion values. We return to this in our discussion of results.

econometric interest, with seminal contributions by Imbens and Angrist (1994); Angrist and Imbens (1995); Angrist et al. (1996); Heckman et al. (2006).

[26] Recent work by De Chaisemartin (2017) demonstrates a modified LATE interpretation still holds under violations of monotonicity provided there are more "compliers" than "defiers" in each strata of the wait time effect distribution. The LATE identifed by 2SLS in this case is specific to the "excess compliers" that exist in each strata.

5 Results

5.1 Graphical Evidence

We begin our presentation of results by providing a graphical representation of the IV approach in Figure 3. All panels draw a histogram showing the distribution of congestion in our sample. Specifically, congestion is included as the residual from a regression of average wait time on fixed effects for hospital-by-procedure and year-by-month, then rescaled to the mean.[27]

Panel (a) illustrates the relationship between congestion and individual wait times, corresponding to the first stage equation (2). The graph plots a local linear regression of individual wait time against congestion. Individual wait time is monotonically increasing in congestion, and is close to linear. This provides some evidence that the monotonicity assumption may be satisfied.

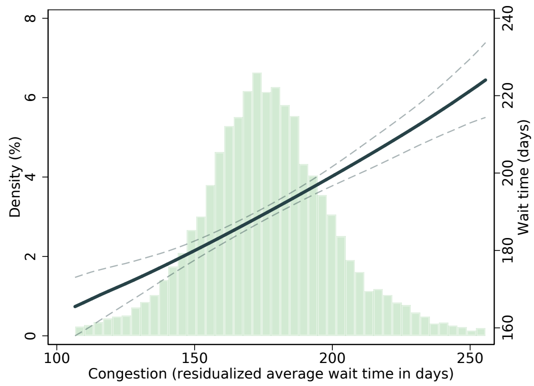
Panels (b), (c), and (d) plot the reduced form effect of congestion on hospital utilization, absence, and DI receipt. The figure shows no evidence of any effects of congestion on hospital utilization: the local linear regression is largely flat over most of the congestion distribution. Absence and 5-year permanent DI receipt on the other hand, is increasing in congestion. Figure 3 thus gives a first indication that wait time increases absence rates, but not health care utilization.

5.2 Main Regression Estimates

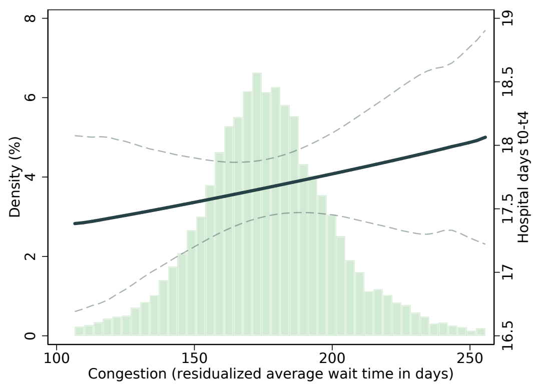
This section presents the estimated effects of wait time on health outcomes and labor market attachment. First, we present our baseline IV estimates on health and labor market outcomes during the five year period following referral for treatment. Next, extended models are estimated to shed further light on the underlying mechanisms.

The first set of models estimates the effects on health outcomes and health care utilization. Table 4 presents the results of estimating Equations (2) and (3); the corresponding OLS estimates are included for reference. All models shown in this table include dummies for hospital-by-procedure and year-by-month.

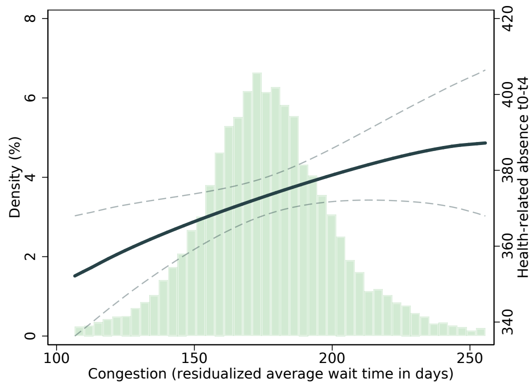
[27] Figure A2 depicts both the instrument (i.e the residual from a regression of average wait time on fixed effects hospital-by-procedure and year-by-month effects); and the raw average wait time.



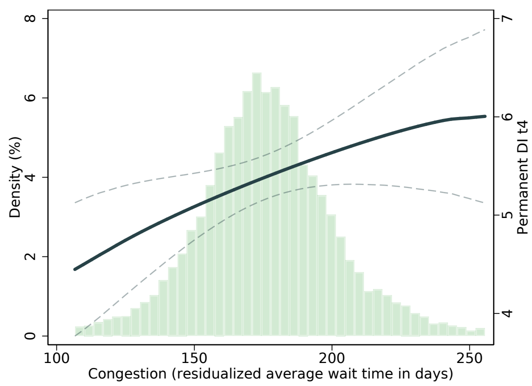
(a) First stage



(b) Hospital days - reduced form



(c) Health-related absence - reduced form



(d) DI receipt year 5 - reduced form

Figure 3. Effect of congestion on individual wait time (first stage) and selected health and labor market outcomes

Notes: Panel (a) illustrates the first stage. The solid line is a local linear regression of residualized individual wait time on congestion. Panels (b), (c) and (d) illustrate the reduced form relationships for hospital days (b) and absence days (c) over the five year period after referral, as well as permanent DI receipt year 5 after referral (panel d). In both figures, congestion is included as the residual from a regression of average wait time on hospital-by-procedure and year-by-month fixed effects. A histogram of congestion is shown in the background of all figures (top and bottom 1% excluded from the graph). Dashed lines represent 95% CI.

OLS estimates (panel A) indicate some statistically significant correlations between wait time and health outcomes: a longer wait time is positively correlated with the number of GP and hospital visits, but negatively correlated with repeat procedures for the same condition and with 5-year mortality. However, point estimates are small: 100 days longer wait time is associated with 0.3 additional primary care visits over the 5 year period. Moreover, interpreting these correlations is complicated by the likely non-random nature of individual wait time.

Column (1) indicates that our first stage is positive and strongly significant (F-value 40.0). A one day increase in congestion predicts an additional 0.36 days of patient wait time. The reduced form estimates indicate that congestion has no significant effects on health outcomes. Similarly, IV estimates are all small and nonsignificant. These estimates are fairly precise, and we are able to rule out any appreciable effects of wait times on health outcomes.

Next, table 5 presents the effects of wait time on labor market outcomes: total absence days over the five years since referral, and disability insurance receipt and earnings in year 5. OLS results shown in Panel A show no significant associations between wait time and total absence or DI receipt. Similarly, there is no significant correlation between wait time and earnings, though there is a small, marginally significant negative correlation between wait time and earnings when we condition on positive earnings in year 5. These estimates are likely to reflect a combination of selection effects, as well as any causal effects of wait time. In particular, an intentional policy of prioritizing patients in need of immediate treatment is likely to introduce a negative selection bias, meaning that the estimated effects on absence and DI receipt would be biased downward.

Panels B and C show the corresponding reduced form and IV estimates. The reduced form models indicate that patients who are referred to surgery in periods with high average wait times experience significantly higher absence rates in the following years, as well as a higher probability of receiving permanent disability benefits five years on. The estimated effects on health-related absence and disability are not only highly statistically significant, but also economically meaningful. Scaling these estimates by the first stage, the IV models find that each additional day spent waiting for surgery increases total health-related absence over the five year period by 0.87 days. Our model of long term disability implies that ten additional days of wait time increases the likelihood of a patient receiving DI by 0.4 percentage points. When the fiscal spillovers of longer wait times are considered, any effects on DI are particularly interesting as DI tends to be a more permanent state, with low rates of recipients returning to work. This latter finding is important as it indicates that

Table 4. Effects of wait time on health outcomes

	(1) Wait time	(2) GP visits	(3) Hospital days	(4) Hospital costs	(5) Resurgery	(6) Emergency admissions	(7) Mortality
Panel A: OLS							
Wait time		0.00319**	0.00357***	0.00759	-0.000109***	-0.000110	-0.00635**
		(0.00139)	(0.00112)	(0.00739)	(0.0000353)	(0.000141)	(0.00252)
Panel B: Reduced form							
Congestion		0.00734	0.00618	0.0235	-0.0000240	0.000173	-0.000876
		(0.00677)	(0.00433)	(0.0286)	(0.000157)	(0.000663)	(0.0170)
Panel C: IV estimates							
Congestion (FS)	0.356***						
	(0.0563)						
Wait time		0.0206	0.0174	0.0660	-0.0000674	0.000486	-0.00246
		(0.0186)	(0.0119)	(0.0801)	(0.000439)	(0.00185)	(0.0474)
Observations	26,410	26,410	26,410	26,410	26,410	26,410	26,410
Dep. mean	190	39.345	17.650	119.006	0.329	1.330	6.778
FS F-stat	40.0						

Notes: The table shows the estimated effects of wait time on health outcomes. All health outcomes except mortality are measured as the total over the five years following referral. Mortality is measured as death within five years of referral. All regressions include year-by-month and hospital-by-procedure fixed effects. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5. Effects of wait time on labor market outcomes

	(1) Wait time	(2) Health-related absence	(3) DI	(4) Earnings	(5) Earnings > 0	(6) Earnings if > 0
Panel A: OLS						
Wait time		0.0211	-0.000915	-0.000171**	-0.0000174*	-0.0000351
		(0.0209)	(0.000710)	(0.0000781)	(0.00000971)	(0.0000326)
Panel B: Reduced form						
Congestion		0.308***	0.0146***	-0.000935*	-0.0000675	-0.000412**
		(0.112)	(0.00522)	(0.000525)	(0.0000752)	(0.000205)
Panel C: IV estimates						
Congestion (FS)	0.356***					
	(0.0563)					
Wait time		0.866**	0.0409**	-0.00263*	-0.000190	-0.00103**
		(0.353)	(0.0163)	(0.00153)	(0.000215)	(0.000504)
Observations	26,410	26,410	26,410	26,410	26,410	22,799
Dep. mean	190	373.127	5.381	6.124	0.919	6.690
FS F-stat	40.0					

Notes: The table shows the estimated effects of wait time on labor market outcomes. Absence is the total number of health-related absence days (sick pay, temporary and permanent DI) in the five years following referral. DI is an indicator equal to one for patients enrolled in permanent disability insurance five years after referral. Earnings is the inverse hyperbolic sine of earnings measured five years after referral. All regressions include year-by-month and hospital-by-procedure fixed effects. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

the estimated effect on absence is not purely transient but may reflect a permanent withdrawal from the labor market.

Columns (3) - (5) present the effects on earnings. Longer wait times lead to significant earnings losses. The estimated effect size in column (3) implies that 10 additional days of wait time reduce earnings five years on by approximately 2.6%; this effect is imprecisely estimated and only marginally significant. The estimated effect on the probability of having any earnings at all is negative, but not statistically significantly different from zero. We do find significant reductions in earnings conditional on employment: a 10 days longer wait time reduce earnings by 1% in this sample, just under half the effect of the full sample.[28]

Overall, the estimated models of health and labor market outcomes indicate that while longer wait times have no lasting effects on health outcomes, labor supply is significantly reduced in the long run. To further examine this, we have estimated a set of IV models of hospital days, health-related absence and permanent DI by years since referral. For all three outcomes, we run seven separate regressions for each year in the $[-2, 4]$ window around the referral date. The estimates for years -2 and -1 serve as a falsification test: assigned wait time should not have any effects on health and labor market outcomes in the years leading up to referral. As a consequence, the estimated coefficients for these years should be close to zero if our identification strategy holds.

Figure 4 plots the estimated coefficients for hospital utilization, days of health-related absence and permanent DI receipt. For the years leading up referral, the estimated coefficients are indeed close to zero, which is reassuring. For hospital days, longer wait times lead to fewer hospital days in the first year (year 0), followed by an increase in hospital utilization the second year. This pattern is consistent with longer wait times shifting the timing of surgery while leaving total health care utilization unchanged. Importantly, results are only marginally significant, and any effects on

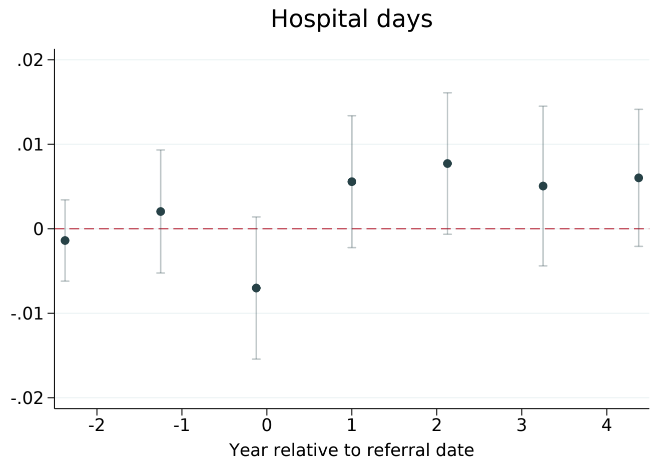
[28]The lack of significant effects on the probability of a patient having positive earnings may at first seem inconsistent with the increase in permanent DI receipt. However, about 50% of the population that receive DI at year 5 also have some labor income, though these are typically small amounts compared to people who do not receive DI. The share of these DI recipients with positive earnings falls over time: in year 6, the share with positive labor income is down to x%.

utilization appear to be transient: by years 3 and 4 after referral, effects are small and not statistically significantly different from zero.

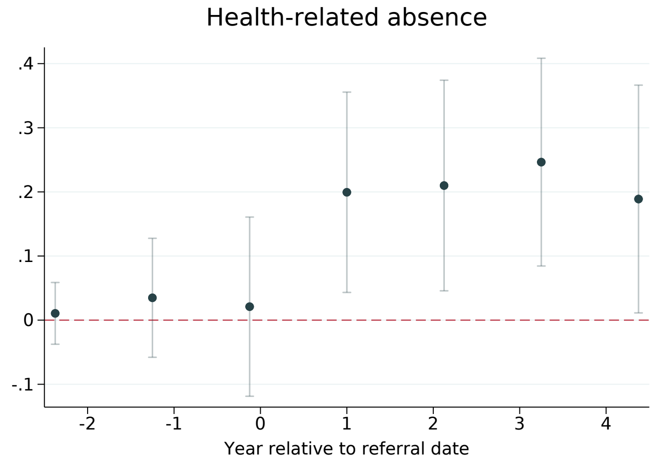
For health-related absence, the pattern is different. On the one hand, longer wait times are expected to increase absence in the first year if patients are unable to work while waiting for surgery. On the other hand, as longer wait times shift the timing of surgery, we should expect a corresponding shifting in absence directly related to surgery and recovery from year 0 to year 1. In the referral year, the models find zero effects of wait times on absence, suggesting that these two effects approximately cancel each other out. For later years, however, the model estimates persistent positive effects: Ten days additional wait time leads to 2-2.5 days additional absence days in each of these years. Crucially, unlike the estimated effects on hospital utilization, the effects on absence do not fade out over time: they are roughly constant between year 1 and year 4, and the estimated effect is still statistically significant in year 4. In order to show how these patterns relate to the aggregate outcomes reported in tables 4 and 5, Appendix figure A3 plots the estimated effects of the running sum of these outcomes over the first 5 years.

Panel C shows the effects on permanent DI - note that for this outcome, we only report one pre-referral year, as inclusion in our estimation sample is conditional on the patient not receiving DI in year two before referral. As expected, effects on DI receipt are small and nonsignificant in the first years after referral, then the estimates start increasing in the third year and become significantly different from 0 in year 4. This is consistent with the institutional setting where permanent DI receipt requires a thorough evaluation period which typically takes years to complete.

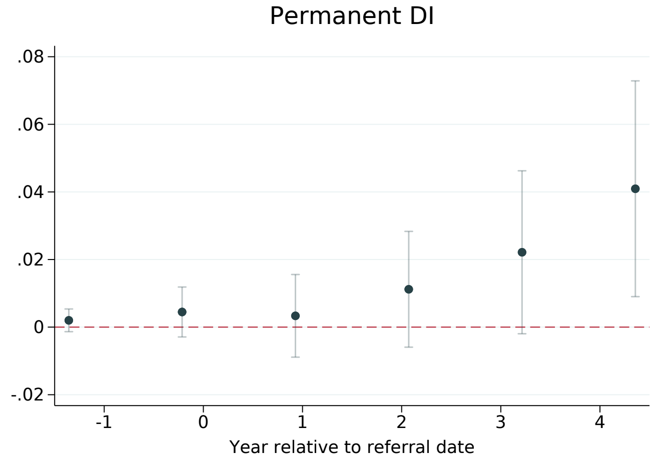
In summary, the results from Table 4 and 5 combined with the time line in Figure 4 reveal that the observed effect of wait time on labor market outcome is not explained through a deterioration in health. It could, in fact, support our hypothesis, outlined in Section 1, that prolonged sick leave due to longer wait times could alter individuals preferences with respect to work/absenteeism or human capital accumulation.



(a) Hospital days



(b) Health-related absence



(c) Permanent DI

Figure 4. Effects by years since referral

Note: The figure plots the estimated effects of wait time on the number of hospital days and days of health-related absence relative to the year of referral.

5.3 Extensions

The (local) average results may mask heterogeneous responses. In Table 6 we explore whether the effects of longer wait time on sick leave days in the five years following the referral date differ depending on patient characteristics. Using detailed demographic characteristics, we split the sample along the following dimensions: education, gender, age (younger/older than 45), occupational category (manual vs office workers), and finally we split the sample according to whether the patient was on sick leave on the day they were referred to surgery.

Table 6 summarizes IV estimates of wait time on hospital days, total absence and permanent DI for the ten subsamples. First, we note that the estimated first stage is positive and statistically significant for all groups. Recall that in order for the IV estimation strategy to be valid, the monotonicity assumption must be satisfied. The fact that the first stage is non-negative for all the estimated subgroups is consistent with monotonicity.

For health outcomes, the analysis indicates that our null result holds across subgroups: the models, presented in panel B, show no significant effects on hospital use for any demographic category. If longer wait times adversely affected the health of patients with more serious conditions, we might expect longer wait times to increase utilization rates for patients who were on sick leave on the time of referral. However, the results in table 6 give no indication that these patients increased their utilization as a result of longer wait times; if anything, the estimated effect on utilization is smaller than for patients who were not on sick leave when they were referred for surgery.

The estimated IV models presented in panels C and D reveal substantial heterogeneity in effects of wait time on absence. When the sample is split according to education, the effects are driven exclusively by workers with high school or less. For people with at least some higher education, the estimated effect is not statistically significant. Similarly, when the sample is split by occupational classification, the models find the estimates are much larger for people in manual occupations com-

Table 6. Heterogeneous effects of wait time on health and labor supply

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(11)	(12)
	Education		Gender		Age		Job		On sick leave	
	Low	High	Male	Female	>45	<45	Office	Manual	No	Yes
Panel A: First stage										
Congestion	0.372***	0.319***	0.335***	0.362***	0.289***	0.443***	0.298***	0.459***	0.360***	0.345***
	(0.0633)	(0.0838)	(0.0514)	(0.0868)	(0.0672)	(0.0851)	(0.0889)	(0.118)	(0.0658)	(0.111)
FS F-stat	34.5	14.5	42.6	17.4	18.5	27.1	11.3	15.1	29.9	9.8
Panel B: Hospital days										
Wait time	0.0130	0.0260	0.00147	0.0282	0.00899	0.0170	0.0182	-0.0228	0.0230	0.00575
	(0.0136)	(0.0267)	(0.0202)	(0.0184)	(0.0214)	(0.0149)	(0.0255)	(0.0205)	(0.0147)	(0.0295)
Dep. mean	17.824	17.269	16.298	19.186	19.740	14.917	17.517	16.429	16.511	21.161
Panel C: Health-related absence days										
Wait time	1.151**	0.0427	0.658	1.125*	1.040*	0.580	0.0474	0.766	0.249	2.719**
	(0.454)	(0.457)	(0.441)	(0.586)	(0.592)	(0.420)	(0.413)	(0.559)	(0.257)	(1.063)
Dep. mean	433.503	240.858	324.232	428.649	410.726	323.965	255.733	464.531	244.566	769.504
Panel D: Permanent DI										
Wait time	0.0463**	0.0196	0.0340**	0.0503*	0.0504	0.0231**	0.0130	0.0542*	0.0107	0.135**
	(0.0206)	(0.0185)	(0.0172)	(0.0279)	(0.0336)	(0.0106)	(0.0182)	(0.0323)	(0.0125)	(0.0656)
Dep. mean	6.656	2.585	4.394	6.501	8.346	1.503	3.172	6.975	2.598	13.961
Observations	18,133	8,277	14,043	12,367	14,965	11,445	11,161	6,581	19,942	6,468

Notes: The table shows subsample estimated effects of wait time on health labor market outcomes. Absence is the total number of health-related absence days (sickness absence, temporary and permanent DI) in the five years following referral. DI is an indicator equal to one for patients enrolled in permanent disability insurance five years after referral. Earnings is the inverse hyperbolic sine of earnings measured five years after referral. All regressions include fixed effects for year-by-month and hospital-by-procedure. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

pared to office workers, though both estimates are fairly imprecise. Given our focus on orthopedic procedures, this is an intuitively plausible finding: manual jobs and jobs that require less formal education may require workers to be more physically fit, yielding a stronger effect of increased wait time on health-related absence. The effect of wait time appears to have higher impact on subsequent sick leave for women than for men, and for senior patients (age>45) than for younger patients, however, neither of the differences are statistically significant. A corresponding subsample analysis of disability insurance receipt in year 5 yields similar patterns, with effects larger for less educated workers.

When the model is estimated separately for patients who were on sick leave at the time they were referred and patients who were working, a clear pattern emerges: patients who were not on sick leave have no significant effects on absence or DI. Meanwhile, for patients who were on sick leave there are large reductions in labor supply following longer wait times. In this group, a 10 days longer wait time increases expected health-related absence over the following 5 years by a total of 27 days, and rates of permanent DI receipt increase by 1.4% - a 10% increase relative to the mean. This pattern is particularly striking given that we find no evidence of adverse health effects.

Appendix table A2 shows the effects for each of the five procedures we study. One group of procedures in particular differs from the others: our instrument does not appear to bind for hip and thigh procedures. While we do not know the mechanism behind this difference, it is worth noting that only 1,783 patients in our sample undergo these surgeries. Moreover, those patients have considerably more hospital days on average over the 5 year period compared to other orthopedic patients, perhaps suggesting that these procedures may reflect more complex medical circumstances where residual variation in hospital congestion is less binding in determining individual wait times. For the other four classes of procedure, the first stage is positive and significant, and of similar magnitude, though the F-stat dips below 10 for hand and wrist procedures. Splitting the sample like this, we do lose precision to the extent that none of the IV estimates are significant at conventional levels. The point estimates are largely similar, however.

Our IV estimates are weighted averages over all one-day increments in treatments induced by our instrument (Angrist and Imbens, 1995).We now examine if responses are different at hospitals characterized by different levels of average wait time. To this end, we first calculate the average wait time at each hospital, before splitting the sample into above and below the median (181 days). The results of this exercise, presented in appendix table A3, indicate that effects are driven entirely by hospitals with above-median wait times. In hospitals with average wait times below the median, longer wait times have no significant effects on overall absence rates or five-year rates of DI receipt.

5.4 Robustness

Table 7 presents a set of robustness and specification tests. In our preferred specifications, the instrument is constructed using the average wait times of patients referred to the focal worker's observed hospital. This may be problematic if patients self-select to hospitals on the basis of expected fluctuations in queue lengths. To address this, we have estimated models where the instrument is defined using catchment areas based on individuals' place of residence. Specifically, we define the instrument as average wait times among patients who live in the same catchment area, regardless of the observed hospitals. Results from this exercise are shown in column (1) of Table 7. As expected, this approach reduces the precision of the estimates as we now introduce additional measurement error. The first stage is weaker, though still passes conventional tests for weak instruments with an F-statistic of 15.5 (vs 40 in the baseline models). Somewhat surprisingly, this specification yields a marginally significant increase in the number of hospital visits. Meanwhile, estimated effects on total absence days and DI remain statistically significant at conventional levels, with point estimates somewhat larger than our preferred specification, though the limited precision complicates the interpretation of this difference.

Our preferred specification includes only controls for year-by-month of referral and hospital-by-procedure fixed effects. The models in column (2) and (3) of table 7 assess the stability of the estimates to adding additional covariates. These are measured the year before referral and include, in column (2): week fixed effects, linear, quadratic and cubic terms for age, earnings and indicators for female, married, foreign born and education status (high school dropout, high school graduate, college). In column (3) we use date fixed effects rather than year-by-month or -week. The results from these models are remarkably similar to the estimates from our baseline models. The stability of results across models with and without additional controls supports the claim that patient characteristics are unrelated to the instrument.

Our finding that longer wait times have significant effects on absence rates up to five years after referral suggests that longer wait times do not simply raise absence rates while patients wait for surgery. Rather, there appear to be significant long-term effects lasting beyond the waiting period and the time of initial recovery. However, this interpretation is potentially problematic as the sample contains some very long wait times, lasting longer than two years. These observations seem unlikely to drive our results, nonetheless: as the results in column (4) of table 7 indicate, omitting patients who wait longer than 2 years for surgery yields slightly larger point estimates for both total absence and five year DI, though the difference is not statistically significant.

When constructing the instrument, we implement a number of admittedly arbitrary decisions with respect to window size and sample size. As a robustness check, we redo the analysis with different versions of the instrument: by changing the queue window; changing the queue size; and trimming the instrument of extreme values. In the main estimation sample, we exclude hospital-procedure groups in which the number of patients in the referral window ever dips below 3. Columns (5) and (6) illustrate how our results change when we require a minimum of 5 or 10 peers in any given window. Imposing these additional restrictions weakens the instrument somewhat, possibly reflecting how these models leave less variation in the instrument as additional groups are excluded from the sample. Overall, IV estimates indicate that our key results are largely robust to choice of threshold, though imposing a threshold of 10 reduces the precision of the estimated effect on wait time to the point

Table 7. Robustness

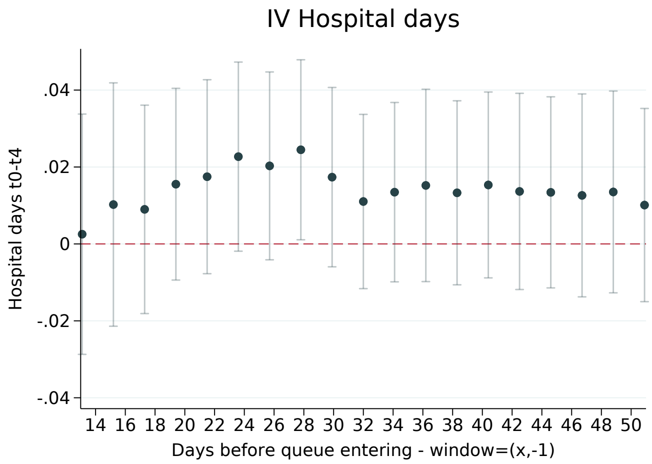
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Time controls		Min N^{peers}	queue sample		
	HRR	Add ctrls	Day	Max wt 2yrs	5 peers+	10 peers+	No LMA restr	Excl. hips
Panel A: First stage								
Congestion	0.263***	0.358***	0.367***	0.293***	0.357***	0.381***	0.339***	0.377***
	(0.0667)	(0.0550)	(0.0539)	(0.0462)	(0.0636)	(0.0854)	(0.0507)	(0.0572)
Dep. mean	190	190	190	171	187	184	197	190
FS F-stat	15.5	42.2	46.4	40.3	31.5	19.9	44.5	43.5
Panel B: Hospital days								
Wait time	0.0419*	0.0171	0.0112	0.0224	0.0217*	0.0194	0.00569	0.0177
	(0.0215)	(0.0123)	(0.0115)	(0.0153)	(0.0127)	(0.0146)	(0.0167)	(0.0115)
Dep. mean	17.650	17.650	17.650	17.575	17.569	17.387	21.659	17.020
Panel C: Health-related absence days								
Wait time	1.318**	0.812**	0.783**	1.167***	0.876**	0.752	0.822**	0.769**
	(0.641)	(0.357)	(0.335)	(0.451)	(0.404)	(0.468)	(0.371)	(0.334)
Dep. mean	373.127	373.127	373.127	373.176	370.412	356.861	626.165	365.446
Panel D: Permanent DI								
Wait time	0.0702**	0.0408**	0.0367**	0.0535**	0.0350**	0.0340*	0.0445**	0.0323**
	(0.0319)	(0.0162)	(0.0154)	(0.0211)	(0.0170)	(0.0196)	(0.0222)	(0.0145)
Dep. mean	5.381	5.381	5.381	5.423	5.249	5.032	20.600	5.141
Observations	26,400	26,410	26,410	25,816	24,768	20,728	39,399	24,627

Notes: All regressions include year-by-month and hospital-by-procedure fixed effects. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

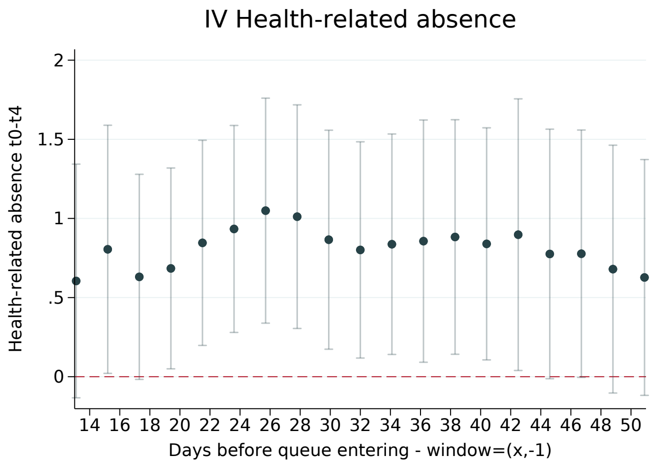
where it is no longer significant at conventional levels. In column (7) we remove the requirements of labor market attachment, and estimate the models in the full sample of patients. Again, the effects are very similar to the baseline model. Finally, in column (8) we exclude the hip replacement procedure from our analyses as our instrument does not appear to bind for hip and thigh procedures (see appendix A2). Here, too, with this sample restriction the estimated effect is very similar to our baseline model. Appendix Table A4 Columns (1) - (3) present models estimated on a sample excluding patients with a history of orthopedic surgery in different windows before referral. We might worry that the identifying assumptions of our model are less likely to hold for these patients – for instance, they, or their referring doctors, might have greater access to information as to which hospitals have shorter queues. However, there appears to be no difference between our baseline estimates and estimates from samples which exclude patients with an orthopedic history.

Our baseline estimations use a time frame of 30 days before the referral date of patient i to estimate patient i's average wait time. Figure 5 illustrates the effects of varying this window, plotting IV estimates of the effects on absence days and disability where the instrument is constructed using pre-referral windows of 14 to 50 days. Overall, results are robust to choice of window, though estimated effects tend to be less significant for very short windows (14 days), possibly reflecting increased noise associated with small sample sizes.

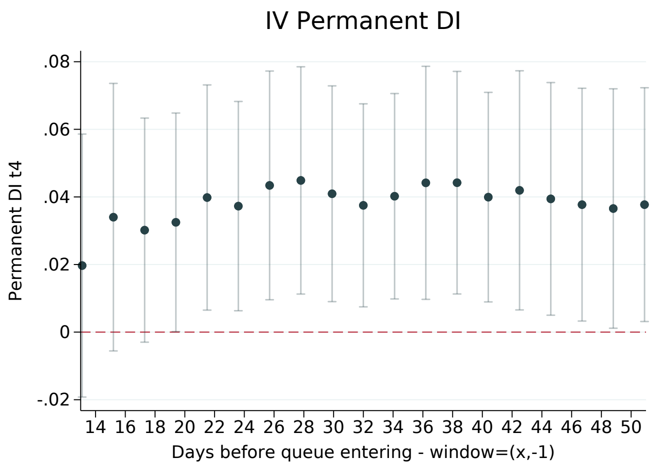
As discussed in section 4, some patterns of scheduling/rescheduling could lead to violations of monotonicity. The local linear regression of wait time on congestion (presented in figure 3) and the non-negative estimated first stage coefficients across subgroups in table 6 give some indications that monotonicity holds. Moreover, excluding patients with delayed procedures from the queue sample yields very similar results (see Appendix table A6). Given independence and monotonicity, the distribution of wait times for patients with high congestion should stochastically dominate the distribution of wait times for patients with low congestion (Angrist and Imbens, 1995; Fiorini and Stevens, 2014). In Appendix figure A7, we have plotted the empirical cumulative distribution functions of wait time for people with high



(a) Hospital days



(b) Health-related absence days



(c) DI year 5

Figure 5. Varying window defining instrument.

Notes: Each point represents coefficients (95% CI) from separate IV estimations of hospital days (panel a), health-related absence (panel b), and DI (panel c) on wait time, with varying length of the pre-referral window used to define the instrument. All regressions include fixed effects for hospital-by-procedure and year-by-month.

(above median/fourth quartile) and low (below median/first quartile) wait times. The CDFs do not cross, providing further support for the monotonicity assumption.

In order for the IV estimation strategy to be valid, the instrument must satisfy the exclusion restriction. The congestion instrument should affect our outcomes only through increased wait times. The exclusion restriction would be violated if, for instance, congestion was correlated with the quality of treatment, as this would open up a second causal channel.

Such violations could occur when hospitals face higher than normal capacity constraints, if this results both in patients waiting longer for surgery (longer wait times for planned procedures) and higher volumes of surgery being performed, possibly reducing the quality of each procedure (if there is a quantity-quality trade-off). To examine this, we construct an auxiliary dataset containing all orthopedic procedures performed during the years 2010-2011. This dataset includes emergency admissions and patients who are referred for several procedures in the same referral period. This sample is used to construct datasets containing average wait times for scheduled patients, as well as counts of the total number of procedures in each time period (week/month). We then estimate a set of models for studying the sickness absence of patients undergoing emergency (unplanned) surgery. These patients have, by definition, not spent time in a queue awaiting treatment. As a consequence, the outcomes for this group can be used to estimate placebo models. Specifically, we estimate regressions of five-year absence and DI on the wait times of scheduled patients, controlling for calendar time and hospital-by-procedure fixed effects. If the exclusion restriction holds, we would expect to find zero effects of congestion for this group. Conversely, a positive relationship between congestion and later sickness absence would indicate that congestion influences outcomes through channels other than individual wait times, which would violate the exclusion restriction.

Results from this exercise are shown in table 8. The model finds no significant congestion effects on absence or DI for patients undergoing unplanned surgeries. This is in line with what we would expect if the exclusion restriction holds. To summarize, we find no evidence that longer wait times have an independent effect

Table 8. Absence, emergency patients

	(1) Hospital days	(2) Health-related absence	(3) DI
Congestion	0.0148	0.0169	0.00440
	(0.00993)	(0.0794)	(0.00482)
Observations	54,951	54,951	54,951
Dep. mean	27.81	118.5	4.606

Note: The table shows models estimated on a sample of patients admitted for emergency orthopedic surgery. In these models, congestion refers to the average wait time of non-emergency patients in the hospital-by-month group. All regressions include year-by-month and hospital-by-procedure fixed effects. Standard errors are clustered at the hospital-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

on treatment quality (e.g. through congestion effects at the hospital).

5.5 Fiscal Implications

Our finding that longer wait times for orthopedic surgery contribute to an increased occurrence of health-related work absence has important fiscal implications through the additional costs placed on social insurance schemes. To investigate this directly, we utilize our IV model to estimate the average causal effect of wait times on the monetary value of transfer payments received by each subject – the sum of any sick leave and disability payments. A detailed description of our approach is presented in Appendix B.

Figure B1 depicts the estimated "transfer effects" each year from the referral date. It is noteworthy that these estimates remain sizable and statistically significant through the fifth year of follow-up. Applying a 3% annual discount rate, the present discounted value (PDV) of additional benefit payments associated with 10 additional days of wait time is estimated to be NOK 6,390 over the five years of observed follow-up time.

This estimate presumably understates the full impact of wait time on transfer payments since our results would lead us to anticipate the transfer effects of wait time to extend beyond the fifth year. The average age in our sample is 46.3, which means the typical subject is 20 years from the standard retirement age (67) in Norway. If we assume the transfer effects of wait time in years 6-20 are the same as those in year 5, the PDV of additional benefit payments associated with 10 additional days of wait time is projected to be NOK 23,390 over 20 years of follow-up. Under a more conservative assumption, that the "transfer effect" decreases annually by 10%, starting in year 6, the estimated PDV of additional benefit payments declines to NOK 14,950. In the exercise below, we will use this more conservative estimate to calculate the total expected reduction in (PDV) transfer payments, on average, from a 10-day reduction in wait time.

Reducing wait times would obviously be costly from the perspective of the Norwegian health system budgets, and predicting the cost of achieving such reductions is beyond the scope of this paper. Nonetheless, these results suggest that even a small intervention to temporarily increase system output (i.e. additional procedures performed) could yield substantial reductions in future transfers.

To demonstrate, consider the impact of an intervention whereby a hospital delivered one additional orthopedic procedure on a specific date, presumably financed through an extra allocation of resources to the hospital. For the sake of simplicity, assume patients are largely treated on a first-come/first-served basis, and that the inflow rate of new patients (into the queue) and workflow rate of the hospital (procedures performed) are not affected by the intervention. Provided the queue is never exhausted at any future point in time, this insertion of an additional procedure into the system would have the effect of permanently reducing the queue length by 1 patient, and would generate a reduction in aggregate wait times equal to one day for each calendar day going forward.[29]

It follows from this that the PDV of the expected transfer savings associated with this intervention can be calculated as NOK 1495 times the infinite sum $\sum_{d=1}^{\infty} \phi^d$, where ϕ is the daily discount factor. Assuming a 3% annual discount rate and applying the power series formula[30], the estimated saving generated by the intervention would be projected to be over NOK 18,467,000.

[29]Suppose the intervention occurred on a Wednesday, causing one patient's surgery to be moved from Thursday to Wednesday, one patient from Friday to Thursday, one patient from Monday to Friday, and so forth. While the Monday-to-Friday patient experiences a three day reduction in wait time, the aggregate reduction in wait time equals "one day" per calendar day.

[30]ϕ takes the value of 0.99991908, so $\sum_{d=1}^{\infty} \phi^d = \frac{\phi}{1-\phi} =$12,356.2.

However, this estimate should be regarded with a fair degree of caution. If shorter queues induce a higher flow of patients into public queues (Martin and Smith, 1999, 2003), or if hospital productivity is endogenous with queue length (e.g. hospital productivity decreases when wait times are shorter), the reduction in queue length resulting from the intervention would be expected to wane over time. Our estimate could also be biased upwards if the transfer effects in years 6 through 20 wane more quickly than we have assumed (10% annually starting in year 6) – though the opposite is also possible - in particular, the effect of wait time on disability insurance could continue to increase beyond our five-year estimation window as more patients complete the DI certification process. In Appendix B, we explore the sensitivity of our estimates with respect to these assumptions. Similarly, if there is a nonzero probability that capacity constraints cease to be binding, the above calculations will overstate the savings induced by an additional surgery. In Appendix B, we discuss how the estimated savings vary with the probability that capacity constraints become non-binding. As a lower bound, we consider the case where capacity constraints immediately cease to bind. This calculation uses the sample average realized wait time rather than the discounted infinite sum of days, assuming that the marginal procedure removes a single patient from the back of the line, fully ignoring spillovers to later entrants. This more conservative assumption still yields substantial savings, totalling more than NOK 250,000 over a 20-year period.

With those caveats in mind, we would nonetheless highlight that the estimated fiscal saving associated with the intervention we described is several orders of magnitude larger than the reimbursements that hospitals receive for treating orthopedic patients. Hospitals in Norway are paid according to a DRG-based payment model, with payment levels intended to approximate the marginal costs hospitals incur in delivering that care, under their current capacity constraints and workflow rates. Over our sampling period, the average costs of an orthopedic procedure was NOK 33,150. Presumably, it would cost more than this to insert an additional procedure into a hospital's existing operations. Generous overtime payments to hospital clinicians and staff might be required, for instance. However, our most conservative estimates - assuming capacity constraints become nonbinding immediately after introducing the marginal surgery - find that the fiscal savings associated with a marginal surgery exceed the average surgery cost by more than 7.5 times.[31]

6 Conclusion

Our paper examined the effect of wait time for orthopedic surgeries on workers' healthcare utilization and labor market outcomes by exploiting the idiosyncratic variation in system congestion that exists when a patient is referred for treatment at a particular hospital. Consistent with the medical literature, we fail to find evidence that longer waits have meaningful effects on patients' future health, with generally nonsignificant associations between wait time and healthcare utilization outcomes, including re-surgery rates. In contrast, longer wait times significantly increase the number of days workers are absent from work for health-related reasons, with 10 additional days of wait time leading to 8.7 more days of health-related absence over the next five years (which includes both temporary sick leave and disability spells). While some of this is due to extended sick leave as patients await surgery, long waits induce higher rates of sickness absence into the fifth post-referral year, well after the surgery and recovery period has ended for the vast majority of our sample. The persistent nature of these labor supply effects is also evidenced by sizable effects of wait time on permanent disability entry by the end of year five, with 10 additional days of wait time causing a 0.4 percentage point increase in disability benefit receipt.

We also uncovered substantial areas of heterogeneity in these effects. The estimated labor supply effects were found to be especially pronounced among patients on sick leave at the time of referral, those with lower education, and those treated in hospitals with higher average wait times. In each case, there is no evidence of larger negative health effects in the subsample exhibiting larger labor supply effects, which further undermines the hypothesis that wait time-induced increases in health-related

[31] Assuming capacity constraints are always binding yields a much larger cost savings ratio of more than 500.

absence were driven by poorer health outcomes.

As discussed in the introduction, the economics literature suggests two mechanisms through which future labor supply and benefit utilization outcomes could be affected by longer wait times even in the absence of long-term health effects. First, to the extent that longer waits result in longer spells on sick leave while awaiting treatment, the interruption from work could contribute to human capital depreciation, including the loss of networks and lower productivity (Rees, 1966; Mincer, 1974; Mincer and Ofek, 1982; Becker, 1991; Calvó-Armengol and Jackson, 2004). If longer separations reduce future earnings capacity, or otherwise impede a worker's career trajectory, this could reduce the utility cost of temporary or permanent work absences in the future for workers who experience longer absence spells while waiting for treatment.

However, we suspect human capital deterioration offers, at most, a partial explanation for the labor supply effects we estimate. Existing research suggests that the average annual rate at which human capital depreciates during separations from work is less than 2% (Arrazola and Hevia, 2004; Weber, 2014), which implies that 10 days of increased workplace absence translates into expected productivity losses of around 0.05%. It seems unlikely that productivity losses of this magnitude could induce a 0.4 percentage point increase in disability entry.

Nonetheless, we cannot rule out human capital depreciation as an important contributing factor. Rates of human capital depreciation likely vary across occupations and across workers of different skill levels and types. For instance, Görlich and De Grip (2008) find evidence that higher-skilled females self-select into occupations for which the wage penalties from career interruptions are smaller, while Weber (2014) finds evidence of higher depreciation rates among less educated workers.[32] It is possible that the subsample of workers entering disability in the aftermath of longer waits are those for whom the human capital depreciation implications of work interruptions are especially severe. Our finding that the labor supply implications of

[32] In contrast to the Weber (2014) result, Edin and Gustavsson (2008) find no evidence of differential skill deterioration among less educated workers when analyzing an explicit measure of literacy skills.

waiting are concentrated among lower-educated workers is broadly consistent with Weber (2014) in that respect.

A second and more provocative explanation of our findings draws on theories of social identity (Sowell, 1975, 1981, 2005; Hofstede and Bond, 1988; Barke et al., 1997; Chiswick, 1983; Murray, 1984) and endogenous preference formation (Bowles, 1998). A worker's sense of identity might be influenced by experiencing a longer period of time in a "debilitated" or "disabled" state, thereby altering the preferences that worker subsequently exhibits towards work and the utilization of health-related benefits. If longer wait times increase the likelihood of a person self-identifying as "disabled" or "work impaired," this could increase that individual's propensity for future sickness-related work absences.

To date, credible evidence pertaining to the importance of social identity in preference formation remains largely limited to experimental settings. A common research design in this literature has been to randomly "prime" subjects in order to make particular aspects of that subject's identity more salient, and then to test how the preferences of primed subjects are affected (Benjamin et al., 2010, 2016). While such studies reveal the general importance of social identity, they say nothing about the extent to which an individual's self-identity might be affected by personal life experiences (like spending a longer time in a debilitated state), or the extent to which such experiences translate into meaningful changes in real-world behavior. Our findings cannot be definitively attributed to wait time-induced changes in workers' self-identity, but they do support the plausibility of such a mechanism and suggest that the real-world implications could be sizable. A prospective path for future researchers on this issue would be to augment the types of data we exploit with psychological survey data to directly investigate the relationship between longer wait times and patients' self-image.

Regardless of the specific mechanism(s) at work, our findings have important policy implications for countries with centrally-run healthcare systems, where wait times for non-emergency hospital services are a persistent concern. While health policymakers have long acknowledged the potential costs borne by patients who are made to wait longer for care, we also find evidence of sizable fiscal spillovers for a country's social insurance programs. Specifically, we estimate that a 10-day increase in patient wait time is associated with an average increase in sick leave and disability payments of around NOK 6400 (or USD 740) over the five years following referral (time discounted to the referral date). When projecting the potential savings over a 20-year period from the referral date, this figure rises to NOK 14,950 (or USD 1730).[33] While we cannot definitively conclude that wait times for orthopedic procedures are too long in Norway, back-of-the-envelope calculations indicate a hypothetical intervention to permanently reduce by one patient the number of patients awaiting orthopedic treatment would yield (discounted) reductions in transfers of over NOK 18 million.

Our findings serve as a general warning to policymakers about the indirect costs borne by social insurance schemes as a result of longer hospital wait times, but we would emphasize that the estimates produced here are context-specific. Average wait times in Norway are relatively long compared with other OECD countries. Sick leave is generously compensated in Norway, and participation in sick leave and disability programs is higher in Norway than in other OECD countries. One might therefore expect smaller labor supply effects in contexts where wait times are generally shorter, or where the institutional barriers and/or cultural deterrents to benefit use are stronger. Moreover, the estimates produced here are specific to orthopedic procedures – a setting where major health consequences due to waiting are not anticipated but where conditions are severe enough to impede work for some patients. Whether the effects detected here can be applied to other national or medical contexts is fertile ground for future research.

[33] As discussed in Section 5.5, this projection assumes the transfer effect estimated in year 5 decreases at a 10% annual rate in each of years 6 through 20.

References

Aakvik, A., T. H. Holmås, and E. Kjerstad (2015). Prioritization and the elusive effect on welfare – a norwegian health care reform revisited. *Social Science & Medicine 128*, 290 – 300. 1

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy 113*(1), 151–184. 4.1

Andrén, D. and D. Granlund (2014). Introducing waiting times for health care in a labor supply model for sickness absence. *Nordic Journal of Health Economics 3*(1), pp–34. 8

Angrist, J. D. and G. W. Imbens (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association 90*(430), 431–442. 25, 5.3, 5.4

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455. 25

Arrazola, M. and J. d. Hevia (2004). More on the estimation of the human capital depreciation rate. *Applied Economics Letters 11*(3), 145–148. 6

Barke, R. P., H. Jenkins-Smith, and P. Slovic (1997). Risk perceptions of men and women scientists. *Social Science Quarterly*, 167–176. 1, 6

Becker, G. S. (1991). *A treatise on the family.* Cambridge, Mass.: Harvard University Press. Enl. ed. 1, 6

Bellemare, M. F. and C. J. Wichman (2018). Elasticities and the inverse hyperbolic sine transformation. Working paper, University of Minnesota. 3.2

Benjamin, D. J., J. J. Choi, and G. Fisher (2016). Religious identity and economic behavior. *Review of Economics and Statistics 98*(4), 617–637. 6

Benjamin, D. J., J. J. Choi, and A. J. Strickland (2010). Social identity and preferences. *American Economic Review 100*(4), 1913–28. 6

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature 36*(1), 75 – 111. 1, 6

Brage, S., J. Nossen, and O. Thune (2013). Flere blir sykmeldt for symptomer og plager. *Arbeid og velferd 2*, 76–86. 1

Calvó-Armengol, A. and M. O. Jackson (2004). The effects of social networks on employment and inequality. *The American Economic Review 94*(3), pp. 426–454. 1, 6

Chiswick, B. R. (1983). The earnings and human capital of american jews. *Journal of Human Resources*, 313–336. 1, 6

De Chaisemartin, C. (2017). Tolerating defiance? local average treatment effects without monotonicity. *Quantitative Economics 8*(2), 367–396. 26

Edin, P.-A. and M. Gustavsson (2008). Time out of work and skill depreciation. *ILR Review 61*(2), 163–180. 32

Fiorini, M. and K. Stevens (2014). Monotonicity in iv and fuzzy rd designs-a guide to practice. Technical report, mimeo, University of Sydney. 4.2, 5.4

Godøy, A. and I. Huitfeldt (2018). Regional variation in healthcare utilization and mortality. 2

Goffman, I. (1963). *Stigma: Notes of the management of spoiled identity.* Englewood Cliffs, NJ: Prentice Hall. 3

Görlich, D. and A. De Grip (2008). Human capital depreciation during hometime. *Oxford Economic Papers 61*(suppl_1), i98–i121. 6

Hamilton, B. H. and R. E. Bramley-Harker (1999). The impact of the nhs reforms on queues and surgical outcomes in england: evidence from hip fracture patients. *The Economic Journal 109*(457), 437–462. 1

Havnes, T. and M. Mogstad (2011a). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics 95*(11-12), 1455–1465. 17

Havnes, T. and M. Mogstad (2011b). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy 3*(2), 97–129. 17

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432. 4.2, 25

Hirvonen, J., M. Blom, U. Tuominen, S. Seitsalo, M. Lehto, P. Paavolainen, K. Hietaniemi, P. Rissanen, and H. Sintonen (2007). Evaluating waiting time effect on health outcomes at admission: a prospective randomized study on patients with osteoarthritis of the knee joint. *Journal of evaluation in clinical practice 13*(5), 728–733. 1

Hirvonen, J., U. Tuominen, S. Seitsalo, M. Lehto, P. Paavolainen, K. Hietaniemi, P. Rissanen, H. Sintonen, and M. Blom (2009). The effect of waiting time on health-related quality of life, pain, and physical function in patients awaiting primary total hip replacement: a randomized controlled trial. *Value in Health 12*(6), 942–947. 1

Hofstede, G. and M. H. Bond (1988). The confucius connection: From cultural roots to economic growth. *Organizational dynamics 16*(4), 5–21. 1, 6

Hoogeboom, T., C. Van Den Ende, G. Van Der Sluis, J. Elings, J. Dronkers, A. Aiken, and N. Van Meeteren (2009). The impact of waiting for total joint replacement on pain and functional status: a systematic review. *Osteoarthritis and Cartilage 17*(11), 1420–1427. 1, 6

Hultin, H., C. Lindholm, and J. Möller (2012). Is there an association between long-term sick leave and disability pension and unemployment beyond the effect of health status? - a cohort study. *PLoS ONE 7*(4), 1 – 7. 1

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average. 25

Kapstad, H., T. Rustøen, B. Hanestad, T. Moum, N. Langeland, and K. Stavem (2007). Changes in pain, stiffness and physical function in patients with osteoarthritis waiting for hip or knee joint replacement surgery. *Osteoarthritis and cartilage 15*(7), 837–843. 6

Malmivaara, A., U. Hakkinen, T. Aro, M.-L. Heinrichs, L. Koskenniemi, E. Kuosma, S. Lappi, R. Paloheimo, C. Servo, V. Vaaranen, and S. Hernberg (1995, Feb 09). The treatment of acute low back pain–bed rest, exercises or ordinary activity? *The New England journal of medicine 332*(6), 351–355. 1

Martin, S. and P. C. Smith (1999). Rationing by waiting lists: an empirical investigation. *Journal of Public Economics 71*(1), 141–164. 1, 4, 5.5

Martin, S. and P. C. Smith (2003). Using panel methods to model waiting times for national health service surgery. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 166*(3), 369–387. 4, 5.5

Mincer, J. and H. Ofek (1982). Interrupted work careers: Depreciation and restoration of human capital. *Journal of human resources*, 3–24. 1, 6

Mincer, J., S. P. (1974). Family investments in human capital: Earnings of women. *Journal of Political Economy 82(2)*, 76–108. 1, 6

Murray, C. (1984). *Losing ground: American social policy, 1950-1980.* Basic books. 1, 6

Nikolova, S., M. Harrison, and M. Sutton (2016). The impact of waiting time on health gains from surgery: Evidence from a national patient-reported outcome dataset. *Health economics 25*(8), 955–968. 1

Parsons, T. (1951). *The Social System.* 3

Rees, A. (1966). Information networks in labor markets. *The American Economic Review 56*(1/2), 559–566. 1, 6

Sivey, P. (2017). Should i stay or should i go? hospital emergency department waiting times and demand. *Health Economics*, n/a–n/a. hec.3610. 4

Sowell, T. (1975). Affirmative action reconsidered. was it necessary in academia? evaluation studies 27. 1, 6

Sowell, T. (1981). *Markets and minorities.* Basic Books New York. 1, 6

Sowell, T. (2005). Are jews generic. *Black rednecks and white liberals*, 65–110. 1, 6

Stephens Jr, M. and D. J. Toohey (2018). The impact of health on labor market outcomes: experimental evidence from mrfit. Technical report, National Bureau of Economic Research. 2, 1

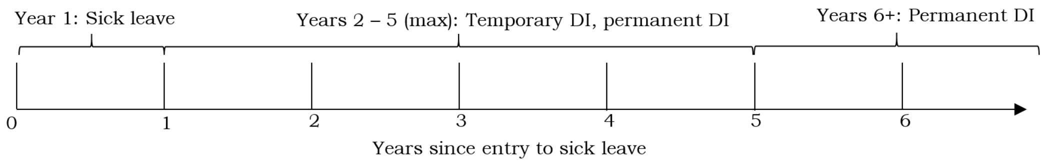
Tuominen, U., H. Sintonen, J. Hirvonen, S. Seitsalo, P. Paavolainen, M. Lehto, K. Hietaniemi, and M. Blom (2009). The effect of waiting time on health and quality of life outcomes and costs of medication in hip replacement patients: a randomized clinical trial. *Osteoarthritis and Cartilage 17*(9), 1144–1150. 1

Tuominen, U., H. Sintonen, J. Hirvonen, S. Seitsalo, P. Paavolainen, M. Lehto, K. Hietaniemi, and M. Blom (2010). Is longer waiting time for total knee replacement associated with health outcomes and medication costs? randomized clinical trial. *Value in Health 13*(8), 998–1004. 1

Weber, S. (2014). Human capital depreciation and education level. *International Journal of Manpower 35*(5), 613–642. 6, 32

Appendix A

Figure A1. Stylized benefits timeline



Notes: Figure illustrates time path of health-related benefits for a patient continuously claiming benefits who enters sick leave with full eligibility. See text for details.

Table A1. NCSP surgical coding - Chapter N Musculoskeletal system

	(1) Number of observations	(2) Average wait time
Ankle and foot	5,488	256.5
Hip joint and thigh	1,783	194.2
Knee and lower leg	11,291	151.8
Shoulder and upper arm	4,478	183.3
Wrist and hand	3,370	218.5
Total	26,410	190.3

Notes: Surgical procedures included in the estimation sample.

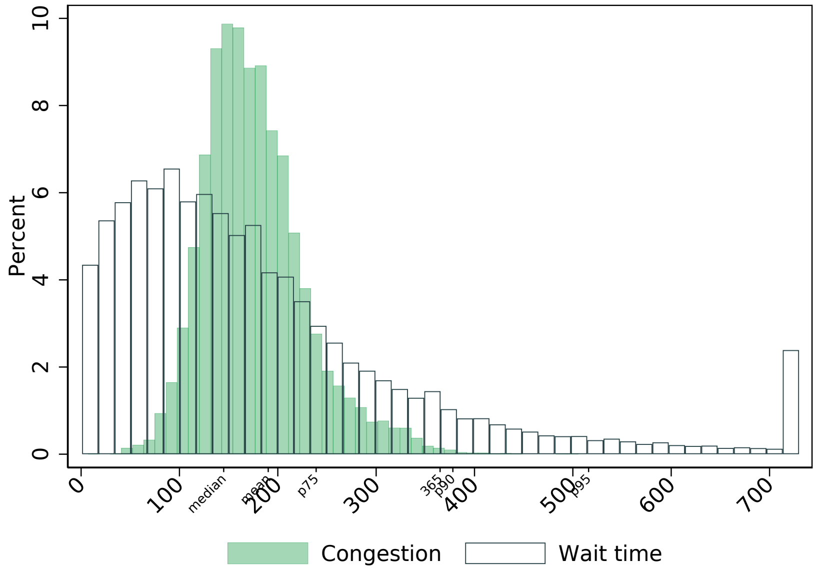


Figure A2. Distribution of individual wait time and congestion

Notes: Congestion is mean wait time for patients referred to the same hospital-by-procedure in the thirty-day window immediately preceding person i's referral (shaded histogram). Histogram for individual wait time in white. Labeled ticks on the x-axis refers to values of the distribution of individual wait time.

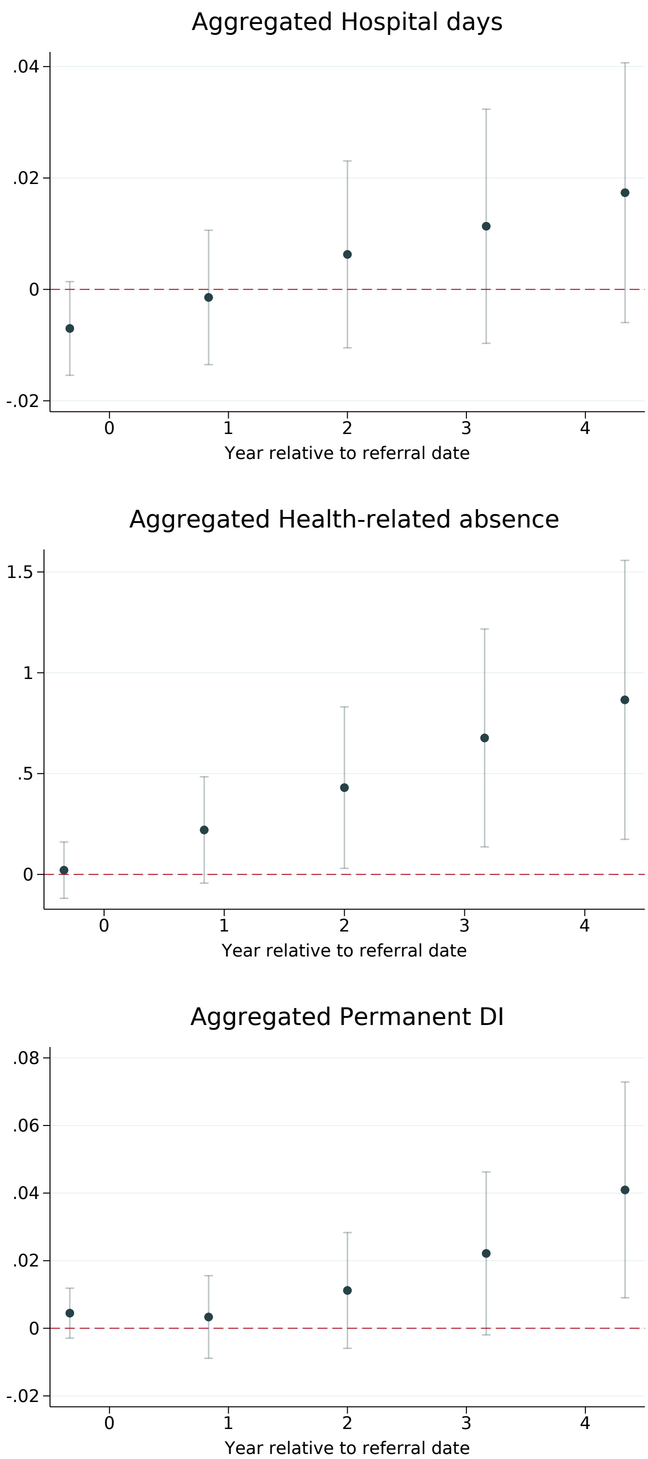


Figure A3. Effects by years since referral
Note: The figure plots the estimated effects of wait time on the number of hospital days and days of health-related absence relative to referral year.

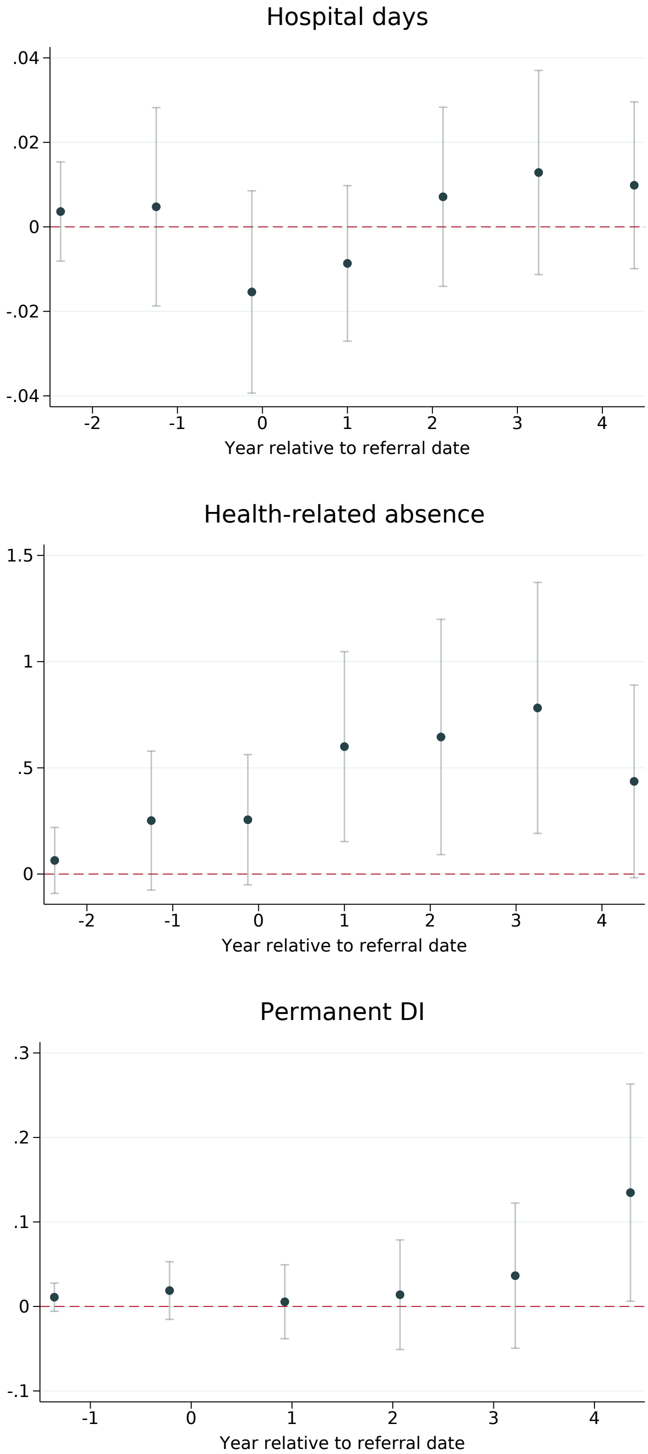


Figure A4. Effects for patients on sick leave on referral date

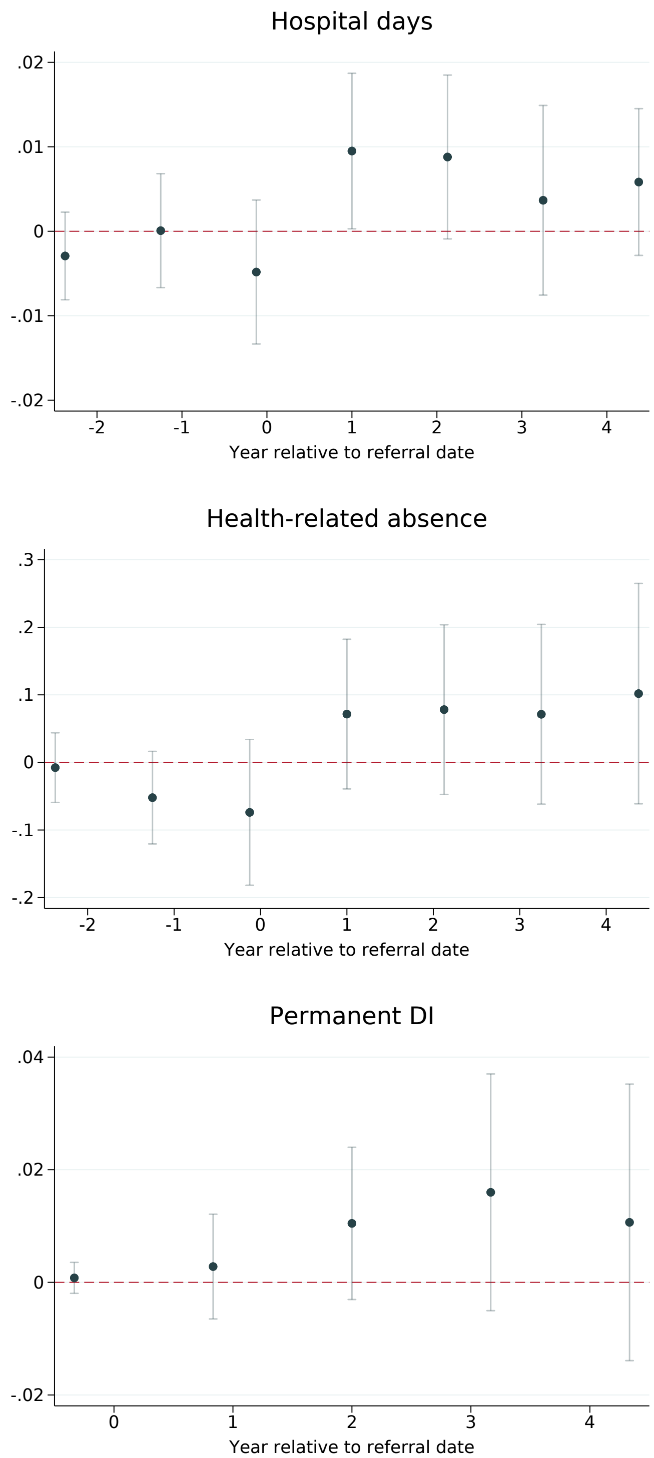


Figure A5. Effects for patients not on sick leave on referral date

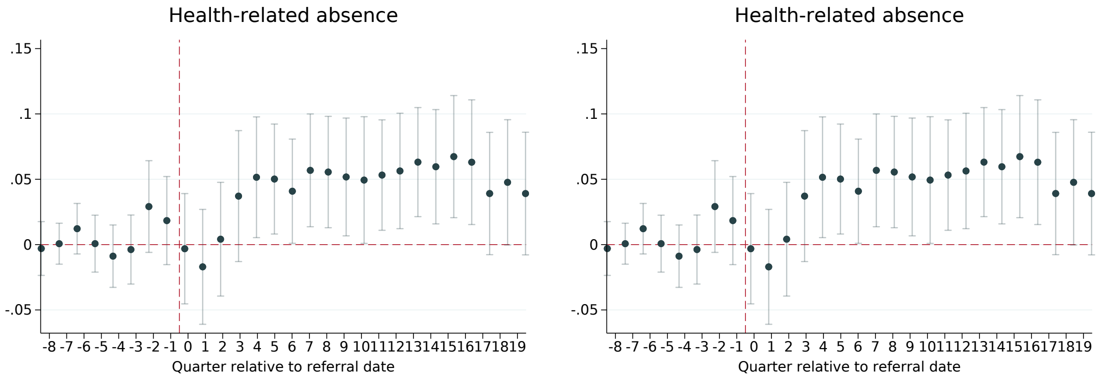


Figure A6. Estimates by quarter

Table A2. Effects by procedure

	(1) Shoulder	(2) Hand/wrist	(3) Hip/thigh	(4) Knee	(5) Ankle/foot
Panel A: First stage					
Congestion	0.339***	0.391***	-0.0534	0.327***	0.338***
	(0.106)	(0.134)	(0.237)	(0.0924)	(0.105)
FS F-stat	10.2	8.6	0.1	12.5	10.5
Panel B: Hospital days					
Wait time	-0.0231	0.00200	-0.0177	0.00331	0.0569*
	(0.0253)	(0.0233)	(0.373)	(0.0248)	(0.0299)
Dep. mean	18.497	18.072	26.350	16.095	17.073
Panel C: Health-related absence days					
Wait time	0.453	0.741	-10.79	0.878	1.144
	(0.854)	(0.644)	(49.11)	(0.734)	(0.728)
Dep. mean	536.378	343.464	479.221	331.200	309.931
Panel D: Permanent DI					
Wait time	0.00605	0.0611	-0.893	0.0383	0.0193
	(0.0340)	(0.0414)	(3.972)	(0.0335)	(0.0206)
Dep. mean	8.039	5.964	8.693	4.375	3.845
Observations	4,478	3,370	1,783	11,291	5,488

Notes: All regressions include fixed effects for year-by-month and hospital-by-procedure. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3. Effects by hospitals above/below median average wait time

	(1) Low congestion hospitals	(2) High congestion hospitals
Panel A: First stage		
Congestion	0.240***	0.415***
	(0.0753)	(0.0728)
FS F-stat	10.1	32.5
Panel B: Hospital days		
Wait time	0.0263	0.0166*
	(0.0399)	(0.0101)
Dep. mean	17.187	18.218
Panel C: Health-related absence		
Wait time	-0.344	1.174***
	(0.858)	(0.413)
Dep. mean	359.790	389.505
Panel D: Permanent DI		
Wait time	-0.0135	0.0547***
	(0.0335)	(0.0196)
Dep. mean	5.104	5.720
Observations	14,556	11,854

Notes: Sample is split by the hospital level median wait time. All regressions include fixed effects for year-by-month and hospital-by-procedure. Standard errors are clustered at the hospital-by-procedure level. Stars indicate significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01

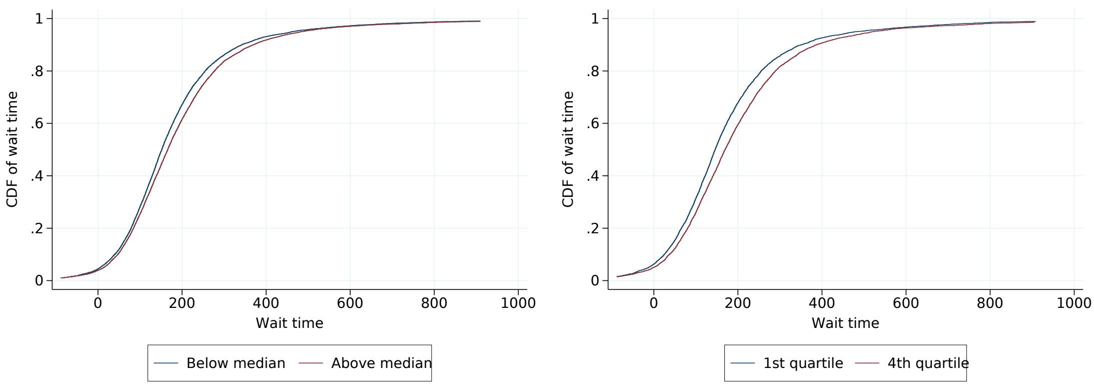


Figure A7. CDFs of wait time by quantile of instrument

Table A4. Robustness

	(1)	(2)	(3)
		No orthopedic surgery last	
	180 days	365 days	730 days
Panel A: First stage			
Congestion	0.334***	0.337***	0.331***
	(0.0599)	(0.0615)	(0.0634)
Dep. mean	191	191	190
FS F-stat	31.1	29.9	27.2
Panel B: Hospital days			
Wait time	0.0125	0.00895	0.00865
	(0.0127)	(0.0129)	(0.0139)
Dep. mean	17.471	17.384	17.253
Panel C: Health-related absence days			
Wait time	0.830**	0.873**	0.863**
	(0.385)	(0.384)	(0.391)
Dep. mean	367.888	365.535	366.371
Panel D: Permanent DI			
Wait time	0.0379**	0.0359**	0.0317*
	(0.0184)	(0.0179)	(0.0187)
Dep. mean	5.240	5.134	5.109
Observations	25,459	24,426	23,020

Notes: All regressions include fixed effects for year-by-month and hospital-by-procedure. Standard errors are clustered at the hospital-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5. Effects of congestion on delays

	(1)	(2)	(3)	(4)
	Any reason	Capacity	Delay due to patient's discretion	Medical reasons
Congestion	0.00441	-0.00655	0.0118	0.00435**
	(0.0141)	(0.00582)	(0.0123)	(0.00217)
Observations	23,925	23,925	23,925	23,925

Notes: All regressions include fixed effects for year-by-month and hospital-by-procedure. Standard errors are clustered at the hospital-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6. Exclude delays from IV

	(1)	(2)	(3)	(4)
	First stage	Hospital days	Total absence	Permanent DI
Congestion	0.347***			
	(0.0580)			
Wait time		0.0154	0.693**	0.0299**
		(0.0119)	(0.328)	(0.0133)
Observations	23,925	23,925	23,925	23,925
Dep. mean	189	17.579	369.731	5.296
FS F-stat	35.8			

Notes: All regressions include fixed effects for year-by-month and hospital-by-procedure. Standard errors are clustered at the hospital-procedure level. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix B: Cost-benefit analysis

Part 1: We first want to estimate the fiscal savings expected to accrue from reducing a worker's wait time by one day. Consider a representative worker, listed at the start of year 1, who waits T days for surgery instead of $T-1$ days. Using the average age (46.3) in the sample, we project estimated changes in benefit payments through year 20.

For each of the first 5 years after referral, we estimate IV models of annual benefit payments. Table B1 shows results from these models.

Our data does not allow us to estimate effects beyond $y = 5$. Instead, our baseline calculations assume that the effects estimated for year 5 continue through retirement. That is, assume $\beta dy = \beta d5$ for $y = 6, \ldots, 20$.

We then take all these parts and calculate the PDV of the stream of yearly effects. Let $\phi < 1$ represent the annual discount factor. The estimated change in the PDV of benefit payments can be expressed:

$$P\hat{D}V = \sum_{y=1}^{20} \phi^y \beta_y$$

This $P\hat{D}V$ gives the PDV for the expected "fiscal savings" resulting from reducing by one day the wait time for a single representative worker.

Part 2: We use this result to infer the fiscal savings that would accrue if an additional surgery was added on day 0, thereby allowing all subsequent patients to move up the queue by one place. We begin with the case where capacity constraints are always binding. In most cases, moving one place up the queue will not affect the treatment date. But one patient who would have been treated on day 1 now instead gets treated on day 0; and one patient who would have been treated on day 2 instead gets treated on day 1; and so forth. Note that if the system operated every day of the year, then for the year starting on day 0, we would observe 365 patients who had their wait times reduced by one day. The total change in patient wait times would be 365 days. If the system does not operate every day of the year (i.e. weekends, holidays), fewer patients in the year would see wait time reductions, but the total change in wait times would still be 365 days. That is, on every day going forward, there is a one-day reduction in some patient's wait time.

We can use this logic to calculate the PDV that accrues from all these wait time reductions as follows:

$$P\tilde{D}V = P\hat{D}V \sum_{d=1}^{\infty} \tilde{\phi}^d$$

where the $\tilde{\phi}$ denotes the daily discount factor (not the annual discount factor, ϕ, though the two are functions of one another). This $P\tilde{D}V$ gives the PDV for the expected "fiscal savings" arising from the insertion of one additional procedure into the system.

Part 3: While we do not have data on the cost needed to insert one additional procedure into the system, data on average costs may be suggestive. That is, we use the DRG payment levels as an approximate measure of the marginal cost hospitals incur for each procedure they perform under current capacity constraints and throughput levels. Presumably, it would cost more than this to insert an additional procedure into a hospital's operations.

Taking the average spending per procedure (NOK 33,150) as a benchmark, we construct some hypotheticals. For instance, suppose the cost of inserting an additional procedure into the system is twice that amount, because (say) the system needs to pay generous overtime wages to those contributing the extra work. Then we could compare $P\tilde{D}V$ to NOK 66,300 to determine whether inserting the additional procedure would yield a net cost reduction for the government. More generally, we can pose the policy question in the following way. How much more than its normal DRG rate should the system be willing to pay to insert an additional procedure into the system? The answer to this is given by $P\tilde{D}V/33,150$.

These calculations rely on two admittedly strong assumptions.[34] First, we assume that capacity constraints are always binding. This assumption implies that

[34] Moreover, we do not take account of the deadweight loss of distortionary taxes (i.e. we understate costs), though we can think of this as entering into the multiplicative factor linking marginal cost to average cost. We also do not take account of lost income tax revenue (understate benefits).

performing one additional procedure reduces wait times in perpetuity. However, if demand for surgical procedures fluctuates over time, there may be periods of excess capacity. Once we reach the point when capacity constraints are not binding, later entrants are no longer affected (even when capacity constraints become binding again). To address this, we calculate variations of total savings where the discount factor is adjusted to account for non-zero annual probabilities that hospitals reach a point of excess capacity. The results of this exercise are shown in columns (2)-(4) of table B1.

Similarly, as a lower bound, we can calculate what would happen if we were to use the marginal procedure to take out the patient at the back of the line, completely ignoring any spillover effects on later entrants. In this exercise, summarized in column (5), we use the sample average wait time rather than the discounted infinite sum of days when calculating total savings.

Second, for years 6-20, we assume that the effect of wait time on transfer is equal to the estimated effect in year 5. This may not hold true: effects in later years could be larger or smaller. In particular, it could be the case that longer wait times shift the timing of DI enrollment forward - patients with shorter wait times may still access DI in later years, in which case the effect would diminish over time. Panel B of table B1 illustrates how the calculations change when we assume that effects on transfers fall by 10% each year starting in year 6.

Table B1. Cost-benefit calculations

		(1)	(2)	(3)	(4)	(5)
Panel A: Baseline						
A.	PDV years 1-20	2339	2339	2339	2339	2339
B.	Daily discount factor	0.99992	0.99967	0.99884	0.99820	-
C.	Sum to infinity	12356	2988	858	555	168
D.	Total saving = A x C	28,895,631	6,987,633	2,007,342	1,297,886	391,707
E.	Cost saving ratio: D/33,150	871.66	210.79	60.55	39.15	11.82
Panel B: Convergence						
A.	PDV years 1-20	1495	1495	1495	1495	1495
B.	Daily discount factor	0.99992	0.99967	0.99884	0.99820	-
C.	Sum to infinity	12356	2988	858	555	168
D.	Total saving = A x C	18,467,143	4,465,783	1,282,888	829,477	250,339
E.	Cost -saving ratio: D/33,150	557.08	134.71	38.70	25.02	7.55
	Pr(Cap stops binding)	0	0.1	0.5	0.9	n/a

Note: The table shows calculations of the NPV value of reducing wait times by inserting 1 additional procedure (rows A-D) together with the cost-saving ratio defined as the largest ratio of marginal to average cost of surgery where the expected benefit exceeds the marginal cost.

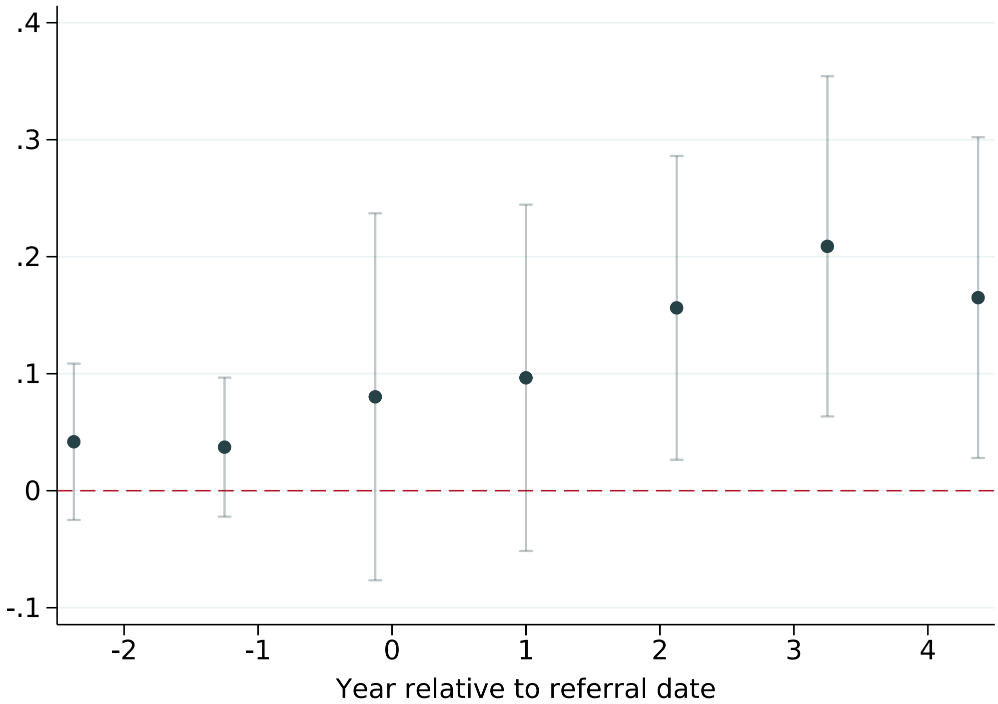


Figure B1. IV estimates for transfers.
Note: The figure plots the estimated effects of wait time on the number of hospital days and days of health-related absence relative to referral year.