Does the Internet improve health care? Evidence from the UK*

Sofia Amaral-Garcia,[†] Mattia Nardotto,[‡] Carol Propper,[§] and Tommaso Valletti[¶]

Abstract

We study the effect of internet diffusion on childbirth procedures performed in England between 2000 and 2011. We exploit an identification strategy based on geographical discontinuities in internet diffusion generated by technological factors. Our design shuts off the role of financial incentives for suppliers. We show that broadband internet access increased Cesarean-sections: mothers living in areas with better internet access are 1.8 percent more likely to have a C-section than mothers living in areas with worse internet access. The effect is driven by first-time mothers who are 2.5 percent more likely to obtain an elective C-section. The increased C-section rate is not accompanied by changes in health care outcomes of mothers and newborns. Therefore, health care costs increased with no corresponding medical benefits for patients. Heterogeneity analysis shows that mothers with low income and low education are those more affected: thanks to the internet, they progressively close the C-section gap with mothers with higher income and education.

*We thank Caterina Alacevich, Rita Santos, Daniel Prinz and seminar participants at the American-European Health Economics Study Group (2017), DIW Berlin, Erasmus University Rotterdam, INSEAD, Queen Mary London, and Universities of Bologna, Leuven, Lund and Melbourne for comments and suggestions.

†Hasselt University and KU Leuven. E-mail: sofia.amaralgarcia@uhasselt.be

‡KU Leuven and CESifo. Email: mattia.nardotto@kuleuven.be

§Imperial College London, University of Bristol, CEPR and IFS. Email: c.propper@imperial.ac.uk

¶Imperial College London, University of Rome II, CEPR and CESifo. Email: t.valletti@imperial.ac.uk

1 Introduction

In recent years, the diffusion of the internet has reduced informational frictions in several markets, with consumers having access to unprecedented sources of information about the price and quality of products. A high proportion of internet users look online for health information (Fox and Duggan 2013) and, with the spread of the internet, patients have found a new source of information on medical conditions, drugs, and medical procedures.[1] Health care is a market characterized by informational asymmetries between providers and consumers. Access to health information via the web is redefining the roles of 'supplier' and 'consumer', as the flow of information is no longer only from doctor to patient (Hartzband and Groopman 2010).

In principle, having access to broader information should lead to better decisions. However, as noted by Phelps (1992), the merits of more information in health care are complex. There are concerns about the quality of the information provided being difficult to interpret, if not misleading, about the capacity of users to make use of it (Eysenbach et al. 2002), and about patients without internet access (Wagner et al. 2005). As a result, one of the US Healthy People 2010's objectives was devoted to the quality of health information online, recognizing that the potential for harm from inaccurate information can be significant.

The aim of this paper is to assess if access to more information via the internet has changed health care treatment. Our setting is the UK in the 2000s and we focus on the impact of the internet on consumer choice of a common health care procedure. Examining the impact of internet access on health care treatments is difficult due to potential endogeneity of internet diffusion. Internet subscription is positively correlated with several observable demographic characteristics (such as income and education) that are also positively correlated with health care use and health outcomes. Hence, it is possible that unobservable demographic characteristics may be correlated with both internet access and health outcomes.

We address this issue with an identification strategy based on exogenous discontinuities in internet quality and access. These discontinuities stem from two key elements of the Asymmetric Digital Subscriber Line (ADSL) technology that was, by far, the most important way to gain access to the internet in the UK in the years we consider. The first is the decay of the digital signal, which means that the quality of all versions of ADSL internet connections strongly depends on the distance between the starting node of the connection (the local exchange, LE) and the delivery point (the house). The second is that the areas served by LEs are irregularly shaped because the topology of the network

[1] See Cline and Haynes (2001) for a review of the literature on health information seeking on the internet.

was designed in the 1930s to serve a different purpose (telephone voice communications).

Our strategy works as follows. We use a detailed map of the topology of the internet network to identify adjacent small local areas of around 650 households that are served by different LEs. Because LE catchment areas are irregularly shaped and different in size, contiguous small local areas can be located at very different distances from their respective node of the network and so will experience (potentially large) differences in quality of internet access. By matching data on LE coverage to detailed census and hospital discharge data we identify a subset of adjacent small areas that are balanced both in terms of aggregate demographics and patient characteristics but differ with respect to the quality and the availability of internet access (discussed in detail below). This identification strategy is very close to the one employed by Ahlfeldt et al. (2017), Faber et al. (2015), and Falck et al. (2014), who also exploit discontinuities in distance as an exogenous determinant of internet quality and internet access.

We apply our research design to the case of childbirth and assess whether internet diffusion affected the choice of childbirth procedure (C-sections) and subsequent health outcomes in the period when the internet was growing dramatically in England. We focus on childbirth for the following reasons. First, childbirth is one of the most commonly performed medical procedures worldwide, meaning that C-sections are one of the most common surgical procedures. The number of these cases not only gives us the statistical power to perform our analysis, but it also means that these cases are quantitatively relevant, particularly as there is widespread concern over the rise in C-sections which has occurred in many health care markets. Second, conditional on pregnancy, women have to deliver their children and in the UK, the vast majority of mothers deliver in hospitals,[2] which eliminates concerns of a possible selection bias. Third, mothers know their pregnancy status well in advance, giving them time to search for information should they wish to do so. Fourth, childbirth in the UK is tax-financed and free at point of delivery, so all consumers face the same zero price.

A number of empirical studies seek to understand why C-section rates are increasing over time. This literature emphasizes the role of the supply side and the role of physicians. The most common explanations are financial incentives for doctors or malpractice pressure (e.g., Currie and MacLeod 2008; Frakes 2013; Shurtz 2014; Amaral-Garcia et al. 2015). In contrast, we consider the role of the demander. Our design shuts-off the role for financial incentives for suppliers. In our setting, hospital staff are salaried and employed by one hospital. Hospitals are paid to provide maternity care through fixed budgets and have no incentive to provide more C-sections (in fact, as C-sections are more expensive than vaginal delivery, to the extent that suppliers take costs into

[2]In England and Wales, 2.3% of pregnant women give birth at home. http://www.nhs.uk/conditions/pregnancy-and-baby/Pages/where-can-i-give-birth.aspx.

account, they have incentives to provide fewer). Further, our design means we essentially look at only within-hospital variation as 97.5 percent of the mothers in our adjacent pairs of small areas use the same set of hospitals. In other words, the hospitals used by mothers in the adjacent small area are almost always the same hospital and in those hospitals the mothers will be treated by exactly the same staff.[3]

We find that mothers with better, faster access to the internet are more likely to have a C-section. Our effect comes only from an increase in elective C-sections: we find no effect of the internet on the likelihood of performing an emergency C-section (a decision made by the medical supplier and not the demander). We find no effect on either mothers or newborns' health outcomes. Finally, the increase is driven by first-time and low income mothers.

The estimated effect of having access to a broadband connection (of at least 2Mbit/s) compared with no internet access is to increase the C-section rate by 3.1 percent. Overall, the internet has been a non-negligible factor contributing to the observed increase in C-sections in the first decade of 2000s, when the C-section rate increased from 20.8% in 2000 to 24.3% in 2011 (a total increase of 16.8 percent). A back of the envelope calculation of the financial cost associated with the increase in C-sections due to internet diffusion is, at minimum, around £65m pounds per annum.

While we do not observe the exact channels by which mothers found information online, our findings can be explained by the role of the internet as a new source of diverse information. A number of consistent facts corroborate this conclusion. First, the ongoing discussion on the effects of the internet on health care choice, both in the medical literature and in the press, highlights the rising role of consumers in determining choice of treatment. Second, we document that the internet quickly became an important source of information on pregnancy and childbirth. Third, we show first-time mothers look more for information than multiple-time mothers, both offline and online.

Our findings support a mechanism of information gathering and joint decisions in a setting where the quality of information is heterogeneous. The internet is a very diverse source of information and one much less controlled by experts than information previously available in print and TV media. Mothers are therefore exposed to a wider range of signals online. First-time mothers, compared to other mothers, have less experience and knowledge about the pros and cons of different delivery methods and thus they are more likely to search for, and to be influenced by, online information. In line with this mechanism, we find that the increase in C-sections is driven by first-time mothers who opt more often for an elective procedure, while we do not find a difference for multiple-time mothers or for emergency C-sections. We also show that the effect we find is not driven

[3] We actually control for hospital effects in robustness tests and also examine whether there is any impact of the internet on choice of hospital. We find this not to be the case.

by mothers selecting hospitals with high C-section rates or traveling longer distances to find a hospital to have a C-section. This supports the idea that access to information gives mothers the ability to influence the decision of their normal supplier.[4]

We contribute to two different literatures. First, to the literature on the effect of increased information in health care markets, where informational asymmetries are a concern. Governments in many countries are seeking ways of providing health information to patients. For example, report cards were introduced in several US states, government websites have been introduced in the US to allow consumers to compare health and nursing care providers, and comparative information on the quality of hospitals is available in the UK and the Netherlands.[5] The literature typically focuses on how measures to increase information might affect health care quality and patients' choice. Evidence on whether these initiatives bring benefits is mixed, pointing to possible unintended consequences and to behavioral responses by both suppliers and consumers. For example, Dranove and Sfekas (2008) find that higher-ranking hospitals did not seem to gain significant market shares due to cardiovascular surgery report cards, a finding that is in line with previous literature (see Dranove and Jin (2010)). However, Dranove and Sfekas (2008) also show that when the information provided is different from patients' beliefs, patients respond by moving to higher-quality hospitals. Dranove et al. (2003) find evidence of "cream skimming": hospitals selected less risky patients and performed more cardiac surgeries on healthier patients with negative effects on overall health. Chou et al. (2014) show that online coronary artery bypass graft (CABG) hospital report cards increased hospital's quality competition and health outcomes for Medicare patients in more competitive markets.

For the UK, Gaynor et al. (2016) consider how a new information system which provided information on quality to help patients make more informed choices affected patients' hospital choice for CABG surgery. They find that, even for this relatively old

[4]From a theoretical standpoint, our findings can be rationalized by a standard model of choice with heterogeneous information. Our results are also consistent with a behavioral model of choice based on anecdotal reasoning - e.g. Spiegler (2006). The welfare implications of these two classes of models are quite different. In the behavioral models, in fact, decision makers do not know the true expected utility of each product offered in the market and base their decisions on the outcomes of a sample of cases they learn about during a search phase. Due to this limitation to the information acquisition process, suboptimal products are chosen in equilibrium. In our context, this mechanism can be triggered by the strong presence online of maternity blogs and websites reporting personal experiences and advice that is not always in line with official guidelines (see section 2.1 for a further discussion on mothers' antenatal information search). Our data are not sufficiently rich to disentangle between competing explanations for the documented findings but leave this to further research.

[5]One of the first and most well-known initiative is the New York CABG (coronary artery bypass graft) report card. Currently, many US states have implemented hospital report cards and extended the available information to other types of treatment. For government websites, see for example, https://www.medicare.gov/hospitalcompare/search.html, https://www.medicare.gov/nursinghomecompare/search.html, http://www.cqc.org.uk/, http://www.kiesbeter.nl..

and sick group, patients became more responsive to clinical quality and mortality rates decreased by approximately 3%. Gutacker et al. (2016) consider the case of hip replacement surgery and find that patient self-reported outcome measures impact patients' choice of provider.

Two papers focus on consumers of a similar demographic to those we examine here. Bundorf et al. (2009) find that patients respond to quality report cards when choosing providers of assisted reproductive therapies, and Price and Simon (2009) find that mothers respond to new medical information on vaginal birth after delivery, with young mothers being more responsive. The latter paper also notes these age differences are similar to those reported by Goldfarb and Prince (2008) regarding the adoption of new technologies such as the internet.

Secondly, we also contribute to the literature on the impact of internet use and diffusion. Among papers studying the market effects of the internet, Brown and Goolsbee (2002) find that internet use lowers the price of term life insurance by 8 to 15%, and Scott Morton et al. (2003) find that internet consumers pay less for their cars. Among papers studying how the internet changed behavior in other market dimensions, Falck et al. (2014) show the effect of internet diffusion on political participation and Gavazza et al. (2017) study its effect both on political participation and on the size of the government.

This paper is structured as follows: Section 2 describes the institutional setting of childbirth in England; Section 3 describes the process of internet diffusion; Section 4 discusses the data and identification strategy; Section 5 presents some descriptive evidence; Section 6 analyzes the effect of the internet on childbirth procedures and outcomes; robustness analysis is provided in Section 7; Section 8 concludes.

2 Childbirth and the health care system in England

As in most health care systems, childbirth is the most common medical reason for admission to hospitals in England.[6] In terms of costs, it represents approximately 2.8% of NHS spending (National Audit Office (NAO) 2013). There is a trend in many developed countries towards an increasing proportion of deliveries by C-section, much of which is not justified by any change in the risk profiles of mothers, leading to concerns about costs and unnecessary medical intervention (World Health Organization 2015).

In the UK, C-section rates have increased over time, from approximately 20.8% in 2000 to 24.3% in 2011. There are two types of C-section. Elective C-sections have a pre-booked date of delivery and the care pathway is locked-in in advance (Freeman et al.

[6] More than 500,000 women give birth each year in England (Hospital Episode Statistics, Admitted Patient Care, England - 2013-14: Procedures and interventions).

2016). Emergency C-sections are generally performed when labor has started and when it becomes clear that there might be risks for the mother or baby with a vaginal delivery. This type of delivery cannot be planned in advance. Figure 1 shows both elective and emergency C-sections increased approximately by 1.7 percentage points (from 8.4% and 12.4% in 2000, to 10.1% and 14.1% in 2011, respectively). This corresponds to a 20% increase in elective C-sections and a 14% increase in emergency C-sections. While both types of C-sections have increased, elective C-sections have proportionally increased more. In January 2016, the president of the Royal College of Obstetricians raised concerns about the increase in first-time mothers having C-sections, even though many have not had any problems during their pregnancy.[7]

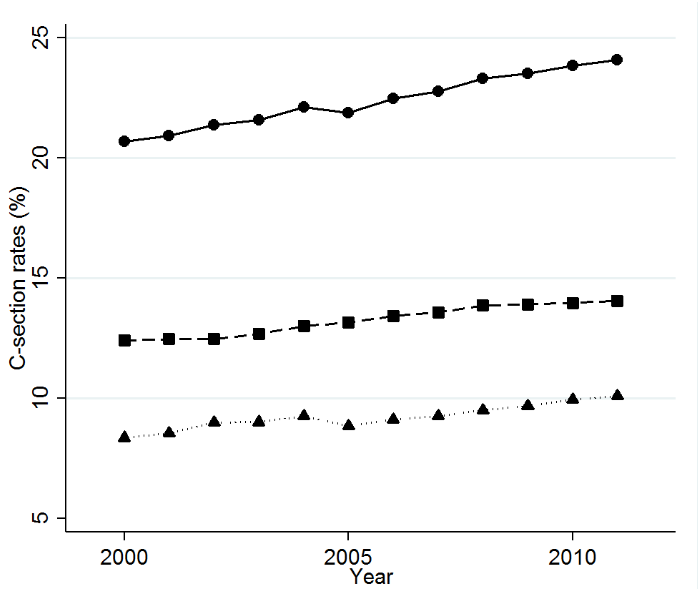

Figure 1: C-section rates over time. Overall C-section rate in circles; emergency C-section in squares; and elective C-section in triangles.

Health care in the UK is tax financed. The vast majority of health care, including maternity care, is provided by the NHS.[8] NHS hospitals are publicly funded and known as NHS Trusts. In order to access maternity care, pregnant women can self-refer to a midwife or go through their General Practitioner (GP), with the vast majority opting for the latter (House of Commons 2003).[9]

During the period of our analysis, choice of hospital was constrained for pregnant women and they almost always delivered in a hospital in their area of residence.[10] Ho-

[7]https://www.theguardian.com/society/2016/jan/31/caesarean-health-risks-c-section-first-time-mothers. A backup copy of this article can be found in the additional materials.

[8]There is a small private sector which specialises in routine elective care for which there are long waiting lists. Very few mothers give birth in private hospitals.

[9]Patients have a very limited choice of GP and they almost always have to choose a GP located near to where they live. See Gaynor et al. (2016).

[10]While hospital choice was implemented for elective care in 2006, maternity care was not included in

wever, there has been promotion of women's right to be involved in decisions and to have a choice in childbirth. While for other types of health care the meaning of choice was related to the right to choose a hospital, in maternity care the emphasis was put on aspects of the delivery, for example, use of an epidural anaesthetic, choice of a midwife unit, or birthing pool (Department of Health 1993; Thomas and Paranjothy 2001). A report from the Royal College of Obstetricians and Gynaecologists (Thomas and Paranjothy 2001) describes that "[o]ne of the priorities of maternity care is to enable women to make informed decisions regarding their care or treatment. To do so, they require access to evidence-based information, to help them in making their decisions". The same report mentions maternal request as a factor that arguably contributes to increases in C-sections.[11]

NHS employees (including midwives and obstetricians) have a fixed salary that does not depend on their performance or results (Freeman et al. 2016) and physicians are generally employed by only one NHS Trust (Gaynor et al. 2016). So while C-sections are more costly than natural deliveries, NHS employees do not benefit financially from performing them. In the period we study, hospitals were funded for maternity care from fixed budgets provided by publicly funded bodies covering specific geographic areas that have the task of buying hospital-based health care from NHS Trusts for their population (Gaynor et al. 2013). This does not provide incentives for hospitals to perform additional surgeries; in fact, this type of funding gives them pressure to reduce costs. Finally, all hospitals are subject to the same rules and incentives, independently of being located in an area with more or less internet diffusion.

2.1 Childbirth information and the internet

The amount of information about childbirth available to pregnant women has increased significantly, with the internet being one relevant source of information. We are particularly interested in understanding how internet diffusion influenced the choices made by mothers and whether certain types of mothers could be more influenced than others. We now consider how internet access might contribute to the formation of mothers' beliefs about childbirth procedures.

Expectant mothers become aware that they will be delivering many months beforehand. This gives them time to search for information if they wish to do so. Evidence

this reform. With respect to maternity care, there were plans to allow pregnant women to have access to choice of hospital after 2009. However, evidence indicates that in 2010 this was still not implemented (National Audit Office (NAO) 2013).

[11] There is some evidence suggesting that an increasing number of mothers are requesting a C-section delivery but the motivations for this demand are unclear (NIH 2006, NICE 2004). http://www.bbc.com/news/health-14806315 (a backup copy of this article can be found in the additional materials). This topic is a long standing part of the medical debate; see, for instance, Paterson-Brown et al. (1998).

shows that internet users search for health information online (Fox and Duggan 2013), and this holds true for pregnant women. Pregnant women are in age groups that are more likely to use the internet. According to the second US nationwide survey collected in 2006 on women's childbearing experiences, 76% of respondents stated they had accessed the internet for information about pregnancy and childbirth (Declercq et al. 2006). Similarly, Kolko (2010) finds that the introduction of broadband in the US was concentrated on relatively few online applications: downloading movies and music, online purchases, and researching health information. In the UK, a national survey of women's experience of maternity care found that 42% of women used non-NHS websites during their pregnancies (Redshaw and Heikkila 2010).[12]

In addition, there is evidence of a sizable difference in the search for information – across all sources – between first- and multiple-time mothers. In 2010 Redshaw and Heikkila (2010) surveyed a representative sample of approximately 5,000 women who gave birth in England, asking whether they read or consulted the NHS pregnancy book, visited the NHS website, visited other websites and blogs, as well as if they attended a pre-natal course. Results summarized in Figure 2 show that first-time mothers visited

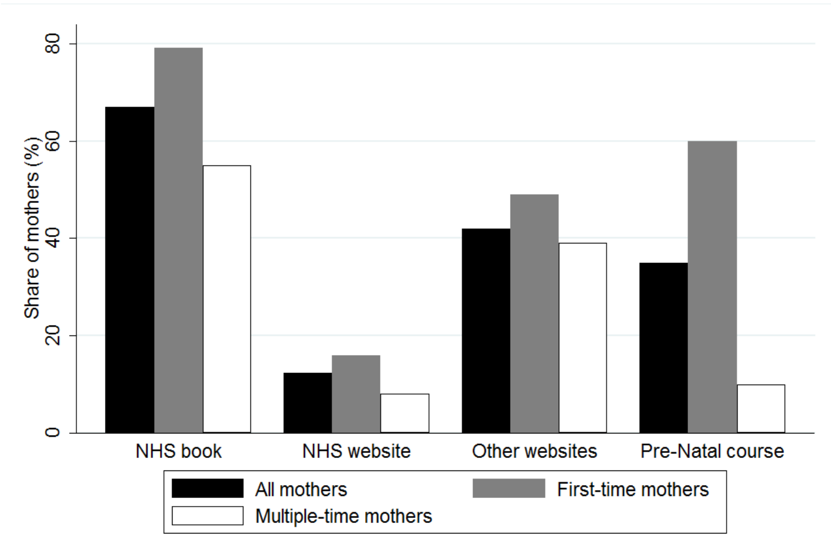

Figure 2: Share of mothers reporting access to different sources of antenatal information. Source: Redshaw and Heikkila (2010).

more information outlets than multiple-time mothers. On average, 67% of mothers read or consulted the NHS pregnancy book with a predominance of first-time mothers

[12]We also verified, using google trends (which covers January 2004 onwards), the recurrence of queries such as 'childbirth' or 'C-section'. They all show a rising trend in the interest index, particularly from 2006. While this does not distinguish between treated and control groups with respect to internet availability, it does confirm that the internet was used as a source of specific health information in the UK during the period we study, and progressively so, as the interest index is normalized to the total number of queries. Copies of the google trend graphs can be found in the additional materials.

whose share is 79%. Similarly, both the NHS website and other pregnancy websites and blogs attracted more first-time mothers (16% and 49%, respectively) than multiple-time mothers (8% and 39%, respectively).[13] Finally, an important source of antenatal information were pre-natal courses, which were attended mostly by first-time mothers.

Differences between first-time and multiple-time mothers with respect to sources of information during pregnancy were also found in the US (Declercq et al. (2006)). Experienced mothers primarily relied on their previous experience (48%), doctor or midwife (18%), the internet (13%), and books (12%). First-time mothers tended to rely primarily on books (33%), friends and relatives (19%), doctor or midwife (18%), and the internet (16%). Even though the difference between the percentage of first and multiple-time mothers who relied primarily on the internet is not significant, it is noteworthy that the internet was the primary source of information for more than 10% of all mothers. Moreover, for those mothers using the internet, usage was heavy.

The evidence shows that mothers with less experience search more for information, both offline and online. While exactly what women searched for is not known, there is considerable information about the risks and benefits of different modes of childbirth delivery. The most commonly cited benefits of a C-section include greater safety for the baby, less pelvic floor trauma for the mother, avoidance of labor pain, mother's fear of having a vaginal birth, the belief that there is a lower risk of fetal injury/death (Wiklund et al. 2007), preference for having a known date of delivery, and a decreased risk of postpartum hemorrhage (National Institutes of Health 2006). Some of the most common disadvantages are excessive bleeding, higher risk of infection, increased risk of complications in future pregnancies, and a higher risk of maternal death. Vaginal deliveries tend to have shorter hospital stay and recovery time but may entail a greater risk of obstetric tears and trauma to the mother along with incontinence (see Lavender et al. 2012 and references therein).

Future mothers can gather online information from different sources (*e.g.*, parenting websites, blogs, newspapers, scientific journals). However, this poses a challenge in terms of the accuracy and interpretation of this information. For instance, while scientific articles are more reliable, they are also more difficult to understand and access for the lay reader. On blogs and websites, the quality of the information presented is generally not verified. The exception is official medical websites, where there is an effort to update online content so that pregnant women can use it as a source of information. This is the case, for instance, for the US National Institutes of Health and the NHS in the UK. However, these efforts only started in 2010. An advantage of blogs and non-governmental websites is that pregnant women can read online about other mother's experiences during pregnancy and birth and share stories and concerns in an informal

[13] Note that in 2010 broadband internet penetration in the UK was approximately 75% of households.

way. Internet diffusion spurred the proliferation of blogs and websites supporting both childbirth methods. Those supporting C-sections emphasized the right to have one (not all mothers are aware of the possibility to ask for this procedure) as well as the physical and psychological consequences that may follow from a natural delivery.[14]

During the period we study, the UK media reported several cases of celebrities who chose to have a C-section. This lead to the expression "too posh to push".[15] This may have created the idea that mothers who make this choice are wealthier, famous, highly educated, or from a higher social class.[16] Weaver and Magill-Cuerden (2013) analyze media articles in main UK national weekday newspapers from 1999 to 2011 that used the expression "too posh to push". They argue that the media contributed to the idea that it is common to have a C-section on maternal request and that the association of the expression with celebrity contributes to the interest in the topic. Alves and Sheikh (2005) analyze hospital discharge data in the UK and conclude that there is evidence for the "too posh to push" hypothesis, as mothers from richer areas are more likely to request an elective C-section.[17]

Finally, internet diffusion should not have an impact on physicians' likelihood of performing a C-section. Physicians have expertise from many years of training and know the costs and benefits associated with both delivery methods. They do not need to rely on the internet to obtain this information. However, even if some physicians do use the internet more, our design is such that the mothers we compare give birth in the same hospitals. Thus differential access by providers of care should not affect our results.

[14]The right to have a C-section seems to be particularly important for tokophobic mothers, i.e., *"women who dread and avoid childbirth"*. (Hofberg and Brockington 2000). Nama and Wilcock (2011) identify tokophobia as a relevant reason for requesting a C-section.

[15]For instance, http://content.time.com/time/magazine/article/0,9171,610086-1,00.html. A backup copy of this article can be found in the additional materials. Weaver and Magill-Cuerden (2013) note that the first article using this expression was published in the Daily Mail in 1999, with the title "Are you too posh to push?".

[16]For instance, http://www.dailymail.co.uk/health/article-1388203/Too-posh-push-Caesareans-common-middle-classes.html. In the UK, a parenting website even published an article on celebrities who had a C-section http://www.parenting.com/parenting-advice/celebrities/12-celebrities-who-had-c-section-births. Backup copies of these articles can be found in the additional materials.

[17]It is also possible to find articles in the UK media about tokophobia. These articles tend to mention that this is a common reason to request a C-section. http://www.independent.co.uk/life-style/health-and-families/features/explaining-tokophobia-the-phobia-of-pregnancy-and-childbirth-9726809.html, http://www.telegraph.co.uk/lifestyle/wellbeing/healthadvice/4703329/Fear-of-the-nine-month-time-bomb.html, and https://www.theguardian.com/lifeandstyle/2010/mar/04/i-have-phobia-of-pregnancy. Backup copies of these articles can be found in the additional materials.

3 Internet diffusion in England

The diffusion of broadband internet in England started in the late 1990s. During the time period considered in this paper, most internet access was through the telephone network (with an aggregate market share of 80 percent), the only technological alternative being the cable network (accounting for the remaining 20 percent).[18]

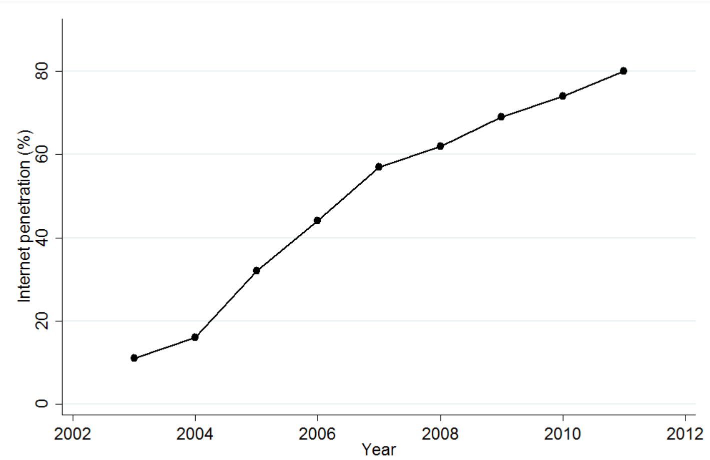

Figure 3: Broadband internet penetration between 2003 and 2011. Source: Eurostat

Providing internet access over the telephone network required an upgrade of the copper telephone infrastructure, but its broad footprint remained unchanged. The network is made of several nodes, called local exchanges, connected to each other. Each LE serves a number of houses connected via a fixed line, often called "the last mile". The ADSL technology, which enables the transmission of digital signals over the telephone network, requires the installation of special equipment in the LEs and replacement of the connections between LEs and the backbone with faster lines. This process of upgrade took several years, both because of the size of the investments and the regulatory framework.

The market for internet access in the UK was subject to open-access regulation that was similar to that which occurred in other European countries as a response to the fears that the market power of incumbent telephone operators would translate into similar market power in internet access.[19] The regulation introduced by the British regulator (Ofcom) consisted of three elements: first, it required British Telecom (BT), the owner of the telephone infrastructure, to functionally separate into two entities: Openreach, which became in charge of the maintenance of the network, and BT Wholesale, which became in charge of leasing the lines to entrant internet operators. This separation followed

[18] Mobile and fiber network jointly account for 1% of internet access over this time period.

[19] The European Commission asked the Member States to regulate their markets for internet access under the general framework outlined in the Regulation EC 2887/2000 and Directive 2002/19/EC. Each Member State then introduced its own legislation to guarantee entrant telecommunication operators the possibility to operate through the national telephone networks.

the repeated claims of entrant operators lamenting the attempts of BT to exclude them from the market and became effective in 2005. Second, BT was mandated to enable all LEs to host entrant operators who had the opportunity to make investments in the LEs by installing their own equipment, thus directly serving their costumers using the network of BT-Openreach. This technological option is known as *local loop unbundling*, LLU hereafter). It took approximately 3 years (2003 to the middle of 2005) to enable the LEs. Third, the wholesale price of leasing an internet line to entrants was regulated and, subsequently, twice revised downwards in 2004 and 2007.

The effect of the regulatory interventions, which changed the market in the middle of the 2000s and was followed by strong market entry, helped the UK to close the gap in internet access with other countries. The number of internet service providers grew rapidly and the market share of LLU operators went from only 2.2% at the end of 2005 to almost 40% at the end of 2009, thus becoming (in aggregate) the most popular choice among internet users. Entrant operators adopting LLU technologies also boosted improvements in the quality of internet access, offering substantially faster internet connections than those of the incumbent.[20]

Figure 3 reports the evolution of broadband internet penetration in England. The change in the pace of the diffusion process is remarkable. In the 5 years between 1999 and 2004, broadband internet penetration rose from almost zero to 16%, whereas in the following 5 years it jumped to almost 70%. However, this rise in broadband access was not homogenous. Urban areas led the process of internet diffusion. Another source of digital divide, but at the local level, results from the characteristics of the ADSL technology. The main challenge when transmitting a digital signal over the telephone network is the decay in its strength, as strength decreases with the distance traveled. The decay is not only a function of distance but it depends also on the material of the line. The telephone network, being designed to provide (analog) voice services, was made of copper lines, which experience a strong decay when used for transmitting digital signals (contrary to optical fiber, which has almost no decay). This forced the deployment of fast (fiber) connections between LEs to guarantee that, at least into the LE, the transmission speed was high enough to provide broadband services. After that point, between the LE and the house, the copper connection suffers from a decay but

[20] For more details on the effect of open access regulation see Nardotto et al. (2015). Cable was the only alternative to the ADSL technology in the years we consider. Virgin Media, the cable operator, built a completely private network during the 1990s to sell cable-TV. This technology has two advantages over ADSL: first, it is easier to adapt for the provision of internet access and second, connection speed is higher than with ADSL, the gap being larger with the first versions of ADSL. However, the network deployed by Virgin Media was limited to the most densely populated areas of the country and was accessible to only approximately 50% of households in the UK. In total, the market share of Virgin Media was approximately 20%. Virgin Media was not subject to open access regulation and never granted entrant internet providers access to its network.

the speed can be acceptable if buildings are sufficiently close.[21]

In its 2011 annual report, Ofcom stated: *"A characteristic of ADSL broadband is that performance degrades due to signal loss over the length of the telephone line. This means that the speeds available to different customers vary significantly, with those with shorter line lengths (i.e. who live closer to the exchange) typically able to achieve higher speeds than those with longer line lengths. [...] We found that the average download speed received for 'up to' 20Mbit/s or 24Mbit/s ADSL packages was 6.6Mbit/s, and 37% of customers had average speeds of 4Mbit/s or less"*. This strong decrease in speed performance indicates the different conditions households had with regard to the quality of their connection depending on location. For many years after the introduction of ADSL, being located further than 2km away from the LE meant having very poor internet access quality. We exploit the discontinuities in quality and availability produced by this technological limitation to identify the internet's effect on consumers' health care choices.

4 Data and identification

Our empirical analysis seeks to identify the effect of internet diffusion on the health care decisions of expectant mothers and their health care outcomes. For this purpose, we combine several sources of data. First, we use data on the structure of the telephone network and broadband penetration to identify small areas where we can exploit exogenous discontinuities in internet penetration. Second, we match the selected areas with census data in order to verify their balance with respect to a wide range of demographic characteristics. Third, we combine the selected areas with a dataset of all childbirth cases in England between 2000 and 2011. In the following sections, we describe these data sources and discuss our identification strategy.

Internet data. The data on internet diffusion are provided by Ofcom and have been previously used by Nardotto et al. (2015) and Ahlfeldt et al. (2017). The data contains detailed information on the telephone infrastructure. In particular, it reports the exact location of each LE (the geographic coordinates) and all postcodes (7-digits) served by each LE.[22] There are 5,587 LEs in the UK, of which 3,832 are in England.

Ofcom did not collect detailed information on the evolution of the broadband market

[21] The obvious alternative to the telephone network would be the creation of a fully fiber network with all copper connections replaced by fiber ones. This system is referred to as fiber-to-the-home and deployments costs are very high. It is now slowly making progress, again starting with the most densely populated areas.

[22] There are approximately 1.7 million active postcodes in the UK. On average, a postcode covers an area with a radius of 50 meters, but it is often smaller (*i.e.*, a building) in urban areas.

until the end of 2005. The decision to finally start systematic data collection was partly the result of the need to monitor the evolution of the market after the drastic changes in regulation described in Section 3, together with the fears related to the low take up of broadband subscriptions in the first part of the 2000s. The dataset collected quarterly by Ofcom on internet penetration covers the time period between the end of 2005 and the beginning of 2010. For each LE in the network, it reports the number of potentially served households, the number of subscribers for each internet service operator, and the number of subscribers of the cable operator. Thus, with this information it is possible to compute the total internet penetration in the area, defined as the ratio of the number of households with an internet subscription to the total number of households.

Data on demographic characteristics. Data on socio-demographic characteristics come from the UK Census, available at the small area level we examine. We use information on age structure, employment, and ethnicity. We also include two socio-economic variables provided by the hospital discharge data (described in "Childbirth data" below) that characterize small areas in terms of census data. For income, we use the Index of Multiple Deprivation (IMD) Income Domain, a measure of income deprivation of the patient's area of residence. For education, we use the IMD Education Training and Skills Domain, which captures the deprivation level in terms of education, skills, and training. IMD indices are constructed so that higher values imply higher deprivation.[23] We use these variables both as controls and to allow us to examine possible heterogenous effects across neighborhoods with high/low income and education.

Childbirth data. We use data from the UK Department of Health's Hospital Episode Statistics (HES) dataset for the financial years 2000 to 2011. The HES is an administrative dataset containing information on every English NHS hospital inpatient admission. The data contain detailed information on patients' conditions at admission to hospital and during their stay and on the medical procedures received by the patient.[24]

Using these data, we start by constructing variables that indicate risk factors such as diabetes, anemia, cardiac and lung conditions, previous abortion, previous C-section, placenta previa, multiple delivery (twins), eclampsia, hypertension, obesity, baby in breech position, fetal distress, and cephalopelvic disproportion.[25] The presence of high-risk factors increases the likelihood of having a C-section. In addition to identifying

[23] http://webarchive.nationalarchives.gov.uk/20120919132719/http://www.communities.gov.uk/documents/communities/pdf/131206.pdf.

[24] Patients' conditions are classified according to the ICD-10 codes from the International Classification of Disease. Procedures are classified according to OPCS codes, which is a procedural classification for the coding of procedures, operations and interventions performed in the NHS, and equivalent to the CPT codes in the USA.

[25] See for instance Currie and MacLeod (2008) and Frakes (2013).

whether a C-section was performed, we separate multigravida mothers (multiple-time mothers, *i.e.*, those who had at least one previous pregnancy) and primigravida (first-time) mothers. We also examine the utilization of induction of labor and episiotomy during vaginal delivery. Episiotomy utilization is indicated in the presence of certain complications of birth (*e.g.*, abnormal presentation, fetal distress).

We examine several measures of health care outcomes of mothers and newborns. With respect to maternal outcomes, we examine co-morbidities, prolonged labor and mothers' trauma during all vaginal deliveries (the last has been used by regulatory authorities as a measure of quality of maternal treatments).[26] For newborns, we examine prematurity, low birth weight, fetal distress, stillbirth, the need for resuscitation measures, and whether the newborn was discharged home or died/stayed in hospital. Finally, the discharge data contains an important variable for our identification strategy: the lower layer super output area (LSOA) of the patient's home.[27] There are approximately 32,000 LSOAs in England.

Identification. Our identification strategy relies on the discontinuities in internet penetration generated by the characteristics of the ADSL technology and the topology of the telephone network. As discussed in Section 3, the quality of an internet connection depends on its distance to the LE. Being located further from the LE means a larger negative gap between the potential and the actual speed of the connection, which negatively influences the decision to subscribe to an internet service provider or its actual availability. The spatial configuration of the telephone network results in discontinuities in the distance between the residences and the LEs, as the infrastructure was not meant to offer digital services and its roll-out, which took place in the 1930s, did not take into consideration the problem of signal decay.[28]

Figure 4 provides a graphical representation of our identification strategy, which consists of the following steps. First, we identify pairs of neighboring LSOAs that are served, in a sufficiently large proportion, by different LEs.[29] More precisely, each LSOA must have at least 95% of its postcodes served by only 1 LE and the two LEs serving the neighboring LSOAs have to be different. This reduces substantially the number of LSOAs

[26]Traumas are equivalent to Patient Safety Indicators (PSIs) 18 and 19. As defined by the Agency for Healthcare Research and Quality, PSIs are "a set of indicators providing information on potential in hospital complications and adverse events following surgeries, procedures, and childbirth".

[27]See: http://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography for more information on the UK Census geographies.

[28]This is evident from the topography of the network, as LEs are not located in a way that minimizes the average decay suffered by the households they serve. Moreover, these areas can be very heterogeneous in size even in urban areas, which penalizes those households located close to the border of a LE with a large catchment area.

[29]The borders of the catchment areas of LEs do not follow any census or administrative pattern so it is virtually impossible to find LE borders perfectly overlapping LSOA borders.

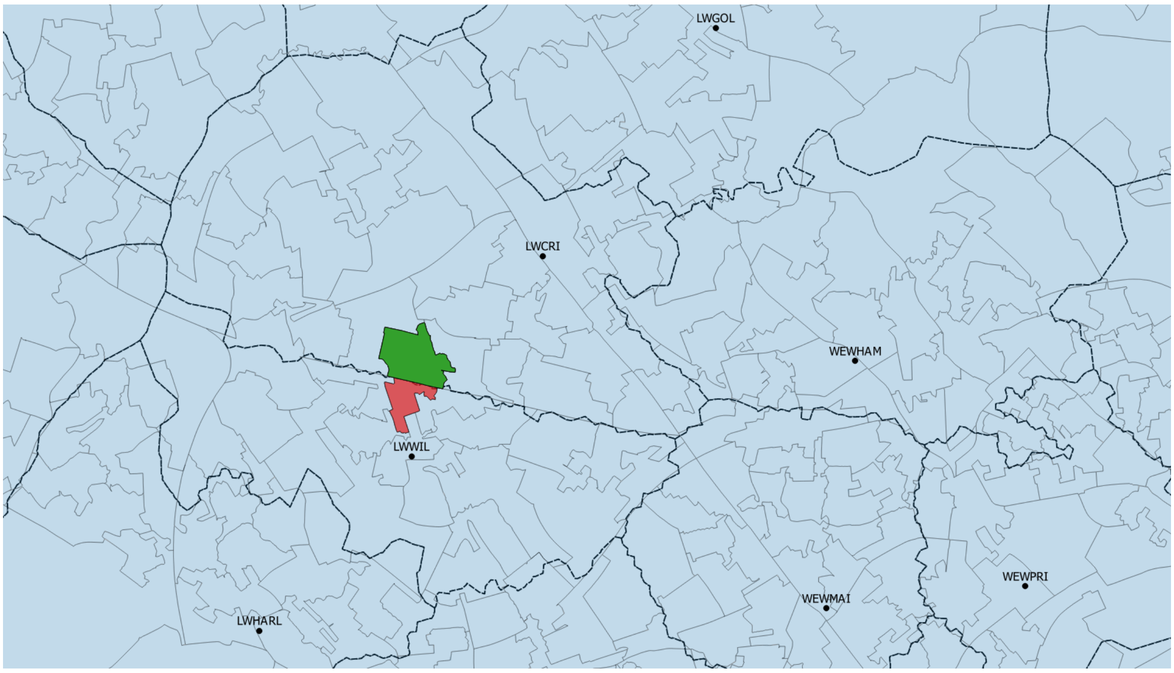

Figure 4: The map shows the borders of LEs (bold dashed lines), the locations of the LEs (black dots with LEs' names), and the LSOAs areas (light blue areas delimited by light grey lines). An example of matched LSOAs is given by the pair of LSOAs filled in red and green. They share a border but belong to two different LEs. The LSOA below the border (in red) is connected to a *close* LE labeled `LWWIL` while the LSOA above the border (in green) is connected to a *far* LE labeled `LWCRI`.

that we can use but leaves a sufficient number, as LSOAs cover smaller geographical areas than LE catchment areas. On average, there are 8 LSOAs in each LE. Matched pairs of LSOAs are very similar in demographic characteristics but differ in the distance to their respective LE and, in turn, in their internet penetration. Second, we compare the health choices of the mothers living in the matched LSOAs.

This identification strategy is very close to those in Ahlfeldt et al. (2017), Faber et al. (2015), and Falck et al. (2014). The strong decrease in the actual speed of internet connections at greater distances is measured in Ahlfeldt et al. (2017), who find a 65% decline in speed between two houses located at 1km and 3km from the LE and exploit the discontinuities in internet speed to measure the impact of better internet conditions on house prices. Falck et al. (2014) use data from Germany between 2005 and 2008 and focus on a 4km threshold, above which there was virtually no broadband internet availability. They compare municipalities located below and above this threshold to study the role of internet diffusion on electoral turnout.

We identified 1,209 pairs of LSOAs that we used to estimate the causal impact of internet access on health care choices and outcomes. Each pair of LSOA consists of one *close* LSOA and one *far* LSOA with respect to the distance to their corresponding LE. In the next section, we show that these pairs have balanced demographic and mothers' characteristics. Importantly for the interpretation of our results, we find that 97.5 percent of mothers in the matched LSOAs deliver in the same set of hospitals, which is consistent with the difference in size between the catchment areas of hospitals and of LSOAs, with the former being much larger that the latter (and also larger than the catchment areas of LEs).[30] We also show that the matched LSOAs, and the mothers living in these LSOAs, are very similar in characteristics when compared with the non-matched LSOAs and with the remaining mothers in the HES data.[31]

5 Descriptive evidence

In this section, we provide a descriptive analysis of the data. First, we report summary statistics of the demographic variables. Importantly for our identification strategy, we focus on the LSOAs that we employ in the regression analysis and we report tests for balanced characteristics. Second, we describe the broad trends in childbirth procedures performed between 2000 and 2011 in England and show that they are similar to those observed in the matched sample. Finally, we provide evidence on the relation between distance and internet penetration.

5.1 Summary statistics and trends in childbirth procedures

Panel A of Table 1 reports summary statistics of the demographic variables collected by the Census at the level of the LSOA. The full Census in England takes place every 10 years, and the last two waves have been collected in 2001 and 2011. The upper part of Panel A reports summary statistics from the 2001 Census while the lower part reports summary statistics from the 2011 Census. The left side of Panel A considers the full sample of English LSOAs – for which we report the mean, standard deviation, minimum and maximum of demographic variables – while the right side of the table considers the subsample of matched LSOAs, *far* and *close* – for which we report the mean of the demographic variables – and corresponding tests for balanced characteristics. The variables we consider are: total population, the share of people aged between 20 and 44 years old, the share of full-time and part-time workers, the share of white, and the share of high-skilled workers. Panel B of Table 1 reports summary statistics of mothers'

[30] The number of LSOAs, LEs and hospitals in England are approximately 32,000, 4,200 and 400 respectively, with size of their catchment areas reflecting (inversely) their number.

[31] The findings in Ahlfeldt et al. (2017) might raise concerns about a possible bias in the estimated effect of internet quality on childbirths. However, a number of reasons contributes to make a strong case for our identification strategy. Firstly, we find balanced socio-demographic characteristics on either side of the geographical discontinuity (i.e., when we compare *close* and *far* LSOAs). Secondly, the choice of the childbirth procedure (our outcome) is not influenced by any price consideration because the service is provided for free and the availability of information is the main driver of choice.

Table 1: Summary statistics.

A. LSOA Demographic characteristics	Full sample N=32,482				Matched sample N=2,418		
Variable	Mean	Std.	Min	Max	*Close*	*Far*	P-value
Census 2001							
Population	1512.80	198.67	999	6537	1509.75	1506.16	0.590
Age 20-44 (%)	35.29	8.36	7.55	83.08	35.10	35.81	0.031
Full-time workers (%)	40.74	8.25	4.14	83.83	41.41	41.27	0.663
Part-time workers (%)	11.88	2.70	0.57	26.97	11.73	11.75	0.895
White (%)	90.99	15.00	4.64	100.00	89.94	89.41	0.387
High skill workers (%)	41.54	9.36	11.54	92.04	42.78	42.17	0.112
Census 2011							
Population	1614.07	301.29	983	8300	1611.72	1604.70	0.521
Age 20-44 (%)	33.87	9.94	9.90	95.20	33.33	33.84	0.161
Full-time workers (%)	38.61	7.77	3.30	79.10	38.89	38.79	0.737
Part-time workers (%)	13.92	2.74	0.90	27.80	13.84	13.87	0.721
White (%)	86.23	18.72	0.60	100.00	84.65	83.83	0.298
High skill workers (%)	43.35	8.94	7.50	85.50	44.43	43.99	0.237
Number of deliveries (yearly)	15.64	14.99	1	2264	15.59	16.09	0.134
Distance LSOA - LE (km)	2.45	1.40	0.08	12.67	1.71	3.19	0.000
B. Mothers' characteristics and risk factors	Full sample N=7,033,942				Matched sample N=522,751		
Variable	Mean	Std.	Min	Max	*Close*	*Far*	P-value
Mean age	29.34	6.05	13	59	29.61	29.32	0.000
Twins (%)	1.57	12.43	0	100	1.6	1.57	0.357
Anemia (%)	4.50	20.74	0	100	5.01	5.1	0.136
Breech position (%)	5.67	23.13	0	100	5.69	5.69	0.959
Cardiac and lung (%)	0.24	4.86	0	100	0.26	0.24	0.189
Cord (%)	2.44	15.44	0	100	0.02	0.02	0.104
Diabetes	2.06	14.22	0	100	2.4	2.46	0.041
Cervix (%)	0.19	4.31	0	100	0.21	0.2	0.443
Hypertens. eclampsia (%)	3.52	18.44	0	100	3.54	3.46	0.128
Previous C-section (%)	7.52	26.36	0	100	7.62	7.64	0.804
C-section in 2000 (%)	20.68	40.50	0	100	20.87	21.10	0.571

Notes: **Panel A**. *Population* is the number of inhabitants in the LSOA. *Age 20-44* is the share of people aged between 20 and 44 years old. *Full-time workers* and *Part-time* workers are the shares of people aged between 16 and 64 years old who are employed with a full-time or part-time job respectively. *White* is the share of whites. *High-skilled* is the share of people aged between 16 and 64 years old working in financial intermediation and business activities, public administration and defense, education, and health care. *Number of deliveries* is the yearly number of deliveries in the LSOA. *Distance LSOA - LE* is the linear distance in kilometers between the geographical centroid of the LSOA and the local exchange that serves the LSOA. **Panel B.** *Age* is the mother's age in years at the moment of the delivery. Variables *Twins* to *Previous C-section* and indicator variables for mothers' risk factors: multiple deliveries (twins), anemia, whether the baby is in breech presentation, whether the baby suffers from a cord related problem, cardiac or lung conditions, diabetes, cervix problems, hypertension or eclampsia, and whether the mother had a previous C-section. *C-section in 2000* is an indicator variable for cesarean delivery where we restrict the sample to mothers who delivered in year 2000.

characteristics from HES. Consistent with the upper panel of the table, the left side of Panel B reports summary statistics for the full sample of mothers (more than 7 million over the 11 years between 2000 and 2011), while the right side focuses on mothers living in the matched LSOAs (more than half a million).

The statistics and the tests reported in Table 1 can be read in two ways. The first is to focus on the difference between the means of the variables of the full samples (both in Panel A and in Panel B) and the means of the variables of the matched sample; these means are very close. For instance, the average LSOA in the full sample has a population of 1512.8 inhabitants in 2001, while the average population of the matched LSOAs is 1509.75 for the *close* LSOAs and 1506.16 for the *far* LSOAs. This indicates that the matched LSOAs are very similar in characteristics to the rest of the LSOAs. This suggests that our findings have external validity.

The second aspect is the balance in characteristics between *far* and *close* LSOAs (Panel A) and between the mothers living in these areas (Panel B). Panel A shows that LSOAs (*i.e.*, the neighborhood where mothers live) have balanced characteristics, as most of the tests do not reject the null hypothesis of equal means. Importantly, this holds both at the start and at the end of the sample period, indicating that over the years *close* LSOAs did not attract households with different characteristics with respect to *far* LSOAs, indicating that migration in response to different internet penetration does not seem to pose a problem to identification in our problem.

The only demographic variable in Panel A for which the test rejects the null hypothesis of equal means is the share of population aged between 20 and 44 years old (p-value is 0.031) in Census 2001. Panel B, which reports summary statistics of mothers' characteristics, also shows good balance between *close* and *far* mothers. Variable's averages are very close and only in two cases we reject the null hypothesis of equal mean: age at delivery and the share of mother with diabetes. The average age of *close* mothers is 29.6 years old while the average age of *far* mothers is 29.3 years old. Despite the difference being small it is statistically significant due to the large sample size. We address this concern in one of our robustness checks (see Section 8.2) where we re-balance the samples with respect to age and verify that our results are not affected.

It is important to notice that, despite small imbalances in some variables, we find that the difference in C-section rates in 2000 for *close* and *far* mothers are not statistically significant (C-section rates are 20.87% and 21.1% respectively, as reported in the last row of Table 1).[32] Furthermore, we do not find a statistically significant difference between treated and control areas in the number of yearly deliveries. Finally, due to our

[32] Notice also that all our regressions include a large set of mothers' characteristics and LSOAs fixed effects. Thus, we are able to control for small differences in observed and unobserved (time-invariant in the latter case) differences in characteristics.

identification strategy, *close* and *far* LSOAs, display a large difference in their distance to the respective LE. *Close* LSOAs are located on average 1.71 kilometers away from the local exchange whereas *far* LSOAs are located on average 3.19 kilometers away, almost twice as much.[33]

6 The effect of the internet on childbirth

In this section, we examine how access to the internet affected childbirth in England. We focus only on the matched LSOAs, in which we distinguish between *close* and *far* LSOAs.

A first piece of evidence on the effect of the internet is in Figure 5 which reports C-section rates over time in the *close* LSOAs (dark grey) and in the *far* LSOAs (light grey). We interpolate the yearly rates of both groups of LSOAs with a kernel local polynomial smoother. The figure shows that in the first part of the period, when overall internet

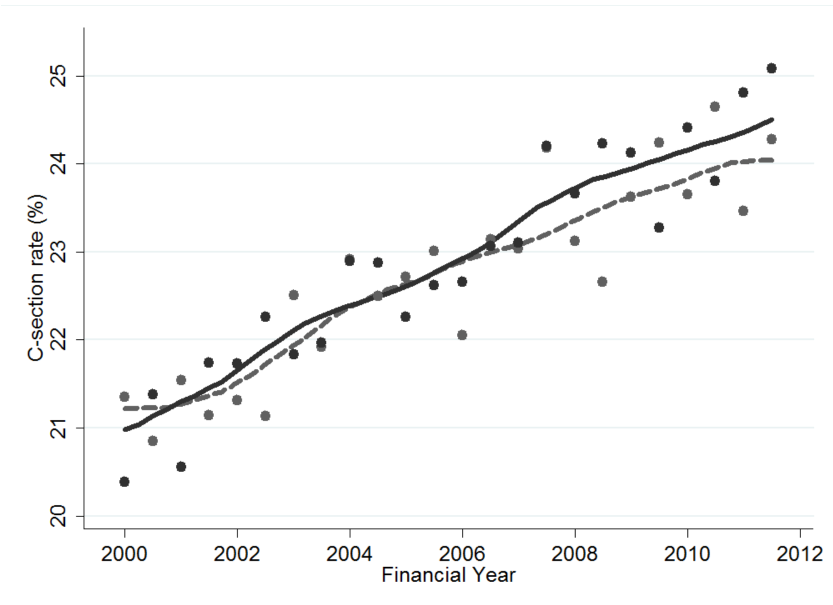

Figure 5: C-section rates over time. *Close* LSOAs are reported in dark grey (solid line) while *far* LSOAs are reported in light grey (dashed line).

penetration was very low, the two groups had very similar C-section rates. With the growth of internet penetration after 2005, *close* LSOAs experienced a more rapid growth of C-section rates compared to *far* LSOAs. To exactly locate the point of divergence in time, mothers become pregnant approximately 9 months before the delivery. Hence,

[33]As discussed in footnote 20, the main technological alternative to ADSL is represented by the cable operator Virgin Media, whose market share is approximately 20%. We test for systematic differences in cable coverage between *close* and *far* LSOAs. We find that *close* LSOAs have less cable coverage (approximately -2%) but this difference is not statistically significant. Thus, we conclude that *close* and *far* LSOAs are balanced also in terms of the alternative technology to access the internet.

the divergence that we observe in the second part of 2006 is due to mothers who got pregnant and looked for information in 2005.

The matching procedure described above allows us to perform a difference-in-differences analysis to measure the effect of internet access on health choices. Our regression model is:

$$y_{ijt} = \delta_1 close_{jt} + \delta_2 post_t + \delta_3 Did_{jt} + \beta X_{ijt} + Time_t + LSOA_j + \varepsilon_{ijt} \quad (1)$$

where i indicates the mother, j the LSOA and t the year.

The dependent variable y is the outcome of interest. The outcomes we focus on are: i) the probability of having a C-section (all, elective and emergency); ii) mothers' procedures and outcomes (induction of labor, episiotomy, co-morbidities, anaesthetic, prolonged labor, prolonged pregnancy, and maternal trauma/tears); and iii) newborns' outcomes (premature, low birth weight, distress, stillbirth, resuscitation measures, discharged or died/stayed in hospital).

The controls we use are as follows. X is a vector of mothers' characteristics, including age of the mother, the number of previous pregnancies, and risk factors. The presence of some risk factors strongly predicts the performance of a C-section. Consistent with previous literature (*e.g.*, Currie and MacLeod 2008, Frakes 2013) we include risk factors such as: diabetes, obesity, cardiac or lung conditions, previous abortion, fetal distress, breech position, cephalopelvic disproportion, multiple pregnancy, placenta previa, eclampsia and hypertension. $Time$ is a linear time trend. $LSOA$ are area fixed-effects that control for time-invariant unobserved factors.

The variables related to the diffusion of the internet are *post* and *Did*. The variable *post* takes value 0 before the financial year 2006 and value 1 thereafter.[34] Our *treatment* period starts in 2006, based on the path of diffusion of broadband internet access in England, documented in Section 3. Internet penetration in the first part of 2000s was very limited. As reported in Figure 3, before 2005 the broadband internet penetration was less than 20%, which means that the vast majority of mothers did not have broadband internet access. *A fortiori*, the difference in internet penetration between *close* and *far* LSOAs was very small before that year, if not zero. The spatial difference in internet diffusion we seek to exploit started with the rapid growth of internet penetration, *i.e.*, from 2005 onwards. Treated mothers are those who had their delivery between April 2005 and March 2006, which means that they got pregnant between August-September 2004 and July-August 2005. The variable *Did* is an indicator variable that is 1 if the mother lives in a *close* LSOA and if the childbirth takes place in the fiscal year 2006 or later. In section 7 we provide a robustness check where we vary the start of the treatment period.

[34] The financial year 2006 runs 1st April 2005 to 31st March 2006.

Table 2 reports the first set of results on the effect of the internet on C-sections. Models (1) to (3) consider all types of deliveries, with the following differences: model (1) considers all mothers; model (2) considers the subsample of multiple-time mothers (at least one previous pregnancy); and model (3) considers the subsample of first-time mothers. There are missing values in the recording of the variables multiple and first-time mothers (which are orthogonal to the treatment) so the number of observations is smaller when these variables are used. Models (4) to (7) exclude emergency C-sections with model (4) considering all mothers; model (5) focusing on multiple-time mothers; model (6) focusing on first-time mothers; model (7) using the sample of first-time and multiple-time mothers; and, finally, model (8) excludes elective C-sections. All models are estimated with clustered standard errors at the level of the matched LSOAs to control for correlation in the error term between neighboring areas.

Table 2: Regressions C-section model.

Dependent variable: Probability of C-section								
	All delivery types			Vaginal deliveries and Elective C-sections				Vaginal deliveries and Emergency C-sections
	All	Mult-time	First-time	All	Mult--time	First--time	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Did	0.44**	0.18	1.40***	0.31**	0.38	0.51*	0.62***	0.29
	(0.22)	(0.31)	(0.47)	(0.15)	(0.25)	(0.27)	(0.23)	(0.21)
Multiple × Did							-0.37	
							(0.25)	
Multiple-time							-2.25***	
							(0.15)	
Mothers covs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LSOA F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.301	0.429	0.219	0.478	0.567	0.364	0.521	0.189
Observations	522751	220519	135581	451531	195356	111448	306804	473989

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. There are missing values in the record of the variables multiple and first-time mothers (which are orthogonal to the treatment). Hence, the number of observations is smaller when these variables are used. Columns (1) to (3) use the sample of all mothers with the following differences: entire sample in column (1); multiple-time mothers in column (2); and first-time mothers in column (3). Columns (4) to (7) consider the sample of natural deliveries and elective C-sections with the following differences: all mothers in column (4); multiple-time mothers in column (5); first-time mothers in column (6); multiple and first-time mothers in column (7); and Column (8) considers the sample of natural deliveries and emergency C-sections.

The estimated coefficient of *Did* in column (1) indicates that the probability of having a C-section is 0.44 percentage points higher for mothers residing in *close* LSOAs compared with mothers residing in *far* LSOAs. This corresponds to a 1.8 percent increase in C-section rate. We consider multiple-time mothers and first-time mothers in columns (2) and (3), respectively. After controlling for mothers' risk factors, internet diffusion does not seem to have influenced mothers who had at least one previous pregnancy. However, we find evidence that first-time mothers are strongly influenced by the internet. The estimated coefficient *Did* in column (3), which is statistically significant at the 1% level, indicates that the probability of having a C-section for mothers living in *close* LSOAs is 1.4 percentage points higher than for mothers living in a *far* LSOA. Our results are robust to hospital fixed effects, as shown in Table 10 in the Appendix. This is as expected given that on average 97.5 percent of mothers in the paired LSOAs deliver in the same hospital.

The main results hold when we exclude emergency C-sections and focus on elective C-sections, with columns (4) to (6) confirming the previous findings. In column (7) we interact the indicator for multiple-time mothers with the *Did* variable. In this model specification the variable *Did* estimates the treatment effect on first-time mothers whilst the sum of $Multiple \times Did$ and *Did* quantifies the treatment effect on multiple-time mothers. We find that the probability of receiving a C-section for first-time mothers living in a *close* LSOA is 0.62 percentage points higher than the one of a counterfactual mother living in a *far* LSOA, which corresponds to a 2.5 percent increase. Instead, we find that multiple-time mothers living in a *close* LSOAs are only 0.25 percentage points more likely to have a C-section than their counterparts living in *far* LSOAs; a difference that is not statistically different from zero. Finally, column (8) reports the regression results when we exclude elective C-sections. In comparison with elective C-sections, emergency C-sections should not be influenced and we find that the coefficient for *Did* is not statistically significant.

Our results show that internet diffusion influenced childbirth procedures and the effect of access to information has been stronger for first-time mothers. Consistently with mothers looking for C-sections as a planned procedural choice, we find that the effect is driven by elective rather than emergency C-sections.

Estimated coefficients reported in Table 2 are the lower bounds for the true effect. This is because the indicator variable for the treatment takes value 1 in case of a *close* LSOA while it takes value 0 in case of *far* LSOA while internet access in the two LSOAs does not jump from 100% access to 0% access. This implies that, ideally, the coefficient should to be scaled by the gap in access between the two groups of LSOAs, which unfortunately cannot be observed. If however we use a measure of predicted internet quality, as reported in Section 6.1, we find that the effect of having access to a 2Mbit/s connection compared with 0 speed (no access) is to increase the C-section rate by 3.1 percent.

In the remainder of this section, we discuss the estimation results of a series of regressions estimating model (1) where the dependent variables are related to other procedures (*e.g.*, episiotomy), mother's outcomes and traumas (*e.g.*, prolonged pregnancy), and newborn's outcomes (*e.g.*, low birth weight).

Table 3 shows the regression results for induction of labor, episiotomy (for vaginal deliveries only), co-morbidities, anaesthetic, prolonged labor, prolonged pregnancy and mother's traumas. Labor induction consists in stimulating uterine contractions in order to try to achieve a vaginal birth. Episiotomy is a surgical incision that is made in the tissue between the vagina and the perineum. It is done during a vaginal delivery with the aim of enlarging the vaginal opening so that the baby can more easily pass through. Complications/co-morbidities is a variable that flags different types of co-morbidities that can occur during or after labor, such as cardiac complications of anaesthesia during labor and delivery, shock during or following labor and delivery, postpartum acute renal failure, and pulmonary complications. Anaesthetic considers cases in which the mother received some type of anaesthesia during labor or delivery. Prolonged labor is also referred as failure to progress. Prolonged pregnancy includes cases of pregnancy over 40 completed weeks to 42 completed weeks gestation. Column (7) presents the regression results for obstetric trauma/tears on mothers. There is no evidence of significant changes in these variables due to internet diffusion, except for prolonged labor, which seems to have increased.

Table 3: Regressions for other mothers' procedures and outcomes.

Dependent variable: Procedure and outcome	Induction (1)	Episiotomy (2)	Co-morbidities (3)	Anaesthetic (4)	Prolonged Labor (5)	Prolonged Pregnancy (6)	Maternal Trauma (7)
Did	0.10	0.07	0.29	-0.41	0.39*	0.04	-0.37
	(0.34)	(0.25)	(0.18)	(0.55)	(0.21)	(0.12)	(0.37)
LSOA F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mothers covs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Distance	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.108	0.112	0.046	0.214	0.033	0.024	0.052
Observations	473989	403193	522751	373259	473989	473989	403193

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. Results on the outcome *episiotomy* (column (2)) consider only vaginal deliveries. The number of observations is smaller for the outcome *anaesthetic* due to missing values. We test for zero correlation between the treatment variable and an indicator variable for missing value, and we do not reject the null hypothesis of zero correlation.

Finally, the outcomes for newborns are shown in Table 4.[35] We assess whether

[35]The total number of observations might not be equal to the total number of deliveries due to missing

there have been significant changes with respect to premature newborns (born before 36 gestation weeks), babies born with low birth weight (less than 2.500kg),[36] stillbirth, use of resuscitation measures, and whether the baby was discharged or died/stayed in hospital. We do not find significant effects for any of the newborn outcomes.[37]

Our results for this large set of maternal and newborn outcomes suggest that the increase in C-sections was not accompanied by improvements in either maternal or newborn health.

Table 4: Regressions for newborns.

Dependent variable:	**Premature**	**Low-birth weight**	**Fetal distress**	**Stillbirth**	**Resuscitation**	**Discharged**
	(1)	(2)	(3)	(4)	(5)	(6)
Did	-0.15	0.03	-0.30	-0.04	-0.10	0.08
	(0.25)	(0.21)	(0.27)	(0.09)	(0.25)	(0.09)
Mothers cov	Yes	Yes	Yes	Yes	Yes	Yes
LSOA F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.127	0.221	0.043	0.060	0.091	0.083
Observations	331618	325452	522751	381782	336880	522641

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. The number of observations is smaller in some columns due to missing values of the dependent variables (and the missing values are orthogonal to the treatment).

6.1 The effect of internet quality on childbirth

We have estimated the effect of the internet on childbirth procedures and outcomes by comparing mothers living in neighboring LSOAs that are *close* and *far* from their respective LEs. This approach relies on the fact that internet access is influenced by availability and quality of internet connections and has the advantage that the only information needed is the location of each LSOA and a detailed knowledge of the telephone network. The results obtained based on this identification strategy can be interpreted as an *intention to treat*, where the exogenous variation that we exploit are the discontinuities in the distances to the LE. These distance gaps translate into internet quality wedges, which have two effects on internet access and usage, firstly through the decision to subscribe to an internet provider and have fixed broadband at home and, secondly, to usage, which is strongly affected by the quality of the connection (faster internet access

information regarding these outcomes. Missing values are orthogonal to the treatment.

[36] We also run the same regression using *very low birth weight* as outcome variable (less than 1.500kg) and the results are the same. For both regressions we add as additional control variable gestational weeks, as premature babies tend to have lower weight.

[37] Some of these outcomes are rare, for example, stillbirth or the need to use resuscitation measures.

translates into easier browsing and quicker web searches).[38]

In this section, we relate the changes in health care choices to direct measures of internet access and of its quality (speed). Unfortunately, broadband penetration data has not been collected by Ofcom at such a disaggregated level as the LSOA.[39] Instead, we can proxy for the quality of local internet connections with two measures: the technology installed in the LE that serves the LSOA and the predicted speed of internet connections in the LSOA.

The first variable we examine to capture the quality of internet connections is the technology installed in the LE to serve the households located in its catchment area. With the start of the broadband internet era, BT and its competitors engaged in a roll-out process of newer technologies in the LEs in order to provide faster and more reliable internet connections. Our data allow us to observe the year of introduction of four technological configurations in each LE in England. This process of upgrade was not simultaneous, with some LEs receiving better technologies earlier than others. The first configuration is Integrated Services Digital Network (ISDN), which cannot be considered broadband, due to the very low data transmission speed of 128Kbit/s (reachable under ideal conditions). The second configuration, which was the first significant upgrade, is given by first version of ADSL. The bulk of the introduction took place between 2003 and 2005, when 80% of the LEs were upgraded.[40] The third technological configuration is the combination of ADSL technologies and the presence in the LE of investments under LLU. As explained in Section 3, LLU is an open access policy introduced by regulators in the UK to promote competition between internet operators. As shown by Nardotto et al. (2015), the main effect of this policy was to promote investments that increased internet access quality (speed). LLU investments started after the introduction of the first ADSL, also due to the regulatory framework, as discussed in Section 3, with most of LEs receiving such investments between 2005 and 2007. Finally, the fourth technological step has been the introduction of ADSL2+, which started in 2008 and continued in the following years.

The second measure of internet quality that we can compute with our data is the average theoretical speed of internet connections in the LSOA, which we can predict based on the estimates reported by Ahlfeldt et al. (2017). In their paper, the authors estimate a model for the *actual* speed of internet connections (not the advertised speed)

[38] As pointed out in Section 4, distance to the LE is the most important determinant of internet access quality, given the advertised speed promised by the providers. Depending on the technology installed in the LE, distance is such an important factor for quality that, beyond certain thresholds, signal decay becomes so significant to make broadband unavailable (see Falck et al. (2014)).

[39] The data we use in this paper, which were also employed in Nardotto et al. (2015), only report the average internet penetration at the level of the LE.

[40] The shares of LEs upgraded in 2003, 2004, and 2005 were 18%, 34%, and 27%, respectively, which add up to total of 79% in these 3 years.

as a function of distance from the LE and the type of ADSL technology installed in the LE, based on a large dataset of speed tests. They fit a fourth-order polynomial in distance whose coefficients are reported in equation (2).[41] LSOAs still served with ISDN (the standard technology employed before the introduction of broadband ADSL) are assumed to have a speed of 128Kbit/s.

$$Log(speed) = \begin{cases} 7.869 + 0.184d - 0.293d^2 + 0.058d^3 - 0.003d^4 & \text{if ADSL} \\ 8.214 + 0.057d - 0.287d^2 + 0.07d^3 - 0.005d^4 & \text{if ADSL+LLU} \\ 8.672 + 0.053d - 0.491d^2 + 0.141d^3 - 0.011d^4 & \text{if ADSL2+} \end{cases} \tag{2}$$

Thus, we use equation (2) and information on the ADSL technology installed in each LE/year, to compute the average predicted speed in each LSOA/year. Based on our calculations, the gap in distance between *close* and *far* LSOAs translates into a speed gap, with *close* LSOAs having a 0.7Mbit/s advantage over *far* LSOAs in 2005 (average predicted speeds are 1.8Mbit/s and 1.1Mbit/s respectively), which became more than 1Mbit/s in 2010 (average predicted speeds are 2.7Mbit/s and 1.6Mbit/s respectively).

The two empirical models we estimate are in equation (3). Here, our unit of observation is the pair of matched LSOAs, indicated with i, in year t.[42]

$$\begin{aligned} \Delta Csec_{it} &= \gamma Tech\ Gap_{it} + \Delta X_{it} + Time_t + \varepsilon_{it} \\ \Delta Csec_{it} &= \beta Speed\ Gap_{it} + \Delta X_{it} + Time_t + \varepsilon_{it} \end{aligned} \tag{3}$$

Starting from the set of control variables that enter both models, ΔX_{it} is the difference in mother's characteristics between the two LSOAs of pair i, and $Time_t$ is a linear time trend. In the first model, $Tech\ Gap_{it}$ is an indicator variable that takes value 0 if the ADSL technologies installed in the LEs that serve the two LSOAs of the pair i are the same, while it takes value 1 if ADSL technologies are different. The dependent variable $\Delta Csec_{it}$ is the difference in C-section rates between the LSOA with better technology and the other LSOA. In the second model, $\Delta Csec_{it}$ is the difference in C-section rates between the two LSOAs of pair i in financial year t, and $Speed\ Gap_{it}$ is the difference in predicted internet speed between the LSOAs of pair i in financial year t.

The two variables of internet quality we employ in (3) have different underlying identifying assumptions. The first model is based solely on the technological gap between matched LSOAs and not on the *close* versus *far* distinction. In other words, we retain the sample of neighboring LSOAs and we relate the difference in their outcomes to

[41] We employ only those coefficients that are statistically different from 0 at least at the 10% level to compute the predicted speed. See Table 1 in Ahlfeldt et al. (2017) for more details.

[42] We compute our variables at the LSOA level, not at the level of the mother. This is particularly sensible in the case of speed, as we do not know the exact location of each mother's house, while we can compute correctly the average distance between the houses in a LSOA and the corresponding LE.

the difference in the internet technology of their connections. Hence, this identification requires the absence of residual correlation between the technologies installed in the LEs that serve the paired LSOAs and the error term. We believe this is a reasonable assumption for the following reasons: (i) due to our design, which produces balanced LSOAs, and (ii) because the LSOA is a relatively small area compared to the LE and thus does not have a strong influence on the technological decisions taken at the LE level. The second model relies on the same identifying assumption of model (1), which requires the border's distance to the LE to be (conditionally) exogenous with respect to the characteristics of the LSOAs on the two sides of the border. The main limitation of this identification strategy is that our measure of speed is predicted based on the model in Ahlfeldt et al. (2017) and, thus, not actual speed (while model (1) only requires a weakly decreasing relationship between distance and speed). Finally, it is important to note that neither model in equation (3) relies on our assumption on the starting year of broadband internet diffusion as they are based on the difference between technology or predicted speed between paired LSOAs over time.

The added value of these alternative identification strategies compared to our baseline estimate of equation (1) is twofold. Firstly, they provide a further test for the main results based on the close-far identification strategy, relying on different identification. Secondly, they allow us to estimate the magnitude of the effect of improving internet access (either in terms of technology installed or internet speed) on the outcomes.

Estimated coefficients of model (3) are reported in Table 5 where columns (1), (3), and (5) report the estimates of the model where we use the technological gap, while columns (2), (4), and (6) report the estimates of the model where we use the predicted speed gap. Estimated coefficients confirm our previous findings. The estimated coefficients of the technological gap and of the internet speed gap reported in columns (1) and (2) show that better internet conditions increase the probability of performing a C-section in the whole sample, while the corresponding coefficients for the sub-sample of first-time mothers reported in columns (5) and (6) confirm that the effect is driven by this subgroup. The size of the coefficient of the speed gap estimated in (2) indicates that a difference in internet actual speed of 2Mbit/s (which is the minimum speed such that an internet connection was considered as *broadband*) determines an average increase in C-section rate of 0.72 percentage points. As the total increase in C-section rate experienced in England between 2000 and 2011 was 3.5 percentage points, it would explain approximately one-fifth of such increase.

Table 5: Internet tecnology/speed and C-sections.

Dependent variable: Δ Probability of C-section						
	All deliveries		Multiple-time mothers		First-time mothers	
	(1)	(2)	(3)	(4)	(5)	(6)
Tech gap	0.65**		-0.74		1.97**	
	(0.33)		(1.22)		(0.81)	
Speed gap		0.36**		-0.28		0.91**
		(0.18)		(0.34)		(0.45)
Mothers covs	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.309	0.309	0.361	0.361	0.209	0.209
Observations	14468	14468	12550	12550	10359	10359

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. Total number of LSOAs/years are not the same between paired columns because in some LSOA/years there were no deliveries by first-time and/or multiple-time mothers.

6.2 Heterogeneous effects of the internet on C-sections.

We now return to our main specification given by equation (1) and consider heterogeneous effects. As at the heart of our story there is the demand-side ability to access, filter and process information that can be found online, we study whether internet diffusion had a different effect depending on income and education. We construct two indicator variables for whether the mother lives in a neighborhood that is below or above the median IMD for income and education, training and skills (we simply refer to the latter as "education").[43]

Table 6 shows the regression results of model (1) where we focus on the subsample of first-time mothers and separate those living in high/low income from those in high/low education areas.[44] The results suggest that internet diffusion increased C-sections performed on first-time mothers residing in low income and low education areas. The *Did* coefficient is not statistically significant for first-time mothers living in neighborhoods with above median income and education.

Figure 6 also shows how C-section rates evolved over time, split by socio-economic background. There are increasing trends for both groups, but these also show interesting differences. Mothers from a higher socio-economic background always had a higher

[43] More precisely, these indicator variables are derived from measures of deprivation. As noted above in Section 4, the measure of income is the percentage of income-deprived households in the LSOA, while the measure of the education level is the percentage of adults who, according to the Office of National Statistics, are considered to have deprived education, training and skills.

[44] We repeated the same test for multiple-time mothers and the results are not statistically significantly different from zero. Available upon request.

propensity to have a C-section, and this increased over time but was not affected by internet availability. In contrast, mothers from a lower socio-economic background started with a lower propensity to have a C-section. Over time, they adopted C-sections more, and relatively more than the other group of mothers. The C-section 'gap' between the two groups reduced over our time period, and almost closed if we consider the split between low-income and high-income mothers. Our results suggest that internet availability contributed to the closing of this gap.

Table 6: Heterogenous effects by Income & Education (first-time mothers).

Dependent variable: Probability of C-section	High income	Low income	High education	Low education
	(1)	(2)	(3)	(4)
Did	0.62	2.30***	0.95	1.93***
	(0.70)	(0.68)	(0.73)	(0.63)
Mothers covs	Yes	Yes	Yes	Yes
LSOA F.E.	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
R^2	0.221	0.220	0.219	0.219
Observations	66475	69106	67419	68162

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

We also assess mother's and newborn's outcomes examining only the sample of those mothers who live in low socio-economic status areas (as proxied by the population having low education levels).[45] Table 7 reports the results for mother's outcomes and Table 8 shows the results for newborn's outcomes. There are no significant effects on any of these outcomes.

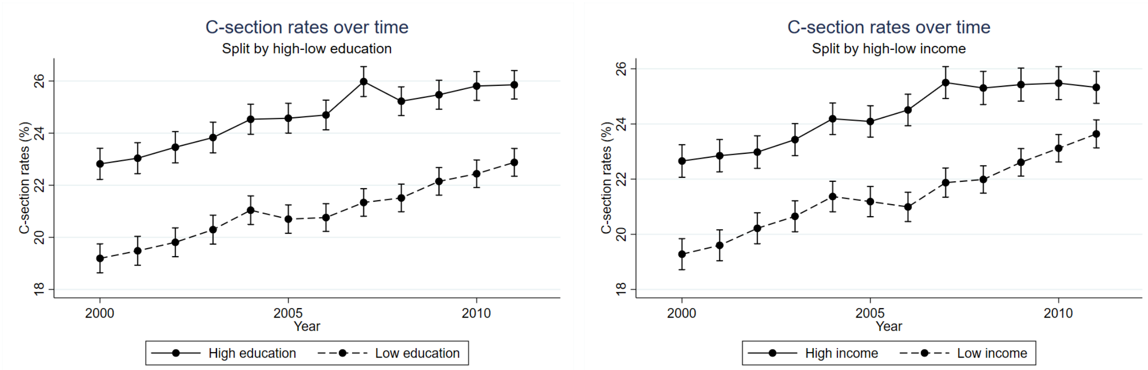

Figure 6: C-section rates over time. LEFT PANEL: split by education. RIGHT PANEL: split by income. Both figures report the 95% confidence interval around the observed frequencies.

[45]Similar results, available from the authors, arise when looking at low/high income mothers.

Table 7: Regressions mother's outcomes - low educated mothers.

Dependent variable: Procedure and outcome	Induction (1)	Episiotomy (2)	Co-mor-bidities (3)	Anaesthetic (4)	Prolonged Labor (5)	Prolonged Pregnancy (6)	Maternal Trauma (7)
Did	-0.21	-0.03	0.10	-0.74	0.46	0.03	-0.24
	(0.67)	(0.39)	(0.32)	(1.41)	(0.31)	(0.21)	(0.65)
LSOA F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mothers covs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Distance	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.119	0.112	0.047	0.244	0.031	0.023	0.047
Observations	205675	163495	205675	148371	205675	205675	163495

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. Results on the outcome *episiotomy* (column (2)) consider only vaginal deliveries. The number of observations is smaller for the outcome *anaesthetic* due to missing values. We test for zero correlation between the treatment variable and an indicator variable for missing value, and we do not reject the null hypothesis of zero correlation.

Table 8: Newborns' outcomes - low educated mothers.

Dependent variable:	**Premature** (1)	**Low-birth weight** (2)	**Fetal distress** (3)	**Stillbirth** (4)	**Resuscita-tion** (5)	**Discharged** (6)
Did	-0.17	-0.00	-0.16	0.08	-0.28	-0.06
	(0.60)	(0.50)	(0.52)	(0.19)	(0.50)	(0.20)
LSOA F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes
Age Groups	Yes	Yes	Yes	Yes	Yes	Yes
Mothers cov	Yes	Yes	Yes	Yes	Yes	Yes
Distance	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.130	0.244	0.045	0.042	0.085	0.083
Observations	139863	137605	205675	158656	139459	205633

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

6.3 Financial costs

The effect of internet diffusion on health care choices can have consequences on the total financial costs borne by taxpayers when the service is publicly funded (as in the UK), or on the final insurance cost borne by private insurers. In our case, while we did not find any impact of internet-driven C-sections on health outcomes, unnecessary C-sections can still generate a significant additional financial burden. Hospital costs are higher for C-sections than for vaginal deliveries, not only because the tariff for a C-section is higher (approximately £1,300 for a normal delivery without complications and £2,500 for a similar C-section), but also because the length of stay in hospital is generally longer. In our sample, the average length of stay is 2 days for a vaginal delivery and 4.4 days for a C-section. Imputing a daily hospital cost of £1,000, the extra cost of a C-section is thus approximately £(2,500−1,300) + £1,000×$(4.4-2)$=£3,600 per delivery. (This is a lower bound as we do not account for the fact that, after a woman delivers by C-section, there is a higher probability that subsequent deliveries will be by C-section). This is a relatively large sum given that we find no immediate associated medical benefits.

Based on our estimated effect, we can thus compute the extra-cost borne by the National Health Service that has been generated by the introduction of broadband internet, through the increase of the C-section rate. Considering an average of 586,000 deliveries in England *per* year from 2000 until 2011, and an average 3.1% increase due to internet diffusion, the yearly extra-cost amounts to approximately £65m pounds.

We are ignoring other benefits such as reducing labor pain, but given existing studies on how much people are willing to pay to avoid pain, there would still be considerable differences in costs. Olafsdóttir et al. (2017), for instance, use subjective well-being methods to estimate the value of pain relief, and find that it amounts to about $100 *per day*, with a lower willingness to pay for lower income people. As a normal delivery lasts for about eight hours, it is doubtful that mothers would pay out of their pockets for a C-section if the objective was only to reduce pain. To find a benefit from the increase in C-sections, we would have to consider other effects, such as the value from reducing gaps between C-section uptakes across different socio-economic groups. However, the medical literature suggests that the rise in C-sections is a concern. Therefore, closing this gap upwards in what is considered an unnecessary procedure for healthy mothers may not be a high social priority.

7 Further issues

7.1 Sensitivity of the treatment effect to varying the starting year

The results reported in Table 2 are based on having the financial year 2006 as the beginning of the treatment period.[46] This in turn builds on the uptake of internet access, which increased rapidly with the market and regulatory changes that took place in 2004 and 2005. Before 2005, overall internet diffusion was very limited both for *close* and *far* LSOAs. In other words, the majority of mothers did not have broadband internet at home - independent of living in *close* or *far* LSOAs.

Figure 7 reports the coefficient for the difference-in-differences variable *Did* in model (1) when we vary the starting year of the treatment. Moving the beginning of the

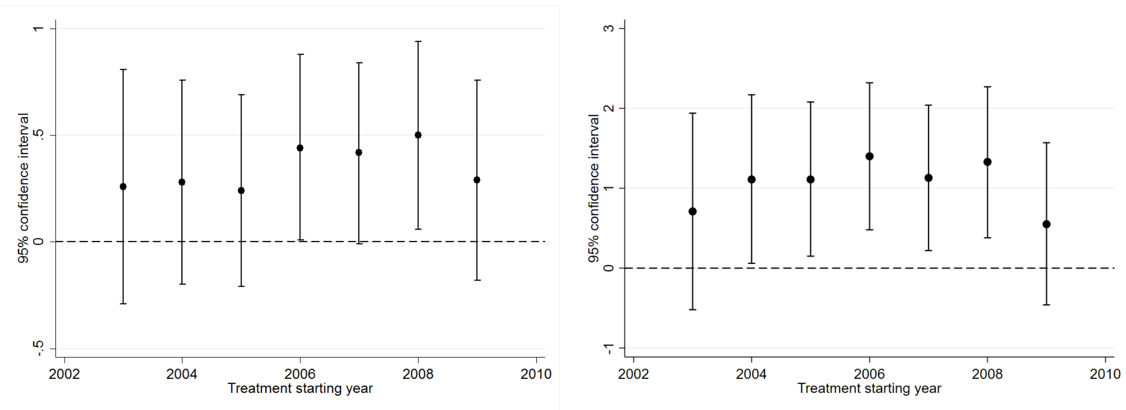

Figure 7: Estimate of the difference-in-differences *Did* coefficient with 95% confidence interval varying the starting year of treatment. LEFT PANEL: all mothers. RIGHT PANEL: first-time mothers.

treatment period toward the early 2000s reduces the estimated coefficient of the effect of the internet. This is consistent with a lower intensity of the treatment, namely the difference in internet access. In contrast, if we move the start of the internet period toward the end of the decade, the estimated coefficient increases until 2008, and then decreases. Two factors play a role in this case. First, by adding additional years after 2005, the pre-treatment period increases in the number of years in which the treatment is actually at work. Second, with our data we cannot know the difference in internet penetration between *close* and *far* LSOAs, which might have peaked after 2006 and then reduced.[47] Figure 7 shows results both for all mothers (left panel) and first-time

[46] As reported in Section 6, mothers who deliver between April 2005 and March 2006 (financial year 2006), got pregnant between August-September 2004 and July-August 2005.

[47] Access to the internet became so quickly diffused that, after 2009, the difference in internet penetration between *close* and *far* LSOAs most likely started to reduce as the vast majority of the population

mothers (right panel), with more statistically significant results for the latter group.

7.2 Choice of hospital vs choice of procedure

One possible consequence of the internet might be that mothers change patterns of care seeking and travel farther to find a hospital with a higher C-section rate. While we find this unlikely given the description of maternity care in England in Section 2, and also given the finding reported in Section 4 that 97.5% of mothers living in neighboring LSOAs deliver in the same set of hospitals, we empirically test for this change in behavior in two ways. First, we examine whether the distance from mother's residence to hospital changed after 2006 in *close* LSOAs compared with *far* LSOAs. Second, we test whether mothers with more exposure to the internet delivered more frequently in hospitals with higher C-section rates.[48]

Table 9 reports the results of a series of regressions where the empirical model is similar to the one in equation (1) but the outcome variables are the distance from mother's residence to the hospital, and two indicator variables for the hospital being in the top 10 or 25 percentile in the distribution of C-section rates in the year before giving birth. Should mothers opt for a hospital with high C-section rates, they would most likely consider rates from the previous year.[49]

Table 9: Choice of Hospital

Dependent variable:	**Distance**	**Hospital top 10% C-section**	**Hospital top 25% C-section**
	(1)	(2)	(3)
Did	0.07	0.00	-0.00
	(0.06)	(0.00)	(0.00)
Mothers covs	Yes	Yes	Yes
LSOA F.E.	Yes	Yes	Yes
Time trend	Yes	Yes	Yes
R^2	0.334	0.059	0.081
Observations	522751	473490	473490

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

We do not find evidence that the increase in C-sections due to internet access is associated with increased mobility. Column (1) shows that distance did not change after 2006 for *close* LSOAs. Column (2) shows the results using as dependent variable a dummy that is equal to one if the mother delivered in a hospital among the top 10

got access to the internet through broadband connections.

[48] Card et al. (2018) find that C-sections for low-risk first births tend to be accompanied by worse outcomes except where mothers sort to high C-section hospitals.

[49] We also consider C-section rates in the current year, and the results are similar.

percentile with respect to C-section rates in the previous year. Column (3) considers the top 25 percentile instead. Based on this evidence, we conclude that mothers with more internet access managed to obtain a C-section through the interactions with their local doctors instead of opting for a different hospital.

8 Conclusion

We examine whether internet diffusion affects health care choices and outcomes by studying how broadband take-up has influenced childbirth. We take advantage of geographical discontinuities in internet diffusion in England to estimate the causal impact of broadband internet access on use of C-sections and health care outcomes of mothers and newborns. Our design focuses on the demand side and shuts off the role of financial incentives for suppliers by examining patients living in adjacent pairs of LSOAs, almost all of whom deliver in the same hospital.

We find that mothers with better internet access have higher C-section rates. The effect is driven by an increase in elective C-sections, rather than in emergency C-sections where the choice is that of the medical supplier. We find no significant changes in procedures that are generally performed during labor and delivery (such as induction of labor, episiotomy, anaesthetics) or in the health care outcomes of mothers and newborns. The differences are driven by first-time mothers, who are less informed by experience about childbirth than multiple-time mothers, and by mothers living in areas where the population has poorer education and lower income.

Overall, this paper provides evidence that increasing access to the internet *per se* does not make patients better off from a strictly medical point of view. Our main findings – higher C-section rate and no sizeable improvements in health care outcomes – suggest a negative impact of internet diffusion on health care costs. More C-sections that are not strictly necessary are performed due to the internet and this is costly for the taxpayers without any gains in terms of health benefits. However, the welfare conclusions are a little more nuanced. We also find that the internet, by giving lower income/lower education mothers the ability to exercise choice over their treatment, considerably reduces the 'elective C-section gap' between higher and lower socio-economic status mothers. In a health care system where those with a higher socio-economic status have access to more tailored care without paying higher prices (Cookson et al. 2016), the internet may allow those who are currently less able to negotiate this to increase their access. Whether this is of benefit or not will depend on the quality of the information available to internet users.

References

Ahlfeldt, G., Koutroumpis, P., and Valletti, T. (2017). Speed 2.0: Evaluating access to universal digital highways. *Journal of the European Economic Association*, 15(3):586–625.

Alves, B. and Sheikh, A. (2005). Investigating the relationship between affluence and elective caesarean sections. *BJOG: an International Journal of Obstetrics and Gynaecology*, 112:994–996.

Amaral-Garcia, S., Bertoli, P., and Grembi, V. (2015). Does Experience Rating Improve Obstetric Practices? Evidence From Italy. *Health Economics*, 24(9):1050–1064.

Bates, D. W. and Gawande, A. A. (2000). The Impact of The Internet on Quality Measurement. *Health Affairs*, 19 (6):104–114.

Bhuller, M., Havnes, T., Leuven, E., and Mogstad, M. (2013). Broadband internet: An information superhighway to sex crime? *The Review of Economic Studies*, 80(4(1)):1237–1266.

Bloom, N., Propper, C., Seiler, S., and Van Reenen, J. (2015). The Impact of Competition on Management Quality: Evidence from Public Hospitals. *The Review of Economic Studies*, 82(2):457–489.

Brown, J. R. and Goolsbee, A. (2002). Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry. *Journal of Political Economy*, 110(3):481–507.

Bundorf, M. K., Chunb, N., Goda, G. S., and Kessler, D. P. (2009). Do markets respond to quality information? The case of fertility clinics. *Journal of Health Economics*, 28:718–727.

Card, D., Fenizia, A., and Silver, D. (2018). The health effects of cesarean delivery for low-risk first births. *NBERw24493*.

Chou, S.-Y., Deily, M. E., Li, S., and Lu, Y. (2014). Competition and the impact of online hospital report cards. *Journal of Health Economics*, 34:42–58.

Cline, R. and Haynes, K. (2001). Consumer health information seeking on the internet: the state of the art. *Health Education Research*, 16(6):671–692.

Cookson, G. and Laliotis, I. (2016). Staffing and the determinants of caesarean sections in the English NHS. *Working Paper*.

Cookson, R., Propper, C., Asaria, M., and Raine, R. (2016). Socio-economic inequalities in health care in england. *Fiscal Studies*, 37:371–403.

Currie, J. and MacLeod, W. (2008). First do no harm?: Tort reform and birth outcomes. *Quarterly Journal of Economics*, 123(2):795–830.

Cutler, D. M., Huckman, R. S., and Landrum, M. B. (2004). The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery. *American Economic Review*, 94(2):342–346.

Declercq, E. R., Sakala, C., Corry, M. P., and Applebaum, S. (2006). Listening to Mothers II: Report of the Second National U.S. Survey of Women's Chilbearing Experiences. *New York: Childbirth Connection*.

Department of Health (1993). Changing childbirth. report of the expert maternity group. Technical report.

Dranove, D. and Jin, G. Z. (2010). Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature*, 48:4:935–963.

Dranove, D., Kessler, D., McClellan, M., and Satterthwaite, M. (2003). Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy*, 111(3)(3):555–588.

Dranove, D. and Sfekas, A. (2008). Start spreading the news: A structural estimate of the effects of New York hospital report cards. *Journal of Health Economics*, 27(5):1201–1207.

Eysenbach, G., Powell, J., Kuss, O., and Sa, E. R. (2002). Empirical Studies Assessing the Quality of Health Information for Consumers on the World Wide Web: A Systematic Review. *JAMA*, 287(20):2691–2700.

Faber, B., Sanchis-Guarner, R., and Weinhardt, F. (2015). Ict and education: evidence from student home addresses. *NBER Working Paper*, (21306).

Falck, O., Gold, R., and Heblich, S. (2014). E-lections: Voting behavior and the internet. *The American Economic Review*, 104(7):2238–2265.

Fox, S. and Duggan, M. (2013). Health online 2013. Pew Internet & American Life Project.

Frakes, M. (2013). The Impact of Medical Liability Standards on Regional Variations in Physician Behavior: Evidence from the Adoption of National-Standard Rules. *The American Economic Review*, 103(1):257–76.

Freeman, M., Savva, N., and Scholtes, S. (2016). Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science*.

Gavazza, A., Nardotto, M., and Valletti, T. (2017). Internet and politics: Evidence from uk local elections and local government policies. *Review of Economic Studies*, Forthcoming.

Gaynor, M., Moreno-Serra, R., and Propper, C. (2013). Death by market power: Reform, competition, and patient outcomes in the national health service. *American Economic Journal: Economic Policy*, 5(4):134–66.

Gaynor, M., Propper, C., and Seiler, S. (2016). Free to Choose? Reform, Choice and Consideration Sets in the English National Health Service. *American Economic Review*, 106(11):3251–3257.

Goldfarb, A. and Prince, J. (2008). Internet Adoption and Usage Patterns are Different: Implications for the Digital Divide. *Information Economics and Policy*, 20(1):2–15.

Gravelle, H. and Sivey, P. (2010). Imperfect information in a quality-competitive hospital market. *Journal of Health Economics*, 29(4):524–535.

Gutacker, N., Siciliani, L., Moscelli, G., and Gravelle, H. (2016). Choice of hospital: Which type of quality matters? *Journal of Health Economics*, 50:230–246.

Hartzband, P. and Groopman, J. (2010). Untangling the web - patients, doctors, and the internet. *New England Journal of Medicine*, 362(12):1063–1066.

Ho, K. (2006). The welfare effects of restricted hospital choice. *Journal of Applied Econometrics*, 21:1039–1079.

Ho, K. and Pakes, A. (2014). Hospital choices, hospital prices, and financial incentives to physicians. *The American Economic Review*, 104(12):3841–3884.

Hofberg, K. and Brockington, I. F. (2000). Tokophobia: an unreasoning dread of childbirth. *The British Journal of Psychiatry*, 176(1):83–85.

House of Commons (2003). Choice in maternity services - ninth report of session 2002-03, volume i. Technical report, House of Commons Health Committee.

Kolko, J. (2010). How broadband changes online and offline behaviors. *Information Economics and Policy*, 22:144–152.

Lavender, T., Hofmeyr, G. J., Neilson, J. P., Kingdon, C., and Gyte, G. M. (2012). Caesarean section for non-medical reasons at term. *The Cochrane Database of Systematic Reviews*, 3.

Nama, V. and Wilcock, F. (2011). Caesarean section on maternal request: is justification necessary? *The Obstetrician & Gynaecologist*, 13(4):263–269.

Nardotto, M., Valletti, T., and Verboven, F. (2015). Unbundling the incumbent: Evidence from uk broadband. *Journal of the European Economic Association*, 13(2):330–362.

National Audit Office (NAO) (2013). Maternity services in england. Technical report.

National Institutes of Health (2006). NIH State-of-the-Science Conference Statement on Cesarean Delivery on Maternal Request. Technical Report 23(1).

NICE (NAtional Institute for Clinical Excellence) (2004). Cesarean section. clinical guideline 13. Technical report.

Ofcom (2011). Uk fixed-line broadband performance - the performance of fixed-line broadband delivered to uk residential consumers.

Olafsdóttir, T., Ásgeirsdóttir, T. L., and Norton, E. C. (2017). Valuing pain using the subjective well-being method. *NBER Working Paper Series*.

Paterson-Brown, S., Amu, O., Rajendran, S., and Bolaji, I. I. (1998). Should doctors perform an elective caesarean section on request? *BMJ*, 317(7156):462–465.

Phelps, C. E. (1992). Diffusion of Information in Medical Care. *The Journal of Economic Perspectives*, 6(3):23–42.

Price, J. and Simon, K. (2009). Patient education and the response to new medical research. *Journal of Health Economics*, 28(6):1166–1174.

Redshaw, M. and Heikkila, K. (2010). Delivered with care: a national survey of women's experience of maternity care 2010. Technical report, The National Perinatal Epidemiology Unit, University of Oxford.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Scott Morton, F., Zettelmeyer, F., and Silva-Risso, J. (2003). Consumer Information and Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities? *Quantitative Marketing and Economics*, 1(1):65–92.

Shurtz, I. (2014). Malpractice law, physicians' financial incentives, and medical treatment: How do they interact? *Journal of Law and Economics*, 57(1):1–29.

Spiegler, R. (2006). The market for quacks. *The Review of Economic Studies*, 73(4):1113–1131.

Thomas, J. and Paranjothy, S. (2001). *National Sentinel Caesarean Section Audit Report.* Royal College of Obstetricians and Gynaecologists - Clinical Effectiveness Support Unit RCOG Press.

Wagner, T. H., Bundorf, M. K., Singer, S. J., and Baker, L. C. (2005). Free Internet Access, the Digital Divide, and Health Information. *Medical Care*, 4(415-20).

Weaver, J. and Magill-Cuerden, J. (2013). "too posh to push": The rise and rise of a catchphrase. *Birth*, 40(4):264–271.

Wiklund, I., Edman, G., and Andolf, E. (2007). Cesarean section on maternal request: reasons for the request, self-estimated health, expectations, experience of birth and signs of depression among first-time mothers. *Acta Obstetricia et Gynecologica Scandinavica*, 86(4):451–456.

World Health Organization (2015). Who statement on caesarean section rates. Technical report.

Appendix

8.1 Hospitals' fixed effects

In Table 10 we report the estimates of the model in (1) for the probability to obtain a C-section, where we include hospital fixed effects as additional control variables. The estimated coefficients confirm the findings reported in Table 2.

Table 10: Regressions C-section model - hospitals' FEs.

Dependent variable: Probability of C-sec	All delivery types			Vaginal deliveries and Elective C-sections				Vaginal deliveries and Emergency C-sections
	All	Mult-time	First-time	All	Mult--time	First--time	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Did	0.49**	0.18	1.37***	0.34**	0.37	0.50*	0.63***	0.32
	(0.22)	(0.31)	(0.47)	(0.15)	(0.25)	(0.27)	(0.23)	(0.21)
Multiple × Did							-0.37	
							(0.25)	
Multiple-time							-2.21***	
							(0.15)	
LSOA F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mothers covs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Distance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hosp FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.304	0.432	0.223	0.479	0.569	0.367	0.523	0.191
Observations	522751	220519	135581	451531	195356	111448	306804	473989

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. There are missing values in the record of the variables multiple and first-time mothers (which are orthogonal to the treatment). Hence, the number of observations is smaller when these variables are used. Columns (1) to (3) use the sample of all mothers with the following differences: entire sample in column (1); multiple-time mothers in column (2); and first-time mothers in column (3). Columns (4) to (7) consider the sample of natural deliveries and elective C-sections with the following differences: all mothers in column (4); multiple-time mothers in column (5); first-time mothers in column (6); multiple and first-time mothers in column (7); and Column (8) considers the sample of natural deliveries and emergency C-sections.

8.2 Sensitivity of the treatment effect re-balancing the age of mothers.

We show in Table 1 that mothers in *close* and *far* LSOAs have a small, although statistically significant, difference in average age at delivery. In order to verify the consequences of this unbalance on our estimates we re-balanced the two samples of mothers through a propensity score matching procedure, as in Rosenbaum and Rubin (1983). With this procedure, we pair each mother living in a *close* LSOA with a mother with similar characteristics (among which there is the age) living in a *far* LSOA.

The re-balancing procedure has two consequences on our sample of mothers: the first one is to re-balance the age variable. On average, mothers living in *close* LSOAs give birth when they are 29.58 years old, while mothers living in *far* LSOAs give birth when they are 29.61 years old, and in this case we do not reject the null hypothesis of equal means. The second consequence is to restrict our sample to the matched mothers, which determines a reduction from 522,751 to 508,926 mothers (i.e., a loss of 2.6 percent of the sample).

We estimate the same model as in equation (1) on this restricted sample. Table 11 shows the estimated coefficients which are the equivalent to those reported in Table 2.

We do not find any relevant difference in estimated coefficients after re-balancing. The effect of better internet access estimated by the *Did* variable in Column (1) of Table 11 is 0.44, i.e., the same as the corresponding coefficient in Table 2. Other coefficients move only slightly, and in all cases within the standard errors of the corresponding coefficient in Table 2.

Table 11: Regressions C-section model - re-balancing mothers' age.

Dependent variable: Probability of C-section								
	All delivery types			Vaginal deliveries and Elective C-sections				Vaginal deliveries and Emergency C-sections
	All	Mult-time	First-time	All	Mult--time	First--time	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Did	0.44**	0.20	1.41***	0.33**	0.40	0.55**	0.59**	0.27
	(0.22)	(0.32)	(0.48)	(0.15)	(0.25)	(0.28)	(0.24)	(0.22)
Multiple × Did							-0.28	
							(0.25)	
Multiple-time							-2.28***	
							(0.15)	
Mothers covs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LSOA F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.301	0.428	0.219	0.477	0.566	0.364	0.521	0.189
Observations	508926	215273	131390	439390	190659	107899	298558	461322

Notes: The standard errors in parentheses are clustered at the level of the matched LSOAs to control for correlation between the neighboring LSOAs. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. There are missing values in the record of the variables multiple and first-time mothers (which are orthogonal to the treatment). Hence, the number of observations is smaller when these variables are used. Columns (1) to (3) use the sample of all mothers with the following differences: entire sample in column (1); multiple-time mothers in column (2); and first-time mothers in column (3). Columns (4) to (7) consider the sample of natural deliveries and elective C-sections with the following differences: all mothers in column (4); multiple-time mothers in column (5); first-time mothers in column (6); multiple and first-time mothers in column (7); and Column (8) considers the sample of natural deliveries and emergency C-sections.