Targeting Precision Medicine: Evidence from Prenatal Screening

Peter Conner, Liran Einav, Amy Finkelstein, Petra Persson, and Heidi Williams*

March 3, 2022

PRELIMINARY AND INCOMPLETE - PLEASE DO NOT CITE OR CIRCULATE

Abstract

Technological advances increasingly offer opportunities to target medical care to patients identified through up-front screening, which raises the question of how broadly or narrowly to screen for potential patients. We explore this trade-off empirically in the context of a new, non-invasive prenatal screen (cfDNA) that can be used to target a follow-up invasive test. We use Swedish administrative data on screening and testing choices for pregnancies conceived between 2011 and 2019. Each pregnancy receives an initial, low-cost non-invasive screen that delivers a predicted risk of fetal chromosomal abnormality. During our study period, cfDNA is introduced, which provides more precise information about risk but is also more costly than the initial screen. We find that insurance coverage for cfDNA dramatically increases its use and lowers rates of subsequent invasive testing, which is about twice as costly and creates an elevated risk of miscarriage. To assess the impact of counterfactual targeting of cfDNA coverage, we develop a stylized model of prenatal testing choices and estimate it based on the choices observed in the data. We find that narrow targeting of cfDNA screening has the (rare) potential to improve outcomes and reduce costs, while broader screening still improves outcomes but also raises costs. These findings point to the potential gains from well-designed targeting of screening, but at the same time highlight the importance of the targeting design.

*Conner: Karolinska Institutet, peter.conner@ki.se; Einav: Stanford University and NBER, leinav@stanford.edu; Finkelstein: MIT and NBER, afink@mit.edu; Persson: Stanford University, Research Institute of Industrial Economics, and NBER, perssonp@stanford.edu; Williams: Stanford University and NBER, hlwill@stanford.edu. We are grateful to Michaela Granfors, Kerstin Petersson, and Jonas Söderling at the Pregnancy Register. We thank Kelsey Moran, Jasmin Moshfegh, Katherine Moulton, Cameron Pfiffer, and especially Gabriel Swagel for excellent research assistance, as well as Sarah Bögl and Iliriana Shala for research assistance at the Research Institute of Industrial Economics.

1 Introduction

There is much interest in the potential for "Big Data" to transform all aspects of the economy, from retail to banking to marketing. Health care is no exception. Precision (or personalized) medicine – which targets treatments to patients based on their genetic, biological, or clinical characteristics – has been widely heralded for its potential to transform both the practice of medicine and the economics of health care. By identifying which patients are likely to benefit from certain treatments and – just as importantly – which are not, physicians can target effective treatments at the relevant sub-population, while sparing the rest the costs, side effects, and false hope of ineffective treatments.[1] Precision medicine thus dangles the tantalizing prospect of what has long been considered to be the holy grail of health care: improving patient health and well-being while simultaneously reducing health-care spending (Aspinall and Hamermesh 2007; Armstrong 2012).

This tantalizing prospect notwithstanding, it is less appreciated – and certainly less celebrated – that personalized medicine requires up-front screening to collect the information needed for personalization. Such information acquisition typically entails costs.[2] Ironically, therefore, it in turn raises questions about how finely to target the acquisition of information that is to be used for further targeting.

We study this issue in the context of prenatal testing, which accounts for the largest share of spending on genetic tests in the US (Phillips et al. 2018). For over half a century, invasive prenatal diagnostic tests - amniocentesis and chorionic villus sampling - have been able to diagnose fetal chromosomal abnormalities, but involve an elevated risk of miscarriage. Since the early 2000s, the development of non-invasive prenatal screening - nuchal translucency (NT) - has offered a lower-cost way to assess the likelihood of the most common chromosomal abnormalities without the risk of miscarriage, and thus inform the decision of whether to undertake invasive testing.

Within this context, we analyze the development of a second-generation form of non-invasive screening - known as cell-free DNA screning (cfDNA) - which substantially increases the informativeness of screening but also considerably elevates screening costs. We empirically explore the costs and benefits of how to target the availability of this second-generation

[1] Examples include targeted therapies for the treatment of breast cancer, lung cancer, and chronic myeloid leukemia.(Goldman et al. 2013; Aspinall and Hamermesh 2007; Berndt, Goldman, and Rowe 2018)

[2] Several recent papers have studied targeting of either screening or treatment technologies based on readily observable individual characteristics, such as age (Einav et al. 2020) or family medical history (Persson, Qiu, and Rossin-Slater 2021). Our analysis however focuses on information that is costly to acquire. Other examples under active discussion include possibly recommending mammogram screening based on breast density rather than (or in addition to) age (Trentham-Dietz et al. 2016) or using biomarkers and genetic testing to tailor cancer treatment (Banerjee et al. 2020).

screen through decisions of whether and when to cover the cost of this screening. This is currently an active – and evolving – policy debate in both the US and many European countries (Minear et al. 2015; Gadsbøll et al. 2020; GenomeWeb 2020).

For our analyses, we draw on detailed Swedish administrative data on prenatal screening and testing decisions for pregnancies conceived between 2011 and 2019. In the population we study, first-generation screening and invasive diagnostic tests are available for free for all pregnancies through our sample, as are downstream medical procedures. All pregnancies in our data receive the first generation non-invasive screen which provides a risk score: the predicted probability of a fetal chromosomal abnormality. It is used by parents to make further testing decisions and by the government to determine coverage for the second-generation screen. Crucially, we observe this risk score, as well as subsequent non-invasive and invasive testing decisions.

Our main focus is on the new cfDNA screening, which becomes available for free at different times in different regions of Sweden over our study period. These different regions also vary in how narrowly they define the set of risk scores that receive free cfDNA screening. Our data and setting thus provide a rare opportunity to empirically evaluate the impacts of targeting a new screening tool and to study the implications for optimal policy design.

We begin with descriptive evidence on testing choices in our setting. Testing rates, both non-invasive and invasive, rise with maternal age and with household income. Naturally, they also rise with the risk score. . Introducing insurance coverage for cfDNA screening substantially increases its use, and decreases use of subsequent invasive testing.

The benefits from covering cfDNA screening depend on the value of information to the parents. The costs depend on the impact of coverage on government spending, whose sign is a-priori ambiguous; cfDNA screening itself requires increased spending, but the screening also decreases spending on subsequent invasive testing, which is about twice as expensive as cfDNA screening. Naturally, both the benefits and costs from coverage will vary depending on the fetus' risk score, as well as parental preferences over the possible pregnancy outcomes.

To analyse outcomes and quantify welfare under alternative regimes we write down and estimate a stylized model of prenatal testing decisions. Each pregnancy has three possible outcomes: a live birth with no chromosomal abnormalities (the vast majority of births), a live birth with a chromosomal abnormality, or no live birth (due to miscarriage, abortion, or stillbirth). All pregnancies receive the initial screen and a risk score. Parents then face three sequential choices: whether to undergo CfDNA screening (which provides more precise information about risk), whether to conduct an invasive diagnostic test (which provides definitive information about chromosomal abnormalities but carries a miscarriage risk), and whether to terminate the pregnancy. The privately optimal decisions depends on the the

fetal risk score, the parents' relative utilities across possible pregnancy outcomes, and the out-of-pocket cost they face for cfDNA screening (when it is not covered).

We restrict our analysis to pregnancies with risk scores that are high enough to have the cfDNA screen covered under the most expansive coverage regime we observe. While this limits our analysis to about 14% of pregnancies in the data, they account (in expectation) for almost all (97%) pregnancies with chromosomal abnormalities. We estimate the model using method of moments, matching the cfDNA screening and invasive testing decisions for different risk scores and different coverage regimes. Despite the fairly stylized nature of the model, it fits the data remarkably well.

We estimate substantial heterogeneity in preferences over pregnancy outcomes, especially in the dis-utility of having a baby with chromosomal abnormalities. We estimate that under perfect information, just under one-tenth of the sample would prefer a live birth with a chromosomal abnormality to losing the baby.[3] We use the estimates to consider outcomes under alternative coverage regimes. We consider a variety of outcomes, including testing rates, birth outcomes, government spending, consumer surplus, and the rate of what we term (ex-post) inefficient outcomes. Inefficient outcomes occur in two forms in our setting: no live birth of a baby that had no chromosomal abnormalities, and live births of babies with chromosomal abnormalities born to parents who would have preferred no live birth.

In the absence of the new cfDNA screening technology, about 27% of the time the parents opt to do invasive testing, despite the miscarriage risk that comes with it. Spending per pregnancy on this testing is $336 US, and just over 0.3% of the pregnancies result in an (ex-post) inefficient outcome. If the new screening technology covers all pregnancies, cfDNA screening rates would be more than 80%, the rate of invasive testing would fall by about 80% to 6%, and the rate of inefficient outcomes would fall by about 90% to 0.03%. Consumer surplus would be about $250 higher per pregnancy, but government spending on screening and testing per pregnancy would rise by about $225 because the increased spending on cfDNA screening is not fully upset by the decreased spending on invasive testing. Thus, with broad coverage, the introduction of cfDNA screening would follow the pattern of most new medical technologies: improving patient welfare but also increasing health care spending.

By contrast, a more narrower targeting of coverage for the new cfDNA screening technology could both improve patient welfare and lower health care costs, compared to not having the new technology. For example, if - as is the case in several of the Swedish regions - the screening technology is covered for only the highest one-third of the risk scores in our

[3]In interpreting this result, it is important to note that our analysis is limited to the approximately three-fifths of pregnancies that undergo the initial non-invasive screening when it is freely available. Our estimates cannot speak to the preferences of the remaining two-fifths of the sample that do not do so, and thus do not receive a risk score.

sample, cfDNA screening rates would be less than 20%, and invasive testing would be only slightly over 5%. As a result, this more narrowly targeted cfDNA coverage both increases consumer surplus (by about $180 per pregnancy relative to the new technology not existing) *and* reduces government spending (by about $115 per pregnancy). Interesting, we also find that the most common coverage regime implemented in Sweden - turns out to be 'too clever by half.' Under this regime, cfDNA screening is only covered for the lowest half of the highest one-third risk scores (on the premise that the very highest-risk pregnancies would opt into invasive testing no matter what the cfDNA screen revealed); we find that produces lower consumer surplus and higher government spending than covering all of the highest one-third of risks scores.

Taken together, our findings illustrate both the promise and perils of screening technologies designed to guide the application of other medical interventions. Appropriately targeted, screening can provide that holy grail – improving patient well-being while saving money. However, if made more widely available, it becomes the more typical form of medical technology: raising both patient well-being as well as health-care spending.

Our findings relate to the large literature on both the promise and peril of medical innovation. In the U.S., it is both widely credited for dramatic secular improvements in health over the last half century, and widely blamed with dramatic increases in health care spending that have occurred over the same period (e.g. Newhouse (1992) and Cutler (2004)). In this context, precision medicine offers the potential to transform technologies that are highly effective for a subset of patients but end up cost-increasing because they are applied too broadly (so-called "type II" technologies in the framework developed by Chandra and Skinner (2012)) into technologies that are both highly health effective and highly cost effective, with little chance of overuse (their "type I" technologies). Whether or not this transformation can be realized is a matter of some debate (Phillips et al. 2014) . Our findings suggest that the ultimate impact of precision medicine is not dependent solely on the nature of the developed technologies, but also on public policies that regulate their access.

Our analysis therefore speaks to policy decisions - both in the U.S. and other countries - regarding coverage for new medical technologies.[4] In this vein, a few academic papers have analyzed the optimal pricing of alternative, more expensive, treatment options (e.g. Hirth, Chernew, and Orzol (2000), Einav, Finkelstein, and Williams (2016), and Hamilton et al. (2018)). Our paper complements this literature by analyzing pricing of a screening

[4]Countries employ different decision-making frameworks with some, like England, trying to employ a strict cost-effectiveness criteria, while others, like Sweden and the Netherlands, also take account of societal values such as the "principle of need" (Sabik and Lie 2008). In the US, Medicare is prohibited by law from considering costs in its coverage decisions, although many have argued that this is misguided (Chandra, Jena, and Skinner 2011)

technology that can be used to target treatments; in doing so, it contributes to a growing interest among economists in the potential impacts of precision medicine and its implications for optimal policy (e.g. Goldman et al. (2013) and Berndt, Goldman, and Rowe (2018)).An important distinction between analysis of optimal policy for therapeutic treatments in prior papers and our analysis for diagnostic screening is that for therapeutic treatments, the value of information is increasing in the probability that the patient will be found appropriate for treatment, while for screening, the value of information may be highest for mid-range risks. Indeed, we see in our setting that the willingness to pay for the new cfDNA screening is hump-shaped both with respect to risk and with respect to preferences .

More narrowly, our paper contributes to a literature at the intersection of medical innovation, family economics, and bioethics concerns over how health technologies that offer the promise of learning more about a pregnancy. Such technologies enable parents to act on their preferences over children's characteristics -– including, but not limited to, child health— but also raise concerns about "designer babies" and the potential eradication of certain traits in the population (Devlin 2019; Ball 2017; Hercher 2021). For example, the development of non-invasive prenatal screening has been accompanied by a substantial decline in the number of babies born with Down Syndrome, prompting an ethical debate over the possible "end of Down Syndrome," with an oft-cited statistic that 95% of fetuses who are prenatally diagnosed with the condition are aborted (Zhang 2020).[5] Another well-documented example is how the introduction of ultrasound technology in prenatal care — originally intended as a tool to monitor high-risk pregnancies (Hvistendahl 2021) — permitted parents with a preference over the child's gender to implement sex-selective abortion, and thereby contributed to a sex ratio imbalance involving over 160 million "missing females" in Asia and elsewhere (see, e.g., Sen (1990), Bongaarts and Guilmoto (2015), and Hvistendahl (2021)).

The rest of the paper proceeds as follows. The next section presents the setting and describes the data. Section 3 lays out the model and its estimation. Section 4 describes the results. The last section concludes.

[5]Interestingly, both our model and our results suggest that this statistic over-states parental preferences for terminating fetuses with Down Syndrome, as it only reflects the decisions of parents who choose to undergo diagnostic testing. We estimate that in our sample of pregnancies that receive an initial screen, expansive coverage of the new cfDNA screening would reduce the rate of live births with chromosomal abnormalities by about one-third.

2 Setting, Data, and Descriptive Statistics

2.1 Technological developments in prenatal testing

There has been remarkable progress over the last half-century in prenatal testing. Since the 1970s, invasive diagnostic procedures have been able to provide a definitive diagnosis of any fetal chromosomal abnormality. Amniocentesis, which was developed first, is typically done at 15-20 weeks of the pregnancy, and chorionic villus sampling (CVS), which came into use a couple of decades later, can be done as early as the 10th week (Akolekar et al. 2015). Both involve inserting a needle into the womb to extract fetal cells (from which fetal DNA is subsequently extracted) from the amniotic fluid (amniocentesis) or placenta (CVS). Because they are invasive, they carry a miscarriage risk – a risk that is estimated to be around 0.5% during our sample period (Akolekar et al. 2015).

More recently, from the 2000s on, advances in genetics have contributed to the development of *non*-invasive screenings. These screen for the three most common chromosomal abnormalities: Trisomy 21 (Down syndrome), Trisomy 13 (Patau syndrome), and Trisomy 18 (Edwards syndrome). Down syndrome, which causes mental retardation and structural deformations, is estimated to occur (in our empirical setting) in approximately 1 in 700 pregnancies that are carried at least 11 weeks, with the risk increasing with age (Conner and Malcus 2017). Patau and Edwards syndromes are 3 and 7 times less common, respectively, and are also more severe, with only 5-10% of all newborns surviving beyond the first year of their lives (Genetic and Rare Diseases Information Center (GARD) 2020; MedlinePlus 2021).[6] Compared to an invasive diagnostic test, non-invasive screening poses no risk of miscarriage and costs less, but it is also less informative about the presence of chromosomal abnormalities. In particular, unlike invasive testing, non-invasive screening only provides a risk assessment, not a definitive diagnosis.[7] Non-invasive screening is therefore typically used to inform decisions about whether to conduct a subsequent invasive test.

Within non-invasive screening there has been further technological progress. The first generation non-invasive screen – called nuchal translucency (NT) – uses information from an ultrasound and maternal blood work, together with maternal age and exact gestational age at screening to give a predicted probability of specific chromosomal abnormalities. Our analysis focuses on the second-generation of non-invasive prenatal screening – known as cell-free DNA screening (cfDNA); it is also commonly referred to as non-invasive prenatal testing

[6] More recently, screens have been developed for other, much rarer chromosomal abnormalities (Kliff and Bhatia 2022), but these are outside of the scope of our study.

[7] In addition, invasive testing can diagnose all chromosomal abnormalities, whereas non-invasive screening only provides information on the three most common described above, and - in the case of cfDNA screening, very rare abnormalities in gender chromosomes..

or screening. It was revolutionary in enabling analysis of fetal DNA without extracting fetal cells, and thus without any miscarriage risk. The screening involves a simple blood draw from the mother, from which the laboratory extracts fragments of the fetus' genetic material that are circulating in the mother's blood.

Like many new medical technologies compared to their existing counterparts, cfDNA screening is both more expensive and better than NT screening. It is about three times more expensive than NT screening - reflecting the higher expense of the lab work involved.[8] It is also more accurate in identifying pregnancies associated with a high risk for chromosomal abnormalities. In our setting, cfDNA screening has a 1% false positive rate and a 1% false negative rate for Trisomy 21 (Down syndrome), relative to NT rates of (respectively) 8% and 4%.[9]

Perhaps not surprisingly, therefore, there is no consensus on recommended practice. By the early 2000s, many countries in Europe (including Sweden) and several states in the United States had adopted a universal two-step prenatal testing program. These typically make the NT screen available for free or at a highly subsidized rate, with coverage for subsequent additional services – including provision of detailed information about the test result and free follow-up diagnostic testing – if the fetal risk score were above some threshold (Crombag et al. 2014; Flessel and Lorey 2011; Gifford et al. 2017).[10] With the advent of cfDNA screening, different countries' medical associations made different recommendations. For example, the American College of Obstetricians and Gynaecologists recommends that cfDNA screening be made available universally as a first step, instead of NT screening (GenomeWeb 2020). However Sweden's counterpart, the Svensk Förening för Obstetrik och Gynekologi (SFOG), instead recommends universal NT screening, followed by coverage for cfDNA screening for certain fetal risk scores (SFOG, 2016).[11] In both the US and Sweden, the recommending body recognizes that cfDNA screening dominates NT screening for predicting the likelihood of chromosomal abnormalities. However the Swedish recommendation - unlike the American one - also takes into account the significant cost difference between NT screening and cfDNA screening (SFOG, 2016). The cost of cfDNA screening relative to the information it provides

[8] In our setting costs run from about 1,500 SEK ($174) for NT screening to 5,000 SEK $580 for cfDNA, to 11,000 SEK ($1,275) for an invasive test (SBU 2016; Ingvoldstad Malmgren et al. 2017)).

[9] These rates are computed by converting NT risk scores to positive if they are higher than $\frac{1}{200}$ and negative otherwise. The accuracy advantage is similar for (the less common) Trisomy 13/18 (Patau or Edwards syndromes), with cfDNA false negative and false positive rates of 3% and 1% relative to 27% and 1% for NT (Conner and Malcus 2017).

[10] In some contexts, including Sweden, invasive testing is also covered for women who wished to go straight to such testing without doing the NT screen or even with a low NT score.

[11] Other European countries' recommendations are mixed, with many adopting a recommendation similar to the one in Sweden, but a few recommending universal cfDNA screening like the American approach (Gadsbøll et al. 2020).

is key for deciding how to design optimal coverage policy for cfDNA screening, and is the focus of our paper.

2.2 Swedish Policy environment

Sweden has universal and publicly financed health insurance, in which covered services are provided essentially for free, but there are some differences in covered services across its 27 regions.[12] At the start of our study period in 2011, virtually all regions provided free NT screening for all pregnancies. The SFOG recommended that those who receive a risk score of $\frac{1}{200}$ or higher be offered the opportunity to discuss follow-up invasive testing (SFOG, 2016), although such invasive testing was in principle available for free for all pregnancies, regardless of NT score or even whether they did NT screening.

With the development of the second-generation non-invasive screen, Sweden continued with its recommendation of universal NT screening, but now also recommended that that NT be followed in some cases by cfDNA screening. Specifically, in 2016, the SFOG issued the national recommendation to offer the cfDNA screen to women with NT risk scores that fall between $\frac{1}{1{,}000}$ and $\frac{1}{51}$, while the highest-risk pregnancies (risk greater than $\frac{1}{51}$) are recommended to "skip" cfDNA screening and obtain a definite diagnosis via invasive testing (SFOG, 2016).

The logic behind not offering cfDNA screening to low-risk pregnancies (below $\frac{1}{1{,}000}$) is intuitive: they are unlikely to have chromosomal abnormalities, so the cost effectiveness of (the more expensive) cfDNA screen is lower. The logic behind not offering cfDNA screening for high-risk pregnancies is more nuanced. It reflects an assumption that women receiving this risk score have a very high likelihood of doing an invasive test regardless of the cfDNA screening result. If this is indeed the case then incurring the (non-trivial) cost associated with cfDNA screening is considered a waste (SFOG, 2016).

The SFOG recommendations do not map directly into policy regarding insurance coverage.[13] In particular, in the first few years after the 2016 national recommendation, different regions made different choices about which pregnancies to cover for cfDNA screening. The most common policy regime covered cfDNA screening for NT risk scores between $\frac{1}{200}$ and

[12] Throughout we use the word "free" somewhat loosely. In practice, in some cases, covered services may require a small copay from the patient, but relative to the full cost of the service this copay is negligible.

[13] The same is true for the recommendations of medical authorities in the other countries. In the US, for example, the country's largest Medicaid managed care organization recently updated their coverage policy, making cfDNA screening available for free to enrollees across 24 states (ACOG, 2020), though far from all women are currently offered this screening. In Europe, the out-of-pocket price for cfDNA varies considerably, with different countries offering coverage for cfDNA screening for all pregnancies, so pregnancies, or pregnancies with certain risk scores (Gadsbøll et al. 2020).

$\frac{1}{51}$.[14] Other regions chose to cover risk scores above $\frac{1}{200}$, above $\frac{1}{1{,}000}$, or (in the case of one region) to follow the national recommendation and cover risk scores between $\frac{1}{1{,}000}$ and $\frac{1}{51}$.

2.3 Data and Sample Construction

We provide a brief overview of our data and variable definitions. Appendix A provides more detail on both, as well as on the construction of some of the statistics we report.

Data The backbone of our data is Sweden's *NT database* of pregnancies from 2011-2019, which is part of the Swedish Pregnancy Register. It contains prenatal testing data reported by clinics that perform approximately 80% of the NT screenings during this period; the selection of clinics into the database is based on which algorithm they use to compute the NT score.

The database only contains pregnancies that received an NT screen, which is performed at 11 through 14 weeks of pregnancy. Women who terminate their pregnancy prior to weeks 11 through 14, when the NT screening is performed, do not enter our sample. As more than 90% of all abortions are performed by the end of week 12 (Socialstyrelsen 2021), and as later abortions often are performed after the discovery of a risk of fetal abnormalities (Graviditetsregistret 2020), the set of pregnancies in our data likely consist predominantly of pregnancies where a child is desired.

For each pregnancy we observe the date of the NT screening, gestational age at screening, mother's age at the expected due date, the number of fetuses, and the region where the clinic is located. Crucially, we also observe the result of the NT screen: a risk score q for each fetus that is a cardinal measure of the probability of a chromosomal abnormality.[15] The risk score is censored from below at $\frac{1}{20{,}000}$ and from above at $\frac{1}{2}$. Finally, we observe whether the NT screening was followed by a subsequent cfDNA screening, and/or an invasive test.

We link the NT database to information about downstream pregnancy outcomes from the Medical Birth Register (MBR) on births from 2011 through 2019, and to information on maternal characteristics from several Swedish population registers. The MBR allows us to observe birth outcomes (live birth or stillbirth) and whether there are chromosomal

[14] This is similar to the nationally recommended coverage policy except that it uses $\frac{1}{200}$ as the lower bound for covering cfDNA screening instead of the nationally recommended $\frac{1}{1{,}000}$ lower bound. The regions presumably chose this higher risk score for the lower bound because the recommending body also states that its estimates suggest that offering cfDNA screening is only cost effective for pregnancies with a positive NT screening, i.e., with a risk of $\frac{1}{200}$ or higher (Ingvoldstad Malmgren et al. 2017).

[15] We define q as the maximum of the two risk scores that the screening produces per fetus: the estimated Down syndrome risk (q_1), and the estimated risk that the fetus has either Edward or Patau syndrome (q_2). The relevant risk is the risk of either syndrome, which is $q = q_1 + q_2 - q_1 q_2$, but since it is extremely rare for both risks to be meaningful, the approximation $q \approx \max\{q_1, q_2\}$ is used.

abnormalities for *all* pregnancies carried to 22 weeks of gestation or longer; if the pregnancy does not appear in the MBR, we code it as a termination (we cannot distinguish between miscarriages and abortions). The population registers allow us to measure mother's month and year of birth, her education, household income, and whether she is foreign born.

Sample construction We focus on pregnancies conceived from 2011 through early 2019, so that we can observe birth outcomes through the end of 2019. We limit the sample to the approximately 97% of all pregnancies that are single pregnancies. We also limit our analysis to the approximately 92% of pregnancies that occur in region-months that cover NT screening (see Table B.1). Take-up of NT screening is virtually non-existent in region-months where it is not covered, so in practice this imposes a negligible restriction on our core NT database. In region-months where it is covered, we estimate that about 60% of the pregnancies that survive to the NT gestational age receive NT screening.[16] Our results and their interpretation therefore apply to the (large) subset of women who choose to get NT screening when it is available for free.

After these restrictions, our data contain 297,249 pregnancies, carried by 235,706 unique women. For most analyses, we further restrict attention to the 14% of these pregnancies with a risk score q of $\frac{1}{1,000}$ or higher, which is the lowest risk score for which cfDNA screening is covered. We refer to this as our baseline sample; it includes 40,326 pregnancies, carried by 37,940 unique women. Although this sample excludes the majority of pregnancies, the variation in risk score in this low range is not very meaningful; indeed, using the distribution of risk scores in our data, we estimate that the pregnancies in our baseline sample cover 97% of chromosomal abnormalities.

2.4 Descriptive statistics

Table 1 presents summary statistics for three samples: all singleton pregnancies, the pregnancies in the baseline sample ($q > \frac{1}{1,000}$), and the approximately 70% of pregnancies in the baseline sample that are in regions that adopt the most common coverage regime - cfDNA coverage for risk scores between $\frac{1}{200}$ and $\frac{1}{50}$, which we will sometimes analyze separately. As would be expected given the relationship between maternal age and the risk of chromosomal abnormalities, the pregnant women in our baseline sample are older than women in the overall sample (maternal age of 35.7 vs. 32.6). They also have slightly higher average household income and levels of education. About one-third of our baseline sample does some subsequent

[16] In some region-months, all women are covered for NT screening, and then take-up is about 75%. In other region-months, only women who are 35 and older are covered for NT screening, and then take-up is about 40%.

(post-NT) testing, compared with about 5% of the overall singleton sample. Pregnancies in our baseline sample are also more likely to result in no live birth ((8.5% vs. 3.3%). Despite higher rates of testing and termination, women in our baseline sample are also more likely to give birth to a child with a chromosomal abnormality (0.4% vs 0.01%).

Figure 1 shows the distribution of pregnancies and the rate of any post-NT testing (cfDNA screening and/or invasive testing by the NT risk score). We the distributions for all singleton pregnancies (panels a and c) and for pregnancies in our baseline sample (panels b and c). The sample excluded from our baseline sample (i.e.those with a risk score below $\frac{1}{1{,}000}$) has extremely low (virtually zero) post-NT testing. Rates of post-NT testing also rise sharply at the risk score if $\frac{1}{200}$. As noted in Section 2 risk scores above $\frac{1}{200}$ are considered "positive" and throughout our study period, the SFOG recommended that those who receive a risk score above $\frac{1}{200}$ be offered the opportunity to discuss follow-up testing.

The risk score from the NT screen is a key predictor of further prenatal testing, but other factors such as maternal age and income may also play a role. To investigate this, Figure 2 shows take-up rates of further prenatal testing by mother's age and household income rank, among all singleton pregnancies.[17] Rates of further prenatal testing rise sharply with age (Panel (a)), but this appears to be driven entirely by differences in risk score. Panel (b) shows that conditioning on the NT risk score completely eliminates the testing age gradient; this is not surprising since the algorithm determining the NT risk score takes maternal age into account. Rates of further prenatal testing are also higher at the upper ends of the income distribution (Panel (c)). Conditioning on NT risk score eliminates some, but not all of this income gradient (Panel (d)); the residual income gradient is somewhat more pronounced for non-invasive testing, which is not always subsidized by public insurance.

3 An empirical model of prenatal testing choices

We write down a simple dynamic model of prenatal testing decisions that describes the choice problem(s) parents face. We then estimate the model using data on their decisions, and use the estimates to quantify the implications of counterfactual coverage policies regarding cfDNA screening on birth outcomes, public spending, and consumer surplus.

All pregnancies receive an NT screen and the resultant risk score; parents then face decisions about whether to do additional (cfDNA) screening (if it is available), whether to do invasive testing, and whether to terminate the pregnancy. The key counterfactuals we examine are the introduction of cfDNA screening (i.e. technological change) and policies

[17] Mother's age is measured at the due date. Household income rank is calculated as the within-year rank across the entire Swedish population.

regarding the coverage of cfDNA screening (i.e. whether patients must pay out of pocket or it is covered by the government and free to the patient). Note that our analysis of cfDNA coverage occurs in a setting in which all other medical choices (such as invasive testing and termination) are available for free. The policy design choice thus mirrors one public insurance systems frequently make regarding coverage of a new technology in the context of an existing insurance system with its pre-existing price distortions that are taken as given.

3.1 Model

Setting and notation. The unit of observation is a pregnancy, denoted by i, which is associated with three possible outcomes: a birth of a healthy baby, a birth of a baby with chromosomal abnormalities, or no live birth (due to miscarriage, abortion, or stillbirth). We normalize the utility of the expecting woman from a healthy baby to zero, and denote the (monetized) utility from losing the baby by a_i and the (monetized) utility from having a baby with a chromosomal abnormalities by c_i.

Throughout, we assume that $a_i < 0$ and that $c_i \leq 0$. Assuming that $a_i < 0$ is natural. It implies that all the pregnancies in our sample are desired pregnancies, which as discussed in Section 2 seems likely. Assuming that $c_i \leq 0$ is inconsequential; as will be clear below, decisions of women with $c_i > 0$ (that is, women who prefer a baby with chromosomal abnormalities over a healthy baby) will be identical (in the model) to what a woman with $c_i = 0$ would decide.

We denote by q_i the risk score that all pregnancies receive from the NT screening, and assume that it is known to all expecting women in our sample. We assume that q_i provides an unbiased prediction of the probability that pregnancy i carries a fetus with chromosomal abnormalities. This assumption is critical for our subsequent analyses since it allows us to simulate counterfactual pregnancy outcomes. It is also a realistic assumption as the risk score in our data is produced by an algorithm that is calibrated off of the Swedish pregnancies in our database.[18] Finally, we assume that all women are risk neutral and maximize their expected monetized utility.

Consistent with the institutional setting, we assume that both termination and invasive testing are free to the patient, and that invasive testing is associated with a (known) miscarriage risk g^{Inv}.[19] We denote the out-of-pocket cost of cfDNA screening as f_i; this will

[18] See Kublickas, Crossley, and Aitken (2009) for more information.

[19] We abstract from miscarriages that occur without being induced by invasive testing. While such spontaneous miscarriages are very common in the beginning of pregnancy, recall that our sample only includes pregnancies that are carried at least 11-14 weeks, when the NT screening is done; among such pregnancies, the miscarriage rate (which includes both miscarriages induced by invasive testing and spontaneous miscarriages) is lower than 2 percent (Oster 2014).

depend on the NT risk score and on the coverage regime. For expositional clarity, we omit the i subscript for the rest of this section.

Invasive testing decision in the absence of cfDNA screening. We first consider the invasive testing decision in a world where cfDNA screening does not exist; this captures the technology available at the start of our sample period. In this case, the patient only needs to make two sequential binary decisions. In period 1, she decides whether to do an invasive test or not; if she does the test, she receives the result. In period 2, she decides whether to terminate the pregnancy as a function of the test result (if she did it). We can then solve the model backwards to derive optimal choices.

The patient's expected utility in period 2 is given by (recall, we omit i subscripts for expositional clarity):

$$v_2(p) = max\{a, pc\}$$

where p is her belief about the probability her fetus has chromosomal abnormalities. Intuitively, she faces a decision between abortion (which yields utility of a) and having a baby, which has a p probability of having chromosomal abnormalities, thus yielding expected utility of pc (recall that the utility from a baby with no chromosomal abnormality is normalized to 0).

In period 1, the patient chooses whether or not to have an invasive test. Without it, her expected utility is defined by $v_2(p)$ in equation 3.1. If she does an invasive test, the (definitive) test result implies either $p = 0$ (if negative) or $p = 1$ (is positive). Taken together, and accounting for the miscarriage risk associated with invasive testing, expected utility from taking an invasive test is therefore given by:

$$v_{1|inv}(p) = g^{Inv}a + (1 - g^{Inv})(pv_2(1) + (1 - p)v_2(0)) = g^{Inv}a + (1 - g^{Inv})p \max\{a, c\},$$

where p is the probability of fetal chromosomal abnormalities.[20] The expected utility from an invasive test thus factors in the risk of miscarriage g^{Inv} and its associated disutility a, as well as - in the absence of miscarriage with occurs with probability $(1 - g^{Inv})$ - the utility from the patient's optimal choice $max\{a, c\}$ if the fetus is found to have a chromosomal abnormality, which occurs with probability p .

Thus, in period 1, the patient chooses an invasive test if and only if

$$v_{1|inv}(p) > v_2(p).$$

[20] In the absence of cfDNA screening, $p = q$ (as we are assuming) but we keep the notation general so that it will still apply when we introduce cfDNA screening below.

This optimal decision rule reflects a trade-off: the test provides information about the presence of chromosomal abnormalities but entails a risk of losing the pregnancy. The woman must trade off the risk of a miscarriage with her preference for (not) having a child with chromosomal abnormalities, equipped only with a noisy estimate of the fetus' underlying risk.

To gain intuition, consider first a case where $c > a$, so that the patient prefers having a baby with chromosomal abnormalities to losing the baby. In this case, she would not abort a baby with a confirmed chromosomal abnormality, so she has no benefit from invasive testing for any value of p.[21] In contrast, a woman with $c < a$ strictly prefers to abort a baby with a confirmed chromosomal abnormalities, so there are potential benefits from additional information. Under the optimal decision rule, she chooses to do invasive testing when her preference for abortion relative to giving birth to a baby with chromosomal abnormalities is strong enough relative to her risk, p. Note that under our assumptions, all women who conduct an invasive test and detect the presence of chromosomal abnormalities will terminate the pregnancy; this is consistent with the evidence in our setting that pregnancies are terminated in 99% of the cases in which invasive testing reveals a chromosomal abnormality in our setting (Conner et al. 2012).

Testing decisions in the presence of cfDNA screening. Once cfDNA screening exists, the patient can choose whether to do cfDNA screening after observing her NT screening result q, but before making the invasive testing decision. We thus add an initial, "period 0," binary decision regarding cfDNA screening. The cfDNA screening generates a binary result (positive or negative), with a false positive rate of g^{FP} and a false negative rate of g^{FN}. Recall that cfDNA screening carries no miscarriage risk.

Figure 4 illustrates the decision tree in this expanded setting. We can again solve for optimal decisions backwards, by appending the period 0 decision regarding cfDNA screening to the earlier decision (regarding invasive testing) presented above.

It is convenient to denote by $v_1(p) = \max\{v_{1|inv}(p), v_2(p)\}$, which is the period 1 continuation value of the patient, conditional on having a probability p of having fetal chromosomal abnormalities, where $v_{1|inv}(p)$ and $v_2(p)$ were defined in equations 3.1 and 3.1 respectively. If the patient chooses no cfDNA screening, then $p = q$ and her expected utility is given by $v_1(q)$.

If she chooses cfDNA screening in period 0, the result provides additional information about the risk of chromosomal abnormality. We assume Bayesian updating, and denote the

[21] In practice, knowledge of a chromosomal abnormality may be useful even for a couple who wishes to keep the baby, as it may help in preparing for the baby's arrival. We abstract from this in the model.

patient's posterior (after cfDNA screening) by $p(q,+)$ and $p(q,-)$ for, respectively, positive and negative cfDNA screening results.Naturally, the Bayesian updating depends on the precision of the cfDNA screening as measured by the false positive rate g^{FP} and the false negative rate g^{FN} [22] Her expected utility from cfDNA screening is therefore given by:

$$v_{0|screen}(q) = Pr(+|q)v_1(p(q,+)) + Pr(-|q)v_1(p(q,-)) - f,$$

where $Pr(+|q) = q(1-g^{FN}) + (1-q)g^{FP}$ and $Pr(-|q) = qg^{FN} + (1-q)(1-g^{FP})$ are the ex-ante probabilities (that depend on q) of the two potential outcomes of the cfDNA test, and the last component, f, is the out-of-pocket cost of cfDNA screening (which, as noted depends on q and on the cfDNA coverage regime).

Given the expected utility from cfDNA screening in equation 3.1, the patient chooses cfDNA screening in period 0 if and only if

$$v_{0|screen}(q) > v_1(q).$$

The trade-off is simple: the cfDNA screening result allows the patient to base her invasive testing decision on more precise information about the fetal risk of chromosomal abnormality without harming the fetus, but for patients (depending on f) obtaining this information is not free.

We note that when cfDNA screening is free (i.e. $f = 0$) and the result is inconsequential to the patient's choice of whether to follow up with invasive testing, the patient will be indifferent between choosing cfDNA screening or not (that is, $v_{0|screen}(q) = v_1(q)$). We make the (tie-breaking) assumption that patients do not do cfDNA screening when it is inconsequential. As a result, the model implies that patients who do cfDNA screening follow up with invasive testing if and only if the cfDNA screening result is positive.

3.2 Key Data Moments: Decisions by risk score and policy regime

We will estimate the model by method of moments, with the key moments being patients' choices regarding cfDNA screening and invasive testing as a function of their NT risk score q and the coverage regime. To illustrate these key moments, Figure 3 plots testing decisions by NT risk q, separately before and after the introduction of the most common cfDNA screening coverage policy, which introduced coverage for cfDNA screening for risk scores $q \in (\frac{1}{200}, \frac{1}{51}]$ during 2016 and 2017 (depending on the region); slightly over 70% of our baseline sample is in regions who experience this regime (see B.2). As we move from left to right along the

[22] That is, $p(q,+) = \frac{q(1-g^{FN})}{q(1-g^{FN})+(1-q)g^{FP}}$ and $p(q,-) = \frac{qg^{FN}}{qg^{FN}+(1-q)(1-g^{FP})}$.

x-axis in each panel, the risk of chromosomal abnormality q rises from $\frac{1}{1{,}000}$ to 1.[23] The choices shown in the figures illustrate some of the features of decision making highlighted by the model as well as the sharp changes in behavior produced by coverage of screening.

We focus first on behavior when the cfDNA screen is not covered for anyone (gray dots). Several key patterns emerge. Both cfDNA screening rates (panel a) and invasive testing (panel b) are essentially zero for risk scores below $\frac{1}{200}$, and jump sharply at the $\frac{1}{200}$ threshold. The rate of cfDNA screening jumps about 20 percentage and the rate of invasive testing (which starts slowly increasing around $\frac{1}{300}$) jumps about 30 percentage points at the $\frac{1}{200}$ threshold. These jumps likely reflect the medical practice of describing pregnancies with an NT risk score greater than $\frac{1}{200}$ as a positive test result, and the SFOG recommendation that such patients be offered the opportunity to discuss further testing. As a result, there is a large increase n rates of invasive test (panel b) - which is always free for the patient - and, even when cfDNA screens are not covered, this seems to prompt some patients to pay about $580 out of pocket for the additional cfDNA screen (panel a). Interestingly, as the risk score increases further above $\frac{1}{200}$, rates of cfDNA screening fall, becoming close to zero again for the highest risk score bin (panel a), and mirroring this declining rate of cfDNA screening above $\frac{1}{200}$, rates of invasive testing rise (panel b), from about 50 percent at a risk score of $\frac{1}{200}$ to over 80 percent in the highest risk score. This pattern is consistent with the predictions of the model: if the patient will do invasive testing regardless of cfDNA screening result, there is no value to the cfDNA screening (and in the initial period, there is a cost to the parent of doing so); as risk scores rise above $\frac{1}{200}$, more and more parents find themselves in this situation and therefore choose to skip cfDNA screening and move directly to invasive testing. This non-monotonic value of information as a function of appropriateness for the 'treatment' (invasive testing) is, as we emphasized in the Introduction, quite different for therapeutic treatments where the value of information is increasing in the probability the patient will be found appropriate for treatment.

We next consider the change in testing that occurs coverage for when cfDNA screening is introduced for risk scores between $\frac{1}{200}$ and $\frac{1}{51}$ (black dots). As shown in panel (a), rates of cfDNA screening jump from about 20 percent at $p = \frac{1}{200}$ (gray dots) to 90 percent (black dots); they remain at that level until coverage ends at $p = \frac{1}{51}$ at which time they drop sharply to under 40 percent. [24] The pronounced, approximately 75 percentage point

[23] as noted, in practice the risk scores are censored so that all those above $\frac{1}{2}$ receive a risk score of $\frac{1}{2}$.

[24] Rates of cfDNA screening also rise slightly for risk scores outside of the covered range. This reflects a general time trend in increasing use of the new screening technology. This likely reflects growing awareness of and comfort with the technology among patients and practitioners. Some of this may be the natural rate of secular diffusion of a new technology, and some may be accelerated by the increase use of the technology in the covered risk score range.

increase in cfDNA testing in the covered range (see panel a) is mirrored (see panel b) by a pronounced, approximately 50 percentage point drop in the rate of invasive testing in the covered range, from about 60 percent when cfDNA screening was not covered (gray dots) to about 10 percent when it is (black dots); once cfDNA coverage ends at $q = \frac{1}{51}$, the rate of invasive testing jumps up sharply to over 70 percent (black dots).

Thus coverage of cfDNA screening in the risk score range of $\frac{1}{200}$ to $\frac{1}{51}$ switches the dominant form of testing; prior to cfDNA coverage, about 20 percnt of pregnancies received cfDNA screening and about 60 percent received invasive testing (with some overlap); afterwards, about 90 percent receive cfDNA screening and only 10 percent receive invasive testing. These changes in invasive testing rates for risk scores in which cfDNA screening is covered are once again consistent with the model, and illustrate the key value of the new screening technology: targeting follow-up testing where it is likely most valuable. Given the high false positive rate of the NT screening and the very high accuracy of the cfDNA screen, many of the pregnancies that were flagged as likely positive for chromosomal abnormality in the NT screening (i.e. risk score above $\frac{1}{200}$ will get a more accurate negative prediction from the cFDNA screen; equipped with that additional information, the results indicate that many patients choose to avoid the invasive test with its associated miscarriage risk.

Panel (c) combines the results of the previous two panels by showing the share of pregnancies that do *any* testing after the NT screening (i.e., either cfDNA screening or invasive testing). It shows that coverage of cfDNA screening for risk scores $q \in (\frac{1}{200}, \frac{1}{51}]$ increases the probability of any test from about 75 percent to almost 100 percent. This illustrates the empirical tension in covering cfDNA scsreens: by reducing invasive testing (panel b) it can lower total spending, but by increasing cfDNA testing (panel a) it can raise total spending. This suggests that the total spending impact of covering cfDNA is a-priori ambiguous, and also likely depends on how targeted the coverage is: coverage of risk scores with lower rates of invasive testing in the absence of cfDNA is less likely to be cost reducing than coverage for risk scores with higher rates of baseline invasive testing. Our model estimates are designed to quantify these types of trade-offs on the spending side, as well a quantify the benefits to patients. [25]

We will use these patterns of testing decisions (as well as similar patterns for the other regions that introduced coverage for cfDNA screening for different ranges of risk scores, as described in Table B.2) as key moments for our model estimation. Crucially, the risk score q will allow us to simulate the distribution of pregnancy outcomes - i.e. live birth without

[25] Panel c also shows that after the policy change, the share receiving any test falls slightly at $p = \frac{1}{51}$ from almost 100 percent to about 90 percent; this is due to the the sharp drop in cfDNA testing (panel a) and the sharp rise in invasive testing (panel b) when cfDNA coverage ends at that threshold.

chromosomal abnormality, live birth with chromosomal abnormality, and no live birth - under counterfactual decisions. Although we can also observe pregnancy outcomes directly (recall Table 1), we will not use them as moments in the estimation. Given our sample size, and that the vast majority of pregnancies result in a live birth with no chromosomal abnormalities, the risk score q - which recall was calibrated on our data - provides a more accurate estimate of pregnancy outcomes than our data where, for example, in our baseline sample of about 40,000 pregnancies, only 0.36 (i.e. about 144 pregnancies) result in a live birth with a chromosomal abnormality. [26]

3.3 Specification and estimation

Econometric specification. The key estimable object of interest is the joint distribution of a_i and c_i, which we assume to be drawn from a truncated (from above, at zero) bivariate normal distribution. That is,

$$\begin{pmatrix} a_i \\ c_i \end{pmatrix} \sim N\left(\begin{pmatrix} x_i'\beta^a \\ x_i'\beta^c \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_c \\ \rho\sigma_a\sigma_c & \sigma_c^2 \end{pmatrix} \middle| a_i < 0, c_i \leq 0 \right),$$

where x_i will be a vector of observable characteristics of pregnancy i (such as maternal age, household income, past pregnancies) and the β's, σ's, and ρ are parameters to be estimated.

The values of all other parameters of the model are calibrated based on the institutional setting described in Section 2. Specifically, we assume that the miscarriage risk associated with invasive testing is $g^{Inv} = 0.5\%$, that the false positive rate from cfDNA screening g^{FP} and false negative rate from cfDNA screening g^{FN} are both equal to 1%, and that these rates are all known to the patient. We set the medical cost of invasive testing and cfDNA screening at 11,000 SEK and 5,000 SEK, respectively. Costs of invasive testing are always born by the government. cfDNA screening costs are either born by the government or the patient, depending on the policy regime, so the patient's out of pocket cost of cfDNA screening (f) is either 0 or 5,000 SEK.

A final tweak to the model arises from the observed impact of a medical recommendation to take invasive testing. As we discussed in Section 2, patient with what is referred to as a "positive" NT screening result (that is, NT risk of $\frac{1}{200}$ or higher) receive more information about chromosomal abnormalities and are explicitly offered the opportunity to discuss follow-up testing. As we saw in Figure 3b (gray circles), prior to coverage for cfDNA screening, presumably because of this consultation, there is a large - approximately 40 percentage point

[26] To further illustrate this point, Appendix figure D.1 shows the analog of Figure 3 for birth outcomes. The rate of live births with chromosomal abnormalities (panel b) or no live birth (panel c) are too small to detect meaningful changes with the coverage regime.

- jump in the propensity of invasive testing around an NT score of $\frac{1}{200}$. This occurs even though invasive testing is free for everyone in the sample, both below and above $q = \frac{1}{200}$. In order to empirically account for this effect in the data in a way that is consistent with the model, we introduce one more parameter, $\psi \in (0, 1)$, and assume that for those who receive a risk score $q_i > \frac{1}{200}$, their belief about the probability the fetus has a chromosomal abnormality is q_i^{ψ} instead of q_i. In other words, we model the impact of the recommendation as equivalent to the patient revising upward toward 1 their post-NT screening prior about the probability of chromosomal abnormality. This "tweak," which we apply for the entire sample helps us fit much better the jump in invasive testing rates at $q = \frac{1}{200}$ prior to the introduction of coverage. After cfDNA coverage is introduced, the discontinuous jump at $\frac{1}{200}$ is a combination of the same consultation effect and the fact that cfDNA coverage changes discontinuously at at $q = \frac{1}{200}$.[27]

Estimation and identification. We estimate the model using method of moments, matching the propensity to do cfDNA screening and invasive testing for each NT risk bin, before and after the introduction of cfDNA screening and by different coverage policy regimes (which vary across regions). As in figure 3, we bin the data by risk score into 20 bins, each of width 50. As seen in Table B.2, there are four distinct cfDNA screening policy regimes in our analysis sample. Over 70% of our sample is in a region which adopts cfDNA coverage for $q \in (\frac{1}{200}, \frac{1}{50})$, and whose screening and testing decisions we plotted in figure 3. The remaining sample is roughly evenly split between three other policy regimes: covering cfDNA screening when $q > \frac{1}{201}$, when $q \in (\frac{1}{1{,}000}, \frac{1}{51})$, and when $q > \frac{1}{1{,}000}$.

Specifically, for each policy regime separately, we match (a) the share of pregnancies that received cfDNA screening by NT risk bin once cfDNA screening is covered; (b) the share of pregnancies that received invasive testing conditional on doing cfDNA screening (again, measured only once cfDNA is covered), by NT risk bin; (c) the share of pregnancies that received invasive testing conditional on not doing cfDNA screening, by NT risk bin. This last moment is also measurable in the pre-coverage regime. [28]

Thus, overall, we have 260 distinct testing moments that we try to match: 20 bins by 3 outcomes by 4 policy regimes, in addition to 20 bins of the rate of invasive testing prior to the introduction of coverage for cfDNA screening. To do so, we simulate testing decisions using the model and a given set of parameter values, and search for the parameters the

[27] While this adjustment to the model is only needed (from a model fit perspective) before the introduction of cfDNA, the definition of a "positive" NT screening remains constant over our analysis period, so we apply this adjustment throughout.

[28] We normalize cfDNA screening to zero in the pre-coverage regime, although as seen in Figure 3a, there is some cfDNA screening before it is covered.

minimize the distance between the observed moments and the simulated moments. We use a (standard) squared distance objective function and weight moments by the number of pregnancies within each moment. The optimization is done in two steps, first we run a global search, and then a derivative-based local search. Appendix C.1 provides more details.

To gain intuition for identification, we can consider different types of variation in the data. The recommendation parameter ψ is identified off the discontinuous jump in the propensity to take up invasive testing during the pre-coverage regime (recall that this jump was the motivation to include this parameter). All else equal, the decision to do invasive testing trades off the increased miscarriage risk against the value of obtaining additional information about the risk of the fetus having a chromosomal abnormality. Patients who want the baby regardless (that is $a_i < c_i$) prefer to avoid the miscarriage risk, while parents who get a large amount of dis-utility from having a baby with a chromosomal abnormality relative to no baby (i.e. c_i much less than a_i) do invasive testing for sure, so the propensity to do invasive testing identifies the relative importance of a_i and c_i. The extent to which the insurance coverage for cfDNA screening (which lowers the cost to the patient from 5,000 SEK to 0 SEK) increases the rate of screening identifies the (monetized) magnitude of these preferences, and the extent to which rates of cfDNA screening and invasive testing change with q_i identifies their relative variance and the correlation.

4 Results

4.1 Model fit and estimates

Figure 5 shows the fit of our model. It plots - as a function of the risk score q - various testing probabilities in the data and as predicted by the model, using the estimated parameters. Specifically, it plots by q the probability of an invasive test, both before and after the introduction of coverage for cfDNA screening (panels a and b respectively), and the probability of cfDNA screening after it is covered (panel c). We note that while we match the moments separately by policy regime, for expositional clarity Figure 5 aggregates the estimation moments across the different regimes.[29] We view the fit as quite encouraging, especially given the parsimonious model parameterization we use. [30]

[29] Appendix Figure XX presents a dis-aggregated version of this figure for the most common coverage regime which includes about 70% of our sample.

[30] In particular, note that Figure 5 plots the data and model predictions by bins of risk score with a width of 25, while we estimated the model by matching moments defined by bins of width 50. Within each bin of 50, therefore, we can examine the fit of unmatched moments, which is reassuring. For example, note in panel c that we match the non-monotonicity of cfDNA screening between the third to last and second to last bin, even though that is not a moment we matched in estimation.

Table 2 presents the parameter estimates. They imply that the average monetized dis-utility of losing the baby is approximately \$126,000, and that the average dis-utility of having a baby with a chromosomal abnormality is about \$199,000, or roughly 60% greater (in absolute value).[31] These averages mask a modest level of heterogeneity, particularly in the dis-utility from having a baby with a chromosomal abnormality. Taken together, the estimates imply (in results not shown) that 10.5% of the sample would prefer to have, under perfect information, a CA baby over losing the baby (that is, $c_i > a_i$). Finally, we estimate ψ to be 0.94, which creates the impact of the recommendation. For example, it implies that a woman with an NT score of $\frac{1}{200}$ (which is when the recommendation kicks in) behaves as if her score is in fact $\frac{1}{145}$.

To provide more intuition for these estimates, Figure 6a plots their implications for consumer willingness to pay for cfDNA screening as a function of q_i. We calculate willingness to pay as the difference in expected utility of the pregnancy outcome conditional on cfDNA screening and the expected utilty of the pregnancy outcome conditional on not receiving cfDNA scsreening. The figure illustrates the non-monotonicity of the value of information that we highlighted in the Introduction. For low-risk pregnancies ($q_i < \frac{1}{200}$), most women would not do invasive testing without the cfDNA screen, so the main value of cfDNA screening comes from detecting fetuses that very likely have chromosomal abnormalitiy, making it natural to have the willingness to pay monotonically increasing in q_i. In contrast, for higher-risk pregnancies, most women would do invasive testing absent cfDNA screening, so the value of cfDNA screening is from detecting pregnancies that do not require invasive testing (due to a negative NIPT result), and this value is (ex ante) more likely for women with lower q_i, which is why the willingness to pay is decreasing in q_i in this range.

We will use these model estimates to analyze outcomes under counterfactual scenarios. Specifically, we simulate pregnancies based on the observable variables in the data (including q_i) and the estimated parameters, and then use the model to generate decisions and outcomes. Appendix C.2 provides more details about these counterfactual calculations.

4.2 The value of information and technology

Our first set of counterfactuals uses the model estimates to explore the consequences - and value - of information about the fetus and the technologies that can provide that information. To do so, Table 3 presents outcomes under four scenarios. The first two – "no post-NT testing" (column (1)) and "first best" (column (2)) – are hypothetical worst-case and best-

[31] Recall that we estimate a joint normal distribution, which is truncated from above at zero for both a and c. However, the results makes this truncation quantitatively unimportant, with only 0.01% of the un-truncated distribution falling above the truncation point/s.

case benchmarks. The key comparison is between a world with only invasive testing (column (3)) and a world that also has cfDNA screening available as well (column (4)); for both we assume that any available technologies are covered by insurance, and therefore free to the patient.

In the "no post-NT testing" scenario (column 1), we assume that neither cfDNA screening nor invasive testing are available. Patients must therefore make the decision of whether or not to terminate the pregnancy using only the information provided by the NT screen of the probability q of fetal chromosomal abnormality and their preferences. In this hypothetical "worst case", we estimate that 1.26 out of 100 pregnancies would be actively terminated due to the risk of having a baby with chromosomal abnormalities, thus generating 98.74 live births out of 100 pregnancies.[32] All of the terminations occur among patients who prefer no live birth to a birth with a chromosomal abnormality (i.e. $a_i > c_i$). Of the 1.26 terminations, we estimate that 0.63 (50%) are inefficient: the fetus had no chromosomal abnormalities, so under perfect information the patient would have preferred not to terminate. A second type of inefficiency that arises in this "no post-NT testing" scenario is the live birth of babies with chromosomal abnormalities born to patients who would have preferred to terminate the pregnancy had they known; we estimate that there are 2.74 such occurrences per 100 pregnancies; these account for 2.8% of live births and 86% of live births with chromosomal abnormalities. We measure consumer surplus in this scenario using our model, and normalize it to zero so it can serve as a benchmark.

The hypothetical "first best" scenario in column 2 assumes that the patient knows for sure whether their fetus has a chromosomal abnormality or not. In other words, there exists some technology that perfectly identifies the fetus type without any miscarriage risk or cost (to the individual or to the government). No testing is done therefore, and patients only terminate the pregnancy if the fetus has a chromosomal abnormality and they prefer no live birth to a birth with a chromosomal abnormality (i.e. $a_i > c_i$). We estimate that 3.37 out of 100 pregnancies end in termination, and 0.43 out of 100 pregnancies result in a live birth with a chromosomal abnormality; these make up 0.4 percent of live births. There are no 'inefficient' outcomes (because there is perfect information), but there are fewer live births (96.63 compared to 98.74 in the no post-NT testing scenario in column 1). The avoidance of inefficient outcomes implies a large increase n consumer surplus relative to the no post-NT testing scenario in column (1), of approximately $2,800 per pregnancy, or approximately

[32]This baseline number is artificially much higher than the typical "industry figure" (of, for example, 62 live births per 100 pregnancies (see CDC (1999))) for two reasons. First, as described in Section 2, most pregnancies that do not result in a live birth end prior to weeks 11-14, and thus do not enter our sample. Second, recall that our model abstracts from miscarriages that are not induced by invasive testing (see footnote section 19).

$83,000 per "affected" pregnancy (i.e. the 3.37 (=2.74+0.63) per 100 pregnancies in column 1 that are inefficient). [33]

In column (3) we introduce return to the imperfect information world but, now introduce an option for invasive testing after receiving the NT risk score. We assume (as in the data) that invasive testing is fully covered for everyone, and is paid out of government budget (at a cost of 11,000 SEK or US $1,248.50 per test). From the patient perspective, the invasive test provides a definitive diagnosis on fetal chromosomal abnormality, but elevates the risk of miscarriage. We estimate that about 27% of patients elect to do an invasive test, with the probability of invasive testing increasing risk score. The introduction of the invasive testing technology has a big effect on outcomes, eliminating most of the inefficient outcomes that would have occurred in the "no post NT testing" scenario (column (1)). Of the 0.63 fetuses per 100 that do not have a chromosomal abnormality and are "mistakenly" aborted without testing (see column (1)), 80% of them now have a live birth; the remaining 0.12 per 100 (8%) have no live birth because of the miscarriage risk that accompanies invasive testing. Similarly, the invasive testing allows most women who prefer to abort a baby with a chromosomal abnormality to do so: 92% of the 2.74 babies with chromosomal abnormalities who are born to parents who would have preferred to have no live birth are now aborted (after they are identified as having a chromosomal abnormality by invasive testing), and only 0.23 (8%) of them are born (because their mothers preferred to avoid invasive testing miscarriage risk, so could not know that the fetus had a chromosomal abnormality). Overall, relative to no post-NT testing, invasive testing eliminates 90% of the inefficient outcomes from the no post-NT testing scenario (3.02 out of 3.37) and generates 89% ($2,484 out of $2,792) of the potential consumer surplus from the first best, at a cost of $336 (per pregnancy) to the government. Indeed, invasive testing revolutionized prenatal care when it was introduced in the 1970s, and likely had a bigger impact that we estimate here since it was introduced in a world where NT screening did not yet exist.

Finally, in column (4) we introduce the option of cfDNA screening prior to invasive testing. For now, we assume that cfDNA screening is paid for out of the government budget (at a cost to the government of 5,000 SEK or $567.50 per test) and is therefore available to the patient for free (we relax this assumption in the next section). Because cfDNA screening is free to the patient and has no risk associated with it, the vast majority of patients (85%) do the cfDNA screen. [34] Out of these 85 (per 100) pregnancies that do cfDNA screening,

[33] Note that our estimate that 3.8 (=3.37+0.43) out of 100 pregnancies have a chromosomal abnormality is occurring in the set of pregnancies with risk scores above $\frac{1}{1,000}$, which, as already discussed in Section 2 are only 14% of pregnancies but account for almost all (97%) of pregnancies with chromosomal abnormalities.

[34] The 15% of patients who do not do the cfDNA screening fall into three types: 10.5% of patients who prefer having a baby with chromosomal abnormality to no baby ($c_i > a_i$), who would naturally avoid the

2.84 have a positive result and therefore follow up with invasive testing (in addition to the 3.15% who go immediately to invasive testing without the cfDNA screen). We thus estimate that the introduction of (covered) cfDNA screens reduces the rate of invasive testing from 27% (column (3)) to 6% (column(4)). This decline in invasive testing in turn reduces by 85% (from 0.12 to 0.01 per 100 pregnancies) the rate of inefficient miscarriages from invasive testing (fetuses with no chromosomal abnormality that miscarry). Similarly, the improved targeting of the invasive testing (due to the information obtained from the cfDNA screen) reduces the rate of inefficient births (babies born with a chromosomal abnormality to patients who preferred no live birth) by 91% (from 0.23 per 100 pregnancies to 0.02 per 100 pregnancies). As it turns out, the quantitative impact of both effects is broadly similar, so the rate of live births is barely affected (96.74% in column 3 compared to 96.64% in column 4). There is a 32% reduction (18%, from 0.66 to 0.45) in the share of live births with chromosomal abnormalities. More importantly, almost all (95%) of live births with chromosomal abnormalities are born to patients who prefer this outcome to no birth.

Taken together, the introduction of cfDNA screening in addition to invasive testing (i.e. column 4 vs column 3) eliminates 88% of the inefficient outcomes (0.22 out of 0.25) and generates 91% ($\$281 = 2,765 - 2,484$ out of $\$308 = 2,792 - 2,484$) of the potential (remaining) consumer surplus. This comes at an incremental cost of \$224 (=\$560-\$336) (per pregnancy) to the government. In other words, full coverage of cfDNA screening reproduces the well-known impact of most technological progress in medicine: improved patient well-being and higher health care spending. We now turn to examining whether more targeted coverage of cfDNA screening can improve on this outcome.

4.3 Targeting insurance coverage for cfDNA screening

In considering the consequences of various targeting of insurance coverage for cfDNA screening, we take as given the coverage of invasive testing. As noted in the Introduction, this is a specific example of the types of coverage decisions that public insurance programs frequently have to make regarding coverage for new medical technologies. Table 4 presents the results. In columns (1) and (2) we reproduce the results from columns (3) and (4) in Table 3, which provides benchmarks for the extremes of no cfDNA technology (column 1) and full coverage of cfDNA (column 2). As already seen, full coverage improves consumer welfare but also increases government spending. In the remainder of the table, therefore we explore

miscarriage risk associated with invasive testing regardless of the cfDNA result, patients with high enough q that would make them want an invasive test even after a negative cfDNA result (due to the possibility of a false negative and a much greater a_i than c_i) (3.15%), and patients with low enough q (and sufficiently close values of a_i and c_i even though $a_i¿c_i$) that they would avoid invasive testing even after a positive cfDNA result (due to the possibility of false positive) (1.65%).

the implications of other coverage policies, either no coverage or partial coverage policies. To provide intuition behind the results in this table, Figure 7 plots, by NT risk score, the cfDNA screening rate (panel A) and the invasive testing rate (panel B) when that risk score is not covered for cfDNA screening (gray circles) and when it is covered (dark circles). For cfDNA screening, these rates correspond to the share of the population whose willingness to pay for cfDNA screening is above 0 (when there is full coverage) or above the out of pocket price (when there is no coverage). [35]

Column (3) introduces the cfDNA technology but provides no insurance coverage for it, so that individuals must pay the full cost of the cfDNA screen ($567.50) out of pocket. Rates of cfDNA screening drop to 18%, compared to 85% when it was free (column 2). Most (88%) of the women who now forgo cfDNA screening substitute to doing no testing at all, while the remaining substitute to invasive testing (which is free, but has a miscarriage risk). As a result, the vast majority of the 0.23 (per 100 pregnancies) inefficient live births (those with chromosomal abnormalities born to parents who would prefer no live birth) that occur without the cfDNA technology (in column 1) are not identified and therefore conntinue to occur (0.20 out of 0.23 or 87%), but the reduced invasive testing rate (10.8%, relative to 26% when cfDNA screening is not available) reduces by about two-thirds the rate of miscarriages among babies with no chromosomal abnormalities (0.04 instead of 0.12). A comparison of column (3) and column (1) shows that the introduction of the cfDNA screening - even without insurance coverage- is able to both increase consumer surplus (by about 0 per pregnancy) - and reduce government spending (by about 00 per pregnancy).

The last two columns consider the impact of more targeted coverage of cfDNA screening than coverage for all pregnancies[36] shown in column (2) or no pregnancies (shown in column (3)). For patients whose risk score falls within the covered range, cfDNA coverage is available for free; other patients can still access cfDNA screening but must pay its full cost out of pocket. Column (4) shows the results of providing coverage of cfDNA screening to patients whose risk score is $\frac{1}{200}$ or above, which is the risk score at which an NT score is labelled "positive". Compared to no coverage of cfDNA screening (column 3) there is an increase in cfDNA screening (from 17% to 27%) and a decrease in invasive testing (from 11% to 5.5%). This has two effects. It reduces the rate of live births with chromosomal abnormalities to parents who prefer no birth (from 0.20 to 0.13 per 100 pregnancies) because more people are doing cfDNA screening and detecting fetuses with chromosomal abnormalities, and it reduces the risk of miscarriage of babies without chromosomal abnormality (from 0.04 per 100 to 0.01

[35] For invasive, the interpretation is more subtle; we see that for low values of q, cfDNA coverage slightly increases invasive testing since the patient would not have found invasive testing worthwhile in the absence of cfDNA screening, but once that screen is free and is done, it occasionally triggers a positive.

[36] which given our sample restriction means all pregnancies with risk score above $[1\frac{}{1,000}$

per 100 pregnancies) because there is less invasive testing. This increases consumer surplus by \$137 per pregnancy (from \$2,523 to \$2,660) but also increases spending per pregnancy (from \$220 to \$135), because the decreased spending on invasive testing is not enough to offset the increased spending on cfDNA screening.

Perhaps most interestingly, the targeted coverage regime in column (4) both increases consumer surplus and lowers government spending relative to the absence of cfDNA technology (column 1), unlike full cfDNA coverage (column 2), which increases consumer surplus but also increases government health care spending. In other words, more narrowly targeting the coverage of cfDNA screening turns this new technology from one of Chandra and Skinner's common "type II" technologies to the more desirable - and rarer - "type 1" technologies. Of course, we saw in column 3 that merely introducing the technology without coverage also reduces spending while raising surplus, but targeted coverage increases surplus by \$137 more while raising covernment cost by only \$85, so for a reasonable cost of public funds (e.g. 15 percent), total welfare is higher with targeted coverage.

Finally, in column (5) we consider even more targeting, and assume that cfDNA screening is covered not for any $q_i \geq \frac{1}{200}$ as in column (4), but only for $q_i \in [\frac{1}{200}, \frac{1}{51}]$; this latter policy is the most common one in our data, covering about 70% of pregnancies (see Table B.2). The comparison of column (4) and column (5) is instructive. The motivation for the greater targeting in column (5) is the assumption that for the very highest risk pregnancies ($q_i \geq \frac{1}{50}$), patients are likely to do an invasive test regardless of the result of the cfDNA screening, so that incurring the (non-trival) cost associated with cfDNA screening is considered a waste. Recall that for this reason the Swedish recommendation is not to cover cfDNA screening in this highest risk range. However, the results in column 5 compared to column 4 suggest that, in practice, this assumption is flawed: cfDNA coverage of the very highest risk pregnancies (in column 4 compared to column 5) in fact causes substantial substitution from invasive testing to cfDNA screening; this results in fewer miscarriages of babies without chromosomal abnormality (0.01 per 100 pregnancies instead of 0.04) and higher consumer surplus (by about \$40 per pregnancy), while saving the government money (about \$16 per pregnancy).

Figure 8 summarizes these results visually, by representing the five scenarios presented in Table 4 using their impacts on consumer surplus and government spending. Welfare is higher to the north-east (higher consumer surplus and lower government spending). Introducing the cfDNA technology with full coverage both raises consumer surplus and government spending, the typical pattern for technological change in health care. By contrast, introducing cfDNA technology with either no coverage or partial coverage achieves the holy grail of lowering spending while simultaneously increasing consumer surplus. Within partial coverage regimes, covering risk scores above $\frac{1}{200}$ dominates (on both these dimensions) covering risk scores in

the range $\frac{1}{200}$, $\frac{1}{51}$. Moreover, for any reasonable levels of trade-off between taxpayer dollars and consumer surplus, covering risk scores above $\frac{1}{200}$ has higher welfare than either no cfDNA coverage or full cfDNA coverage.

Of course, there are many other possible ranges for partial coverage of cfDNA screening. Figure 9 therefore examines more granularly the optimal targeting of cfDNA coverage. We explore policies that cover cfDNA screening for all pregnancies with $q_i > x$, where x varies from $\frac{1}{1{,}000}$ (full coverage) to 1 (no coverage). We report both consumer surplus and government spending, and also the sum of them as an approximation of total welfare. . Total welfare is increasing in coverage until q_i is $\frac{1}{200}$, and then is roughly flat for a range, turning down only when coverage is limited to the highest risk pregnancies. This is not coincidental, as, recall, $\frac{1}{200}$ is when the NT test is considered positive, and the patient is referred to consultation, thus generating (in the model) a discontinuous impact on testing behavior. However, moving to more targeted coverage (from left to right) reduces consumer surplus (panel a), with the reduction coming primarily from an increase in live births with chromosomal abnormalities to parents who would prefer no birth (panel b). when considering

5 Conclusions

TBA

References

Akolekar, Ranjit, Jaroslaw Beta, Gemma Picciarelli, Caroline Ogilvie, and Francesko D'Antonio. 2015. "Procedure-Related Risk of Miscarriage Following Amniocentesis and Chorionic Villus Sampling: A Systematic Review and Meta-Analysis." *Ultrasound in Obstetrics & Gynecology* 45 (1): 16–26.

American College of Obstetricians and Gynecologists (ACOG). 2020. *ACOG Win! Major Payer Updates Coverage Guidelines to Align with ACOG NIPT Recommendation.* Accessed July 19, 2021. https://www.acog.org/news/news-articles/2020/10/acog-win-major-payer-updates-coverage-guidelines-to-align-with-acog-nipt-recommendation.

Armstrong, Katrina. 2012. "Genomics and Health Care Disparities: The Role of Statistical Discrimination." *JAMA* 308, no. 19 (November): 1979–1980. ISSN: 0098-7484. https://doi.org/10.1001/2012.jama.10820. eprint: https://jamanetwork.com/journals/jama/articlepdf/1389899/jvp120073_1979_1980.pdf. https://doi.org/10.1001/2012.jama.10820.

Aspinall, Mara G., and Richard G. Hamermesh. 2007. "Realizing the Promise of Personalized Medicine." Section: Competitive strategy, *Harvard Business Review* (October). ISSN: 0017-8012, accessed February 23, 2022. https://hbr.org/2007/10/realizing-the-promise-of-personalized-medicine.

Ball, Philip. 2017. "Designer babies: an ethical horror waiting to happen?" [In en-GB]. *The Guardian* (January). ISSN: 0029-7712, accessed February 24, 2022. https://www.theguardian.com/science/2017/jan/08/designer-babies-ethical-horror-waiting-to-happen.

Banerjee, Sudeep, Abhishek Kumar, Nicole Lopez, Beiqun Zhao, Chih-Min Tang, Mayra Yebra, Hyunho Yoon, James D. Murphy, and Jason K. Sicklick. 2020. "Cost-effectiveness Analysis of Genetic Testing and Tailored First-Line Therapy for Patients With Metastatic Gastrointestinal Stromal Tumors." *JAMA Network Open* 3, no. 9 (September): e2013565–e2013565. ISSN: 2574-3805, accessed February 9, 2022. https://doi.org/10.1001/jamanetworkopen.2020.13565. https://doi.org/10.1001/jamanetworkopen.2020.13565.

Berndt, Ernst R., Dana P. Goldman, and John W. Rowe. 2018. "Introduction to 'Economic Dimensions of Personalized and Precision Medicine'." In *Economic Dimensions of Personalized and Precision Medicine,* 1–7. University of Chicago Press, April. http://www.nber.org/chapters/c13986.

Bongaarts, John, and Christophe Guilmoto. 2015. "How many more missing women?" [In English]. Publisher: Elsevier, *The Lancet* 386, no. 9992 (August): 427. ISSN: 0140-6736, 1474-547X, accessed February 24, 2022. https://doi.org/10.1016/S0140-6736(15)61439-8. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)61439-8/fulltext.

CDC. 1999. *1999 Fact Sheet - Trends in Pregnancies and Pregnancy Rates* [in en-us], December. Accessed January 26, 2022. https://www.cdc.gov/nchs/pressroom/99facts/pregrate.htm.

Chandra, Amitabh, Anupam B. Jena, and Jonathan S. Skinner. 2011. "The Pragmatist's Guide to Comparative Effectiveness Research." *Journal of Economic Perspectives* 25, no. 2 (June): 27–46. https://doi.org/10.1257/jep.25.2.27. https://www.aeaweb.org/articles?id=10.1257/jep.25.2.27.

Chandra, Amitabh, and Jonathan Skinner. 2012. "Technology Growth and Expenditure Growth in Health Care." *Journal of Economic Literature* 50, no. 3 (July): 645–80. https://doi.org/10.1257/jel.50.3.645. https://www.aeaweb.org/articles?id=10.1257/jel.50.3.645.

Conner, Peter, and Peter Malcus. 2017. "Obstetriskt ultraljud och prenatal diagnostik i första trimestern." Läkartidningen. https://lakartidningen.se/klinik-och-vetenskap-1/artiklar-1/temaartikel/2017/08/obstetriskt-ultraljud-och-prenatal-diagnostik-i-forsta-trimestern.

Conner, Peter, Magnus Westgren, Anna Marsk, Sven Gustafsson, and Marius Kublickas. 2012. "Combined ultrasound and biochemistry for risk evaluation in the first trimester." Publisher: John Wiley & Sons, Ltd, *Acta Obstetricia et Gynecologica Scandinavica* 91, no. 1 (January): 34–38. ISSN: 0001-6349, accessed February 9, 2022. https://doi.org/10.1111/j.1600-0412.2011.01268.x. https://doi.org/10.1111/j.1600-0412.2011.01268.x.

Crombag, Neeltje M. T. H., Ynke E. Vellinga, Sandra A. Kluijfhout, Louise D. Bryant, Pat A. Ward, Rita Iedema-Kuiper, Peter C. J. I. Schielen, et al. 2014. "Explaining Variation in Down's Syndrome Screening Uptake: Comparing the Netherlands with England and Denmark Using Documentary Analysis and Expert Stakeholder Interviews." *BMC Health Services Research* 14 (1): 1–11.

Cutler, David M. 2004. *Your Money or Your Life: Strong Medicine for America's Health Care System.* New York, New York: Oxford University Press. http://www.amazon.com/Your-Money-Life-Medicine-Americas/dp/0195181328.

Devlin, Hannah. 2019. "IVF couples could be able to choose the 'smartest' embryo" [in en-GB]. *The Guardian* (May). ISSN: 0261-3077, accessed February 24, 2022. https://www.theguardian.com/society/2019/may/24/ivf-couples-could-be-able-to-choose-the-smartest-embryo.

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi Williams. 2020. "Screening and Selection: The Case of Mammograms." *American Economic Review* 110 (12): 3836–70.

Einav, Liran, Amy Finkelstein, and Heidi Williams. 2016. "Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments." *American Economic Journal: Economic Policy* 8, no. 1 (February): 52–79. https://doi.org/10.1257/pol.20140293. https://www.aeaweb.org/articles?id=10.1257/pol.20140293.

Flessel, Monica C., and Fred W. Lorey. 2011. "The California Prenatal Screening Program:"Options and Choices" not "Coercion and Eugenics"." *Genetics in Medicine* 13 (8): 711–713.

Gadsbøll, Kasper, Olav B. Petersen, Vincent Gatinois, Heather Strange, Bo Jacobsson, Ronald Wapner, Joris R. Vermeesch, The NIPT-map Study Group, and Ida Vogel. 2020. "Current use of noninvasive prenatal testing in Europe, Australia and the USA: A graphical presentation." Publisher: John Wiley & Sons, Ltd, *Acta Obstetricia et Gynecologica Scandinavica* 99, no. 6 (June): 722–730. ISSN: 0001-6349, accessed February 9, 2022. https://doi.org/10.1111/aogs.13841. https://doi.org/10.1111/aogs.13841.

Genetic and Rare Diseases Information Center (GARD). 2020. *Trisomy 13.* Accessed July 20, 2021. https://rarediseases.info.nih.gov/diseases/7341/trisomy-13.

GenomeWeb. 2020. *ACOG Guidelines Recommend NIPT for All Pregnancies Regardless of Risk.* Accessed July 19, 2021. https://www.genomeweb.com/molecular-diagnostics/acog-guidelines-recommend-nipt-all-pregnancies-regardless-risk#.YPIXVxNKiu4, August.

Gifford, Kathy, Jenna Walls, Usha Ranji, Alina Salganicoff, and Ivette Gomez. 2017. "Medicaid Coverage of Pregnancy and Perinatal Benefits: Results from a State Survey." Kaiser Family Foundation. https://www.kff.org/report-section/medicaid-coverage-of-pregnancy-and-perinatal-benefits-survey-results.

Goldman, Dana P., Charu Gupta, Eshan Vasudeva, Kostas Trakas, Ralph Riley, Darius Lakdawalla, David Agus, Neeraj Sood, Anupam B. Jena, and Tomas J. Philipson. 2013. "The Value of Diagnostic Testing in Personalized Medicine." *Forum for Health Economics and Policy* 16 (2): S87–S99. https://doi.org/doi:10.1515/fhep-2013-0023. https://doi.org/10.1515/fhep-2013-0023.

Graviditetsregistret. 2019. *Graviditetsregistrets Årsrapport 2019.* Accessed July 21. https://www.medscinet.com/GR/uploads/hemsida/dokumentarkiv/Graviditetsregistrets%20%C3%85rsrapport%202019_2.0.pdf.

———. 2020. *Graviditetsregistrets Årsrapport 2020.* Technical report. Stockholm, Sweden: Graviditetsregistret. Accessed October 19, 2021. https://www.medscinet.com/GR/uploads/hemsida/dokumentarkiv/GR%20%C3%85rsrapport%202020%203.0.pdf.

Hamilton, Barton H., Emily Jungheim, Brian McManus, and Juan Pantano. 2018. "Health Care Access, Costs, and Treatment Dynamics: Evidence from ¡em¿In Vitro¡/em¿ Fertilization." *American Economic Review* 108, no. 12 (December): 3725–77. https://doi.org/10.1257/aer.20161014. https://www.aeaweb.org/articles?id=10.1257/aer.20161014.

Hercher, Laura. 2021. *A New Era of Designer Babies May Be Based on Overhyped Science* [in en], July. Accessed February 24, 2022. https://www.scientificamerican.com/article/a-new-era-of-designer-babies-may-be-based-on-overhyped-science/.

Hirth, Richard A., Michael E. Chernew, and Sean M. Orzol. 2000. "Ownership, Competition, and the Adoption of New Technologies and Cost-Saving Practices in a Fixed-Price Environment." Publisher: Sage Publications, Inc. *Inquiry* 37 (3): 282–294. ISSN: 00469580, 19457243, accessed February 23, 2022. http://www.jstor.org/stable/29772902.

Hvistendahl, Mara. 2021. *How Ultrasound Changed the Human Sex Ratio* [in en], June. Accessed February 24, 2022. https://www.scientificamerican.com/article/how-ultrasound-changed-human-sex-ratio/.

Ingvoldstad Malmgren, Charlotta, Erik Iwarsson, Niklas Juth, and Peter Lindgren. 2017. "SFOG ger nationella riktlinjer för fosterdiagnostik med NIPT" [in sv-SE]. *Läkartidningen* (Stockholm, Sweden) (May). Accessed October 18, 2021.

Kliff, Sarah, and Aatish Bhatia. 2022. "When They Warn of Rare Disorders, These Prenatal Tests Are Usually Wrong" [in en-US]. *The New York Times* (January). ISSN: 0362-4331, accessed February 24, 2022. https://www.nytimes.com/2022/01/01/upshot/pregnancy-birth-genetic-testing.html.

Kublickas, Marius, Jennifer Crossley, and David Aitken. 2009. "Screening for Down's syndrome in the first trimester: Combined risk calculation, methodology, and validation of a web-based system." Publisher: John Wiley & Sons, Ltd, *Acta Obstetricia et Gynecologica Scandinavica* 88, no. 6 (June): 635–638. ISSN: 0001-6349, accessed November 3, 2021. https://doi.org/10.1080/00016340902859299. https://doi.org/10.1080/00016340902859299.

MedlinePlus. 2021. *Trisomy 18.* Accessed July 21, 2021. https://medlineplus.gov/genetics/condition/trisomy-18/.

Minear, Mollie A, Celine Lewis, Subarna Pradhan, and Subhashini Chandrasekharan. 2015. "Global perspectives on clinical adoption of NIPT" [in eng]. Edition: 2015/09/25, *Prenatal diagnosis* 35, no. 10 (October): 959–967. ISSN: 1097-0223. https://doi.org/10.1002/pd.4637. https://pubmed.ncbi.nlm.nih.gov/26085345.

Newhouse, Joseph P. 1992. "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives* 6, no. 3 (September): 3–21. https://doi.org/10.1257/jep.6.3.3. https://www.aeaweb.org/articles?id=10.1257/jep.6.3.3.

Oster, Emily. 2014. *Expecting Better: Why the Conventional Pregnancy Wisdom is Wrong–and what You Really Need to Know.* Vol. 1. Penguin.

Persson, Petra, Xinyao Qiu, and Maya Rossin-Slater. 2021. "Family Spillover Effects of Marginal Diagnoses: The Case of ADHD." https://www.nber.org/papers/w28334, *NBER Working Paper No. 28334.*

Phillips, Kathryn A., Julie Ann Sakowski, Julia Trosman, Michael P Douglas, Su-Ying Liang, and Peter Neumann. 2014. "The economic value of personalized medicine tests: what we know and what we need to know." *Genetics in Medicine* 16, no. 3 (March): 251–257. ISSN: 1530-0366. https://doi.org/10.1038/gim.2013.122. https://doi.org/10.1038/gim.2013.122.

Phillips, Kathryn A., Patricia A. Deverka, Gillian W. Hooker, and Michael P. Douglas. 2018. "Genetic Test Availability And Spending: Where Are We Now? Where Are We Going?" Publisher: Health Affairs, *Health Affairs* 37, no. 5 (May): 710–716. ISSN: 0278-2715, accessed February 23, 2022. https://doi.org/10.1377/hlthaff.2017.1427. https://www.healthaffairs.org/doi/10.1377/hlthaff.2017.1427.

Sabik, Lindsay M, and Reidar K Lie. 2008. "Priority setting in health care: Lessons from the experiences of eight countries" [in eng]. Publisher: BioMed Central, *International journal for equity in health* 7 (January): 4–4. ISSN: 1475-9276. https://doi.org/10.1186/1475-9276-7-4. https://pubmed.ncbi.nlm.nih.gov/18208617.

SBU. 2016. *Fosterdiagnostik med mikroarray för utökad analys av kromosomer.* SBU-rapport 246. Stockholm: Statens beredning för medicinsk och social utvärdering (SBU). Accessed November 15, 2021. https://www.sbu.se/contentassets/b7ed0fc762fe4d53ad97cceeb6e4633b/fosterdiagnostik_mikroarray_utokad_analys_kromosomer_2016.pdf.

Sen, Amartya. 1990. "More Than 100 Million Women Are Missing" [in en]. *The New York Review of Books* (December): 61–66.

SFOG (Svensk förening för Obstetrik och Gynekologi). 2016. *Analys av foster-DNA i kvinnans blod: icke-invasiv fosterdiagnostik (NIPT) för trisomi 13, 18 och 21. SFOG riktlinje 2016, framtagen av Ultra ARG tvärprofessionellt.* Accessed July 19, 2021. https://www.sfog.se/media/299882/riktlinjer_nipt_2016_06_05_.pdf.

Skatteverket. *Individuals and Employees.* Accessed January 4, 2022. https://www.skatteverket.se/servicelankar/otherlanguages/inenglish/individualsandemployees.4.7be5268414bea064694c3e1.html.

Socialstyrelsen. 2019. *Hälsodataregister - räddar liv och ger bättre vård.* Accessed August 2021. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/ovrigt/halsodataregister-information-om-nyttan-med-register.pdf.

———. 2021. *Statistik om aborter 2020* [in sv]. Technical report. Socialstyrelsen, May. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/statistik/2021-5-7373.pdf.

Statistics Sweden. n.d. *Longitudinal integrated database for health insurance and labour market studies (LISA).* Accessed May 2019. https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/vilka-mikrodata-finns/longitudinella-register/longitudinal-integrated-database-for-health-insurance-and-labour-market-studies-lisa/.

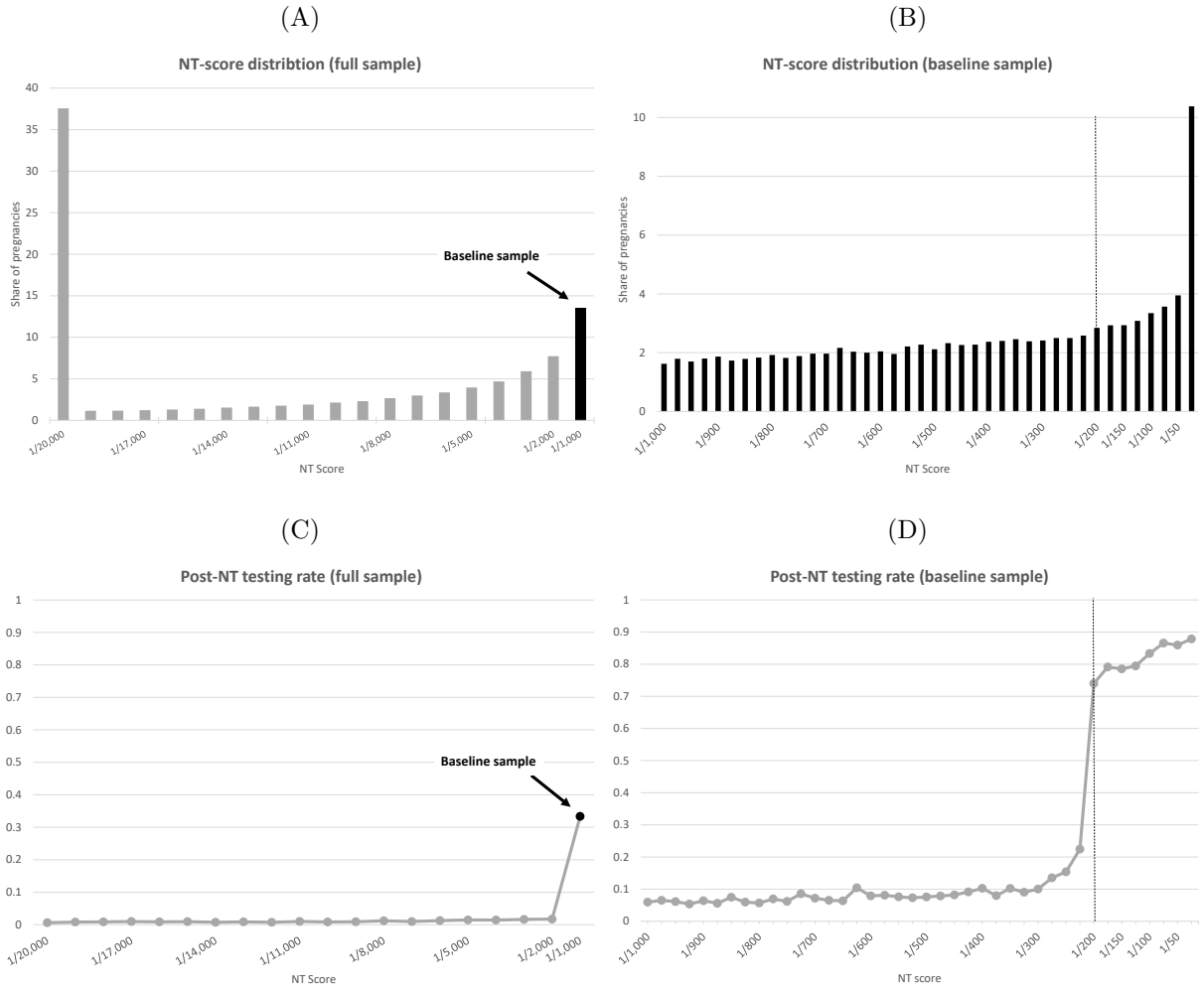
Stephansson, Olof, Kerstin Petersson, Camilla Björk, Peter Conner, and Anna-Karin Wikström. 2018. "The Swedish Pregnancy Register - for quality of care improvement and research" [in en]. *Acta Obstetricia et Gynecologica Scandinavica* 97, no. 4 (April): 466–476. ISSN: 00016349, accessed October 18, 2021. https://doi.org/10.1111/aogs.13266. https://onlinelibrary.wiley.com/doi/10.1111/aogs.13266.

Swedish Research Council. 2020. *The Total Population Register (RTB).* Accessed September 27, 2021. https://www.registerforskning.se/en/register/population-statistics/.

Trentham-Dietz, Amy, Karla Kerlikowske, Natasha K. Stout, Diana L. Miglioretti, Clyde B. Schechter, Mehmet Ali Ergun, Jeroen J. van den Broek, et al. 2016. "Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes." Publisher: American College of Physicians, *Annals of Internal Medicine* 165, no. 10 (November): 700–712. ISSN: 0003-4819, accessed February 9, 2022. https://doi.org/10.7326/M16-0476. https://www.acpjournals.org/doi/full/10.7326/M16-0476.

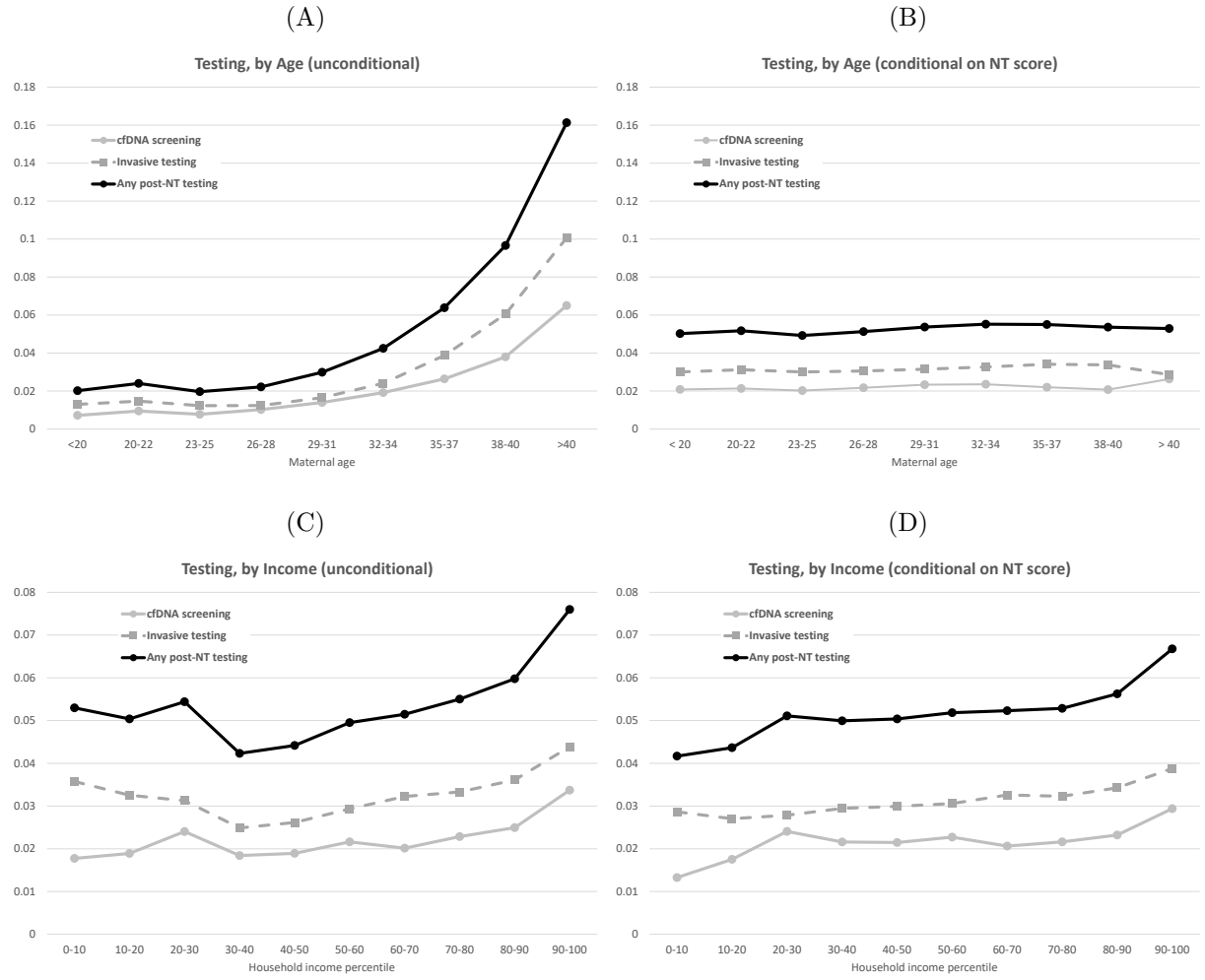
Zhang, Sarah. 2020. "The Last Children of Down Syndrome" [in en]. Section: Health, *The Atlantic,* no. December 2020 (November). Accessed February 24, 2022. https://www.theatlantic.com/magazine/archive/2020/12/the-last-children-of-down-syndrome/616928/.

Figure 1: NT screening risk distribution and rates of post-NT prenatal testing



Note: Figure shows the distribution of NT screening risks and post-NT testing rates for all singleton pregnancies (panels a and c) and the baseline sample of pregnancies with NT screening risk greater than $\frac{1}{1,000}$ (panels b and d). In panels (a) and (c), each bin has width 1,000; in panels (b) and (d) each bin has width 25; x-axis labels show the lower end of that bin. Vertical line in panels (b) and (d) denotes the risk score above which the NT screen is considered "positive" and Sweden recommends that the patient be offered the opportunity to discuss follow-up testing. Singleton pregnancy sample = 297,249 pregnancies; baseline sample = 40,326 pregnancies.

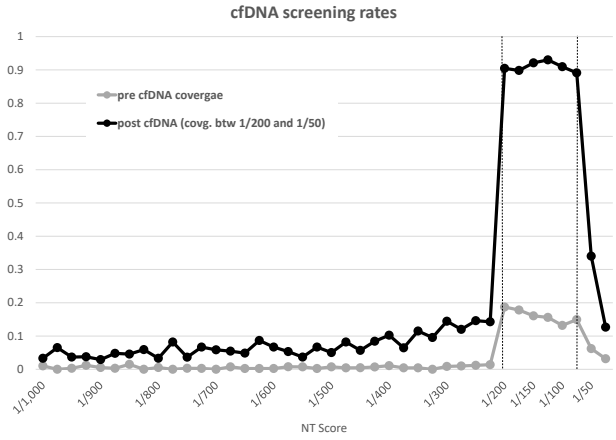
Figure 2: Prenatal Testing following the NT Screening, by Age and Income



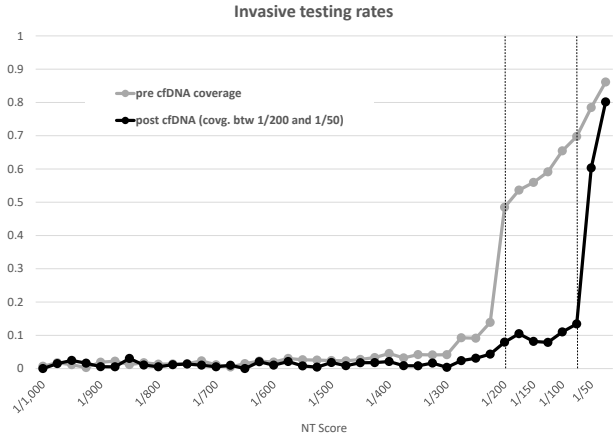
Note: Figure shows the gradient in testing by mother's age (at expected date of birth) and household income rank. Household income is calculated as the average household income across the two years prior to the year of the pregnancy's expected birth date. Household income ranks are then calculated as the within-year rank across the entire Swedish population . Panel A plots testing rates in 3-year age bins and Panel B plots the age gradient after conditioning on 50-point bins of NT risk. Panel C plots testing shares by household income decile and Panel D plots the income gradient after conditioning on 50-point bins of NT risk. Sample: All singleton pregnancies , N = 297,249.

Figure 3: Changes in testing before and after adoption of cfDNA coverage

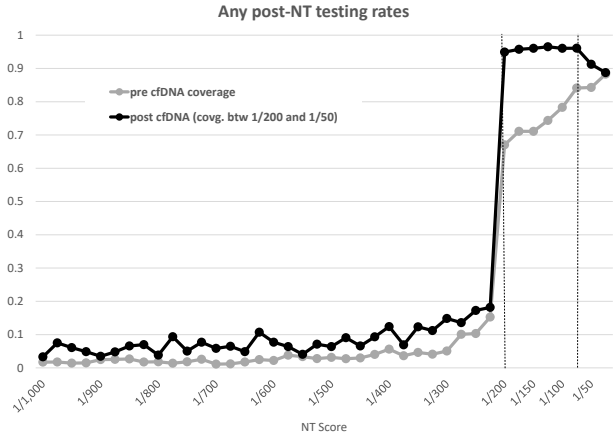
(A)



(B)

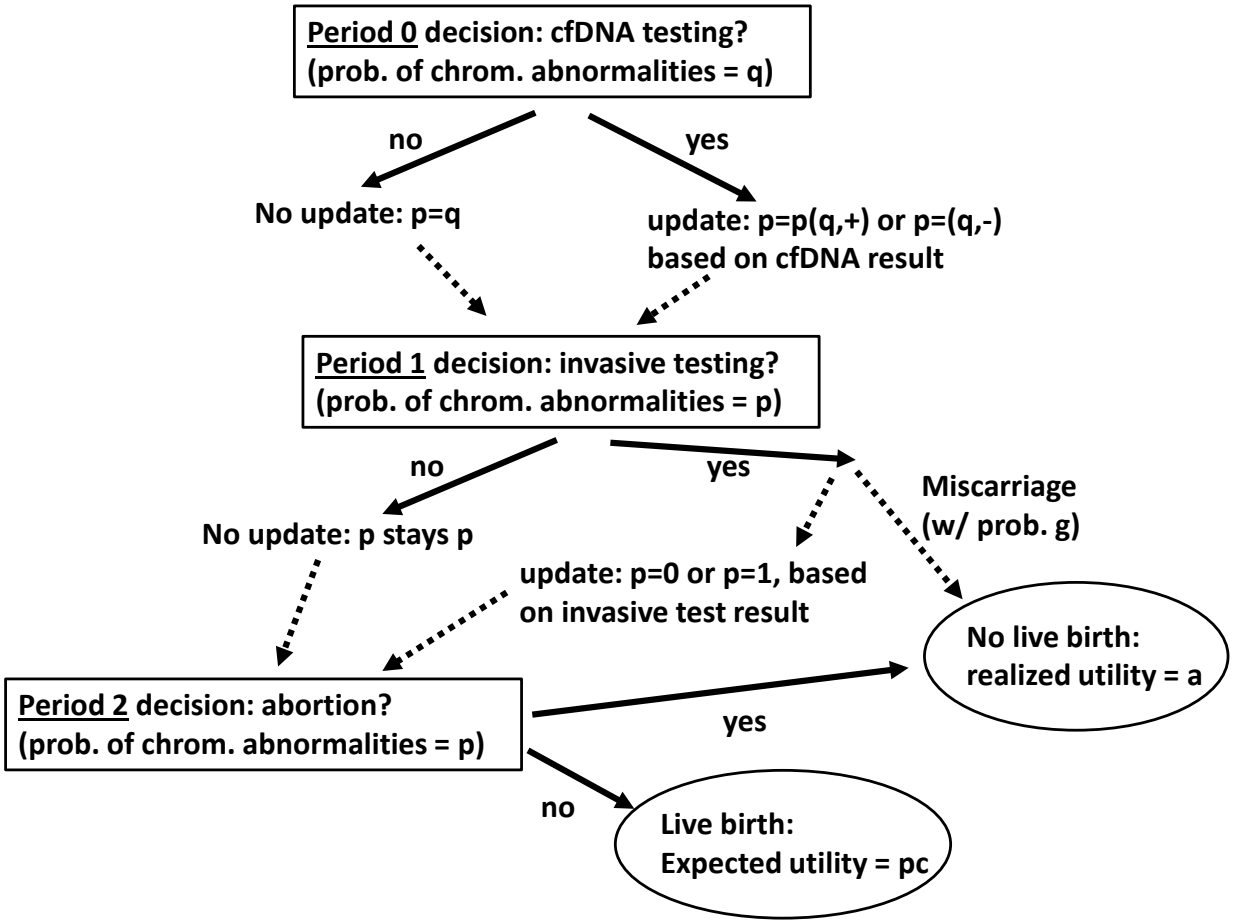


(C)



Note: Figure shows testing rates by NT risk score separately before and after the introduction of coverage for cfDNA. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical lines denote the range of risk scores for which coverage of cfDNA screening is introduced. Sample: Subset of baseline sample that is in regions where cfDNA coverage is introduced for $q \in [\frac{1}{200}, \frac{1}{51}]$ (see Table 1 column 3.) N = 28,851.

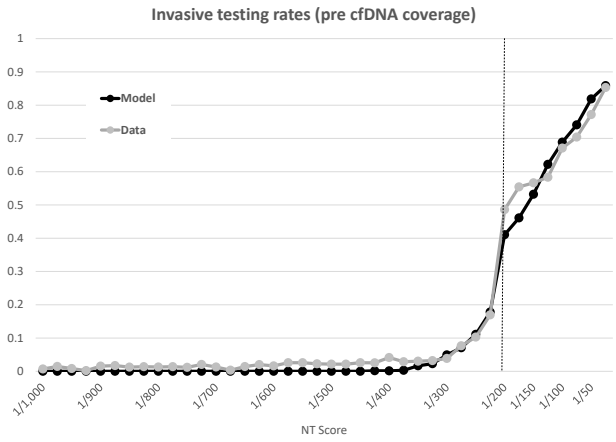
Figure 4: Decision Tree



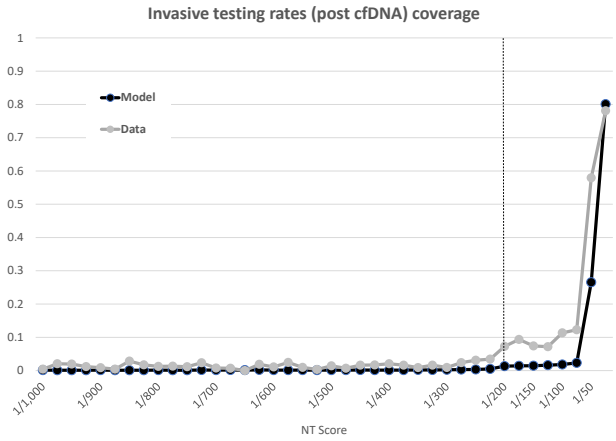
Note: Figure shows the decision tree associated with the model of prenatal testing choices described in Section 3.1

Figure 5: Model Fit to Data - Pooled Across Regimes

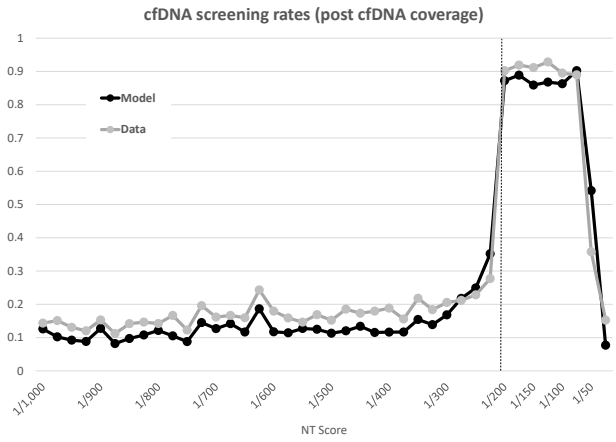
(A)



(B)



(C)



Note: Figure shows testing rates by NT risk score in the data and as predicted by the model, using the estimated parameters from Table 2. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical line denotes the risk score above which the NT screen is considered "positive" and Sweden recommends that the patient be offered the opportunity to discuss follow-up testing. Panels (A) and (B) show invasive testing rates prior to and after the introduction of cfDNA coverage, respectively. Panel (C) shows cfDNA testing rates after the introduction of cfDNA coverage. Testing rates shown are pooled across policy regimes, although we match moments separately by policy regime (See Appendix **??** for the fit of the moments separately for the most common policy regime.). Sample: Analysis sample, N = 40,326.

Figure 6: Willingness to pay for cfDNA and comparative statics

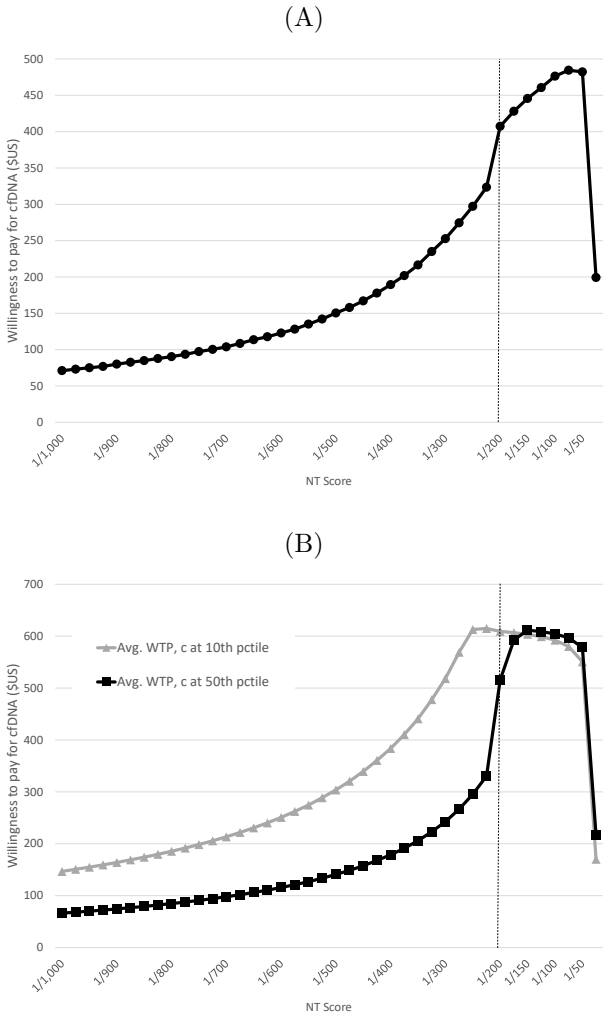
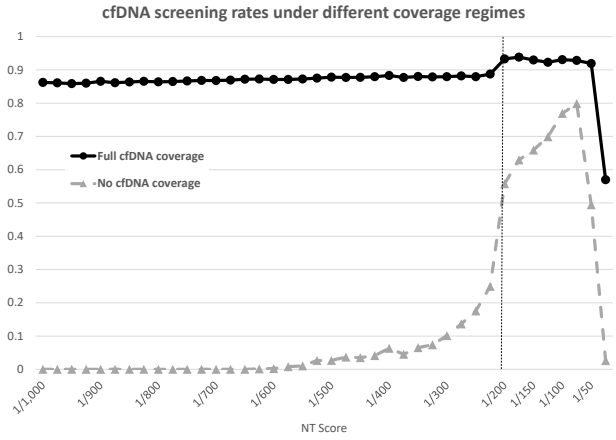


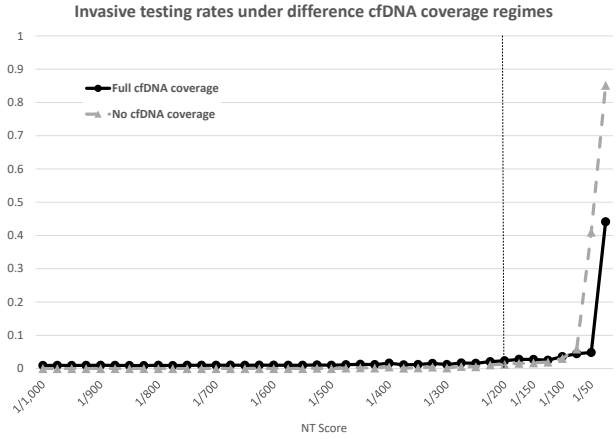
Figure shows the average willingness to pay for cfDNA by NT screening risk under the estimated parameters from Table 2, and comparative statics of average willingness to pay under alternative dis-utilities of a live birth with chromosomal abnormalities. Panel (A) plots the average willingness to pay for cfDNA across the NT screening risk distribution. In Panel (B), we recalculate the willingness to pay for cfDNA under comparative statics of setting the c_i of all pregnancies to the 10th percentile and 50th percentile of the estimated distribution of c_i. We do not show comparative statics over a_i as the estimated σ_a is approximately zero.

Figure 7: Testing rates with and without cfDNA coverage

(A)

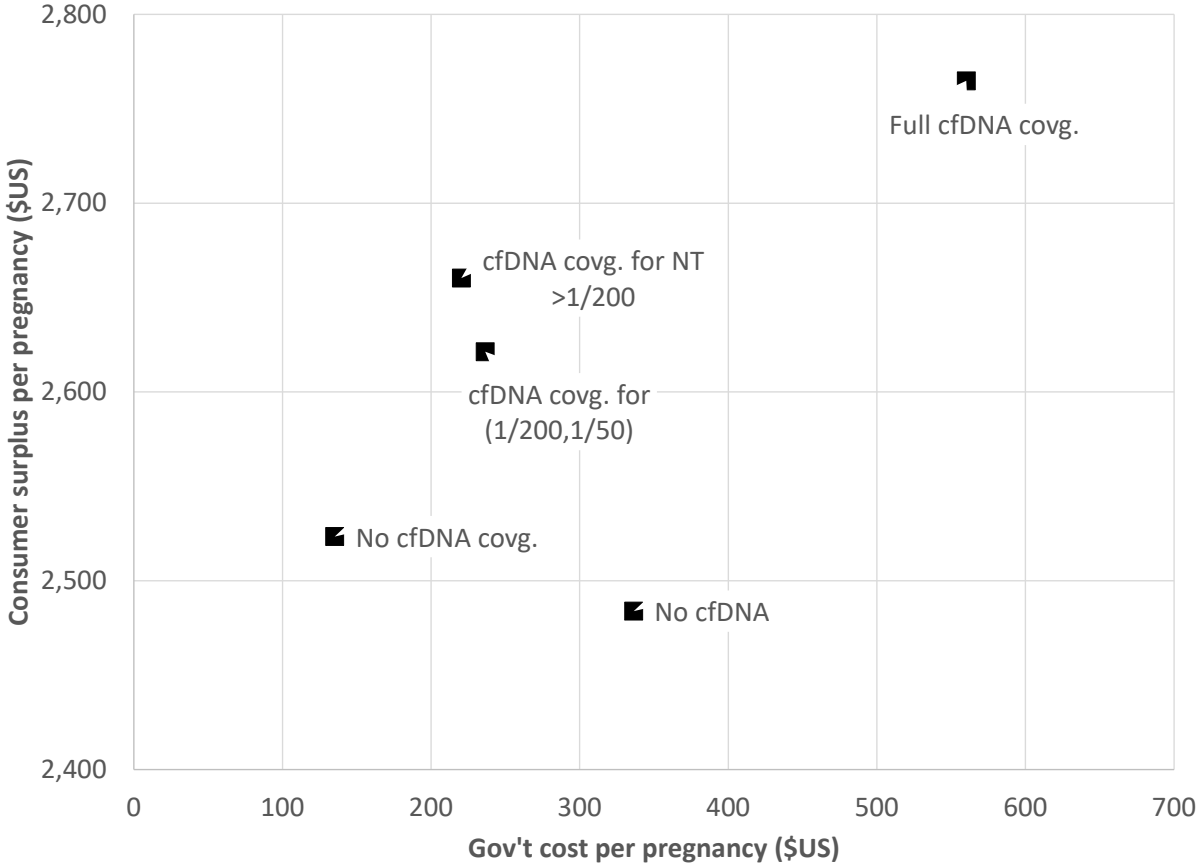


(B)



Note: Figure shows testing rates by NT risk with and without cfDNA coverage. Panel (A) shows cfDNA screening rates and Panel (B) shows invasive testing rates.

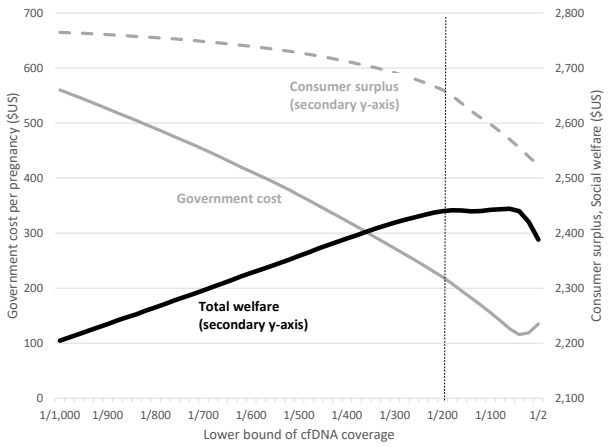
Figure 8: Visualizing tradeoffs for counterfactual cfDNA insurance coverage policies



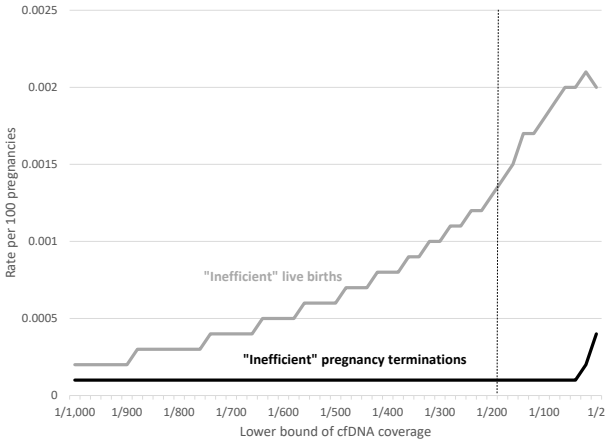
Note: Figure plots estimated average consumer surplus and government cost of prenatal testing per pregnancy (in $US) under counterfactual cfDNA insurance coverage policies. "Full cfDNA coverage" denotes coverage for all pregnancies in the baseline sample (i.e. risk score ¿ $\frac{1}{1,000}$), while other scenarios show cfDNA coverage for no risk scores ("No cfDNA coverage", and coverage for risk scores ¿ $\frac{1}{200}$ or risk scores between $\frac{1}{200}$ and $\frac{1}{50}$. In all of these scenarios, patients whose cfDNA screen is not covered have the option to pay for it out of pocket. By contrast, the "invasive only" scenario assumes that cfDNA screening is unavailable. Consumer surplus is normalized to zero in the scenario where there is neither cfDNA screening nor invasive testing available (i.e. Table 3, column 1.

Figure 9: Optimal cfDNA coverage

(A)



(B)



Note: Panel (A) plots the counterfactual government cost for prenatal testing per pregnancy (left y-axis), average consumer surplus (right y-axis), and total welfare (right y-axis), under different potential lower bounds for cfDNA coverage, Dotted vertical line shows the lower bound that maximizes consumer surplus minus government cost (social welfare). Average consumer surplus is normalized to zero for the counterfactual with no testing from Table 3. Panel (B) shows counterfactual shares of two types of "inefficient" pregnancy outcomes under the different potential lower bounds for cfDNA coverage. "Inefficient" live births are live births with chromosomal abnormalities to mothers who would have preferred to terminate the pregnancy. "Inefficient" pregnancy terminations are terminated pregnancies that would have resulted in a live birth without chromosomal abnormalities. To construct these figures, we compute testing decisions and pregnancy outcomes as the lower risk bound for cfDNA coverage is varied from $\frac{1}{2}$ to $\frac{1}{1{,}000}$.

Table 1: Summary Statistics

	Full sample (all singleton pregnancies)		Baseline sample (= NT score > 1/1,000)		cfDNA covg. for NT in [1/200,1/50] (and NT score > 1/1,000)	
Number of pregnancies	297,249		40,326		28,851	
Demographics:						
Maternal age	32.6		35.7		35.8	
Household income ($US)	63,033		66,904		71,509	
Foreign-born	0.229		0.270		0.272	
No college	0.339		0.321		0.302	
Some college	0.142		0.142		0.147	
College graduate	0.520		0.537		0.551	
Testing:						
Any post-NT testing	0.055		0.341		0.316	
cfDNA testing		0.023		0.131		0.114
Invasive testing		0.032		0.210		0.202
Birth outcomes:						
Live birth	0.968		0.915		0.917	
Live birth, w/o chrom. abnormalities		0.967		0.911		0.914
Live birth, w/ chrom. abnormalities		0.001		0.004		0.003
No live birth	0.033		0.085		0.083	
Ended before 22 weeks		0.030		0.081		0.079
Stillbirth		0.003		0.004		0.003

Note: Table shows summary statistics (means) for maternal characteristics, testing rates and birth outcomes for: all singleton pregnancies (column 1), our baseline sample of pregnancies with NT score $\geq \frac{1}{1,000}$, and the sub-sample of pregnancies in the baseline sample that are regions that ultimately introduce coverage for cfDNA screening for NT scores between 1/200 and 1/50 (column 3); Table B.2 describes which regions are in which policy regimes. Household income calculated as the average household income across the two years prior to the year of the pregnancy's expected birth date, using CPI-adjusted SEK . SEK are then converted to USD using a 1 USD to 8.81 SEK exchange rate.

Table 2: Model Estimates

	Estimate	Std. Err.
μ_a	-126,237	(tba)
μ_c	-198,698	(tba)
σ_a	0	(tba)
σ_c	62,593	(tba)
ρ	0.975	(tba)
ψ	0.938	(tba)

Note: Table shows parameter estimates from the prenatal decision model. Standard errors will eventually be included.

Table 3: Outcomes under Counterfactual Information and Technology

	No post-NT testing (1)	First best (full information) (2)	Free invasive (No cfDNA) (3)	Free invasive & Full cfDNA covg. (4)
Testing (per 100 pregnancies):				
Any testing	0	--	26.94	88.67
cfDNA only	0	--	0	82.67
Invasive only	0	--	26.94	3.15
Both	0	--	0	2.84
Birth outcomes (per 100 pregnancies):				
Live birth	98.74	96.63	96.74	96.64
No chrom. abnormality	95.57	96.20	96.08	96.19
Chrom. Abnormality, a > c	2.74	0.00	0.23	0.02
Chrom. Abnormality, a < c	0.43	0.43	0.43	0.43
No live birth	1.26	3.37	3.26	3.36
No chrom. abnormality	0.63	0.00	0.12	0.01
Chrom. Abnormality (& a>c)	0.63	3.37	3.14	3.35
Cost and surplus ($US per pregnancy):				
Total government cost	0	--	336	560
cfDNA cost	0	--	0	485
Invasive testing cost	0	--	336	75
Consumer surplus	normalized to 0	2,792	2,484	2,765

Notes: Table shows simulated testing decisions and pregnancy outcomes (per 100 pregnancies), and average government spending and consumer surplus (in $US) per pregnancy under alternative assumptions about available technology and information. Consumer surplus is normalized to zero for the counterfactual with no testing. Government spending includes invasive testing ($1,248.50 per test) and cfDNA testing ($567.50 per test). Outcomes that are (ex-post) inefficient are shaded in gray.

Table 4: Outcomes under counterfactual coverage of cfDNA

	No cfDNA (1)	Full cfDNA covg. (2)	No cfDNA covg. (3)	cfDNA covg. for NT > 1/200 (4)	cfDNA covg. for NT in (1/200,1/50) (5)
Testing (per 100 pregnancies):					
Any testing	26.94	88.67	28.24	32.97	32.70
cfDNA only	0	82.67	17.42	27.64	21.79
Invasive only	26.94	3.15	10.12	3.15	10.05
Both	0	2.84	0.70	2.17	0.85
Birth outcomes (per 100 pregnancies):					
Live birth	96.74	96.64	96.80	96.75	96.73
No chrom. abnormality	96.08	96.19	96.17	96.19	96.17
Chrom. Abnormality, a > c	0.23	0.02	0.20	0.13	0.13
Chrom. Abnormality, a < c	0.43	0.43	0.43	0.43	0.43
No live birth	3.26	3.36	3.20	3.25	3.27
No chrom. abnormality	0.12	0.01	0.04	0.01	0.04
Chrom. Abnormality (& a>c)	3.14	3.35	3.17	3.24	3.24
Cost and surplus ($US per pregnancy):					
Total government cost	336	560	135	220	236
cfDNA cost	0	485	0	154	100
Invasive testing cost	336	75	135	67	136
Consumer surplus	2,484	2,765	2,523	2,660	2,621

Notes: Table shows simulated testing decisions and pregnancy outcomes (per 100 pregnancies), and average government spending and consumer surplus (in $US) per pregnancy under alternative assumptions about insurance coverage for cfDNA screening. Throughout, invasive testing is assumed to be available for free. Consumer surplus is normalized to zero for the counterfactual with no testing from Table 3. Government spending includes invasive testing ($1,248.50 per test) and cfDNA testing ($567.50 per test). Outcomes that are (ex-post) inefficient are shaded in gray.

A Data

A.1 Selection into our sample

The NT database is part of the Swedish Pregnancy Register (Stephansson et al. 2018). It is compiled for a subset of clinics in Sweden which cover about 80% of all NT screenings carried out in Sweden from 2011 through 2020 (Graviditetsregistret 2019). The selection of clinics into the NT database is based on which of two algorithms the clinic uses to compute NT scores. The NT database covers clinics that use the dominant algorithm (Graviditetsregistret 2020); see Kublickas, Crossley, and Aitken (2009) for more detailed information about this algorithm, which is calibrated to the Swedish population. Because healthcare in Sweden is provided by the region (county) in which a woman lives, this clinic-based restriction effectively translates into us observing NT screening data for the sub-sample of patients who live close to the clinics covered by this database.

Since the point of entry into the NT database is an NT screen, all the pregnancies in the data have an NT screen and an NT score. Of course, not all pregnancies receive this screening. It is difficult to determine precisely what share of pregnancies that reach the gestational age for NT screening (about 11-14 weeks) receive that screening, since we cannot observe many of the ones that do not. Specifically, we can observe all pregnancies that survive until 22 weeks of gestation or longer in the Medical Birth Register (MBR), compiled by the National Board of Health and Welfare (Socialstyrelsen 2019). However, we have no record in the NT database that are terminated or miscarried before 22 weeks if they do not receive NT screening.

Not all pregnancies in the data are eligible for a free NT screening. Table B.1 shows the regions, years and maternal ages in which NT was covered. It also shows the share of pregnancies that reach 22 weeks in each of the various regimes.

As noted in the main text, we limit our analysis to the 96 percent of pregnancies that, according to the MBR data, occur in region-months that cover NT.[37] The remaining pregnancies are roughly evenly divided between the two NT coverage regimes.

To get a rough sense of what share of pregnancies in the covered region-months receive an NT screen (as well as to measure downstream pregnancy outcomes) we link the NT database to the MBR; that linkage procedure is described below. To calculate the share of pregnancies in covered region-months that receive NT, we limit our analysis to 2019, where we have an estimate that 80 percent of NT screens are in the linked NT-MBR database (Graviditetsregistret 2019). We therefore inflate the number of NT screens in each region by 1.25 to reflect the fact that we only observe 80 percent of them. We estimate that about 60 percent of pregnancies receive NT screens. This number is 72 percent in region-months with universal NT coverage, and about 40 percent in region-months where NT coverage is only for women who are 35 and older; in that setting, rates rise from about 25 percent to 40 percent as maternal age rises from 20 to 35, and then jump to 60 percent for older women (for whom NT is covered).

[37] Almost no pregnancies in the uncovered region-months receive NT (not shown), which is why we exclude them.

A.2 Linkages

As noted above, we link each pregnancy in the NT database to the MBR. We also link each pregnancy in the NT database to additional information in Swedish administrative data on maternal demographics. Specfically, we measure the mother's month and year of birth in Statistics Sweden's Total Population Register (RTB) (Swedish Research Council 2020), and we measure her education, her household income, and whether she is foreign born from the 2009 - 2019 Statistics Sweden's longitudinal database of individual administrative records (Statistics Sweden, n.d.). We determine which prenatal tests were covered for the mother based on the information in the NT database on the clinic's region and the date of NT screening.

Our data use agreement allows us to observe women (and their children) who were registered in the Swedish Population Register (Skatteverket) at some point between 2000 and 2016. The restrictions this sample imposes are to exclude women who immigrated to Sweden in 2017 or later, as well as women who emigrated before 2000.

All linkages are performed using the mother's individual identifier (which is observed in all data sources).[38]

A.2.1 NT database - MBR linkage

Matching screenings in the NT database to births recorded in the MBR requires matching on the mother's ID and the timing of the pregnancy (a mother may have different pregnancies recorded in each dataset and/or multiple pregnancies recorded in either dataset). We use the NT screening date (observed in the NT database), birth month (observed in the MBR), and the expected due date (observed in the NT database and the MBR) to distinguish between pregnancies for mothers with multiple pregnancies and determine correct screening-birth matches.

We first drop observations from the NT database that are missing estimated risk scores and we drop non-single pregnancies from both the NT database and MBR. We then construct screening IDs and birth IDs that uniquely identify the observations in the NT database and MBR, respectively.

For each mother, we create a dataset that includes all potential combinations of births from the MBR and NT screenings from the NT database. For example, if a mother has two births recorded in the MBR and received an NT screening (that was recorded in the NT database) for each of those births, there will be four "candidate" screening-birth observations for that mother in the dataset. Two correct matches, and two incorrect matches. If a mother had two births in the MBR but had only received NT screening (in the NT database) for one of those pregnancies, there would be two candidate screening-birth observations.[39] We drop any births from the MBR for mothers not found in the NT database, and set aside screenings from the NT database for mothers that do not have a birth recorded in the MBR (we will append these back to the data in the final stage of the linkage). The remaining candidate

[38]This individual identifier is a scrambled version of the true social security number, and created by Statistics Sweden.

[39]Note that all of a mother's candidate observations may be incorrect matches. A mother could have a birth recorded in the MBR, and have received NT screening for a separate pregnancy that was not carried 22 weeks.

screening-birth observations comprise a mix of correct and incorrect matches. The goal is to reduce this to a pregnancy-level dataset of all pregnancies that received NT screenings recorded in the NT database, linked to the correct MBR birth record (if there is a matching birth record).

We identify correct matches as screening-birth observations where the NT screening date is no more than 250 days prior to the 15th of the month of birth. The NT screening is generally performed between week 11 (which begins on day 77, 203 days until term) and week 14 (which begins on day 98, 180 days until term) of gestation, but we allow for a longer link period to account for the potential for earlier screenings, late births, and our inability to observe the exact date of birth. We confirm that this procedure identifies a maximum of one correct match for each screening ID.

For screenings where there are no candidate screening-birth observations with a birth month within 250 days of the screening date, we perform an additional step to ensure that our 250 day cutoff is not leading us to miss a correct match. If the 15th of the birth month is within 250 to 270 days of the screening date (i.e. close our 250 day cutoff but not identified as a correct match by the primary matching step), and the expected due date variables from the MBR and NT screening database are within 45 days of each other, we determine that this screening-birth observation is a correct match.

If the two-step matching procedure above finds a correct matching screening-birth for a given screening, we drop the extra incorrect screening-birth candidates for that screening. For screenings with no correct matches, we drop all but one of the candidate observations (and all birth information from that observation), and code that this screening was not matched to a birth (i.e. the pregnancy was not carried at least 22 weeks). Additionally, our matching procedure reveals a small number of cases (15 pregnancies) where the mother obtained two screenings for the same pregnancy. In these cases, we keep the screening with the later test date.

As a final step, we append back the screenings from the NT database for mothers that do not have a birth recorded in the MBR (and code these as pregnancies that were not carried 22 weeks).

A.2.2 NT database - Swedish administrative data linkage

We link the NT database to information on the mother's month and year of birth in RTB using the mother's ID. We link other demographic information from Statistics Sweden's longitudinal administrative database using the mother's ID and the year of the expected due date of the pregnancy.

A.3 Variable definitions

Expected due date: We date pregnancies from their expected due date. We use the expected due date for consistency across pregnancies of different outcomes; pregnancies that were not carried to term do not have a typical "birth date" nine months after conception. While we do not observe the expected due date directly in the raw data, we calculate it for each pregnancy as:

1. The NT screening date plus 270 minus the gestational age at testing (as recorded in NT testing data)
2. Date of last period plus 280 days
3. 15th of month of birth as recorded in MBR (if pregnancy was carried past 22 weeks)
4. The NT screening date plus 196 days[40]

Pregnancy outcomes: The MBR contains all pregnancies carried to 22 weeks of gestation or longer in Sweden from 2011 through 2019. The MBR indicates whether a pregnancy was stillborn or a live birth. We code pregnancies not in the MBR as ended prior to 22 weeks (we cannot distinguish between abortions and miscarriages).

Chromosomal abnormality diagnoses: The MBR contains the (ICD-10) diagnosis codes associated with the neonatal care that a pregnancy received. For pregnancies that resulted in a live birth, we determine that a child has a chromosomal abnormality if the diagnosis codes contain any of: Q90-Q99 (chromosomal abnormalities, not elsewhere classified).

Pre/Post cfDNA coverage introduced: We assign the pregnancy to the "Post-cfDNA" period if the beginning of the 10th week of gestation (210 days prior to the expected due date) is after the introduction of cfDNA coverage in the pregnancy's county. See Appendix B for cfDNA coverage introduction dates by county.

[40]This assumes the test occurs at end of 12th week.

B Swedish coverage policies

Table B.1: NT coverage policies

Regime	Region	Months (from 2011-2019)	Share of pregnancies in MBR
Universal NT coverage	Stockholm	01/2011 - 11/2019	46.70
	Uppsala	01/2016 - 11/2019	
	Ostergotland	01/2011 - 11/2019	
	Jonkoping	01/2011 - 11/2019	
	Kronoberg	03/2012 - 11/2019	
	Kalmar	01/2012 - 11/2019	
	Dalarna	11/2017 - 11/2019	
	Jamtland	09/2017 - 11/2019	
	Halland	05/2017 - 11/2019	
	Varmland	01/2011 - 11/2019	
	Orebro	01/2011 - 11/2019	
	Vasternorrland	01/2015 - 11/2019	
	Jamtland	04/2016 - 11/2019	
35+ NT coverage	Uppsala	01/ 2010 - 01/ 2016	43.91
	Sormland	01/2012 - 11/2019	
	Gotland	01/2011 - 11/2019	
	Blekinge	01/2011 - 11/2019	
	Skane	01/2011 - 11/2019	
	Vastra Gotaland	01/2011 - 11/2019	
	Dalarna	01/ 2011 - 11/2019	
	Gavleborg	01/2011 - 11/2019	
	Vasternorrland	03/2012 - 12/2014	
	Jamtland	01/2011 - 03/2016	
	Vasterbotten	01/2011 - 11/2019	
38+ NT coverage	Blekinge	01/2011 - 06/2016	0.93
No NT coverage	Sormland	01/ 2011 - 12/2011	8.45
	Kronoberg	01/2011 - 02/2012	
	Kalmar	01/2011 - 02/2012	
	Kronoberg	01/2011 - 12/2011	
	Halland	01/2011 - 04/2017	
	Vastmanland	01/2011 - 11/2019	
	Vasternorrland	01/2011 - 02/2012	

Note: Table shows which region-months (from January 2011 through November 2019) are in which of the three NT coverage regimes listed. It also shows the share of pregnancies in the Medical Birth Records in each regime.

Table B.2: cfDNA coverage regimes

Regime	Region	cfDNA coverage introduced	Share of analysis sample
$[\frac{1}{51}, \frac{1}{200}]$	Stockholm	10/2016	71.5
	Sormland	09/2016	
	Gotland	02/2016	
	Halland	05/2017	
	Vastra Gotaland	02/2017	
	Vastmanland	02/2017	
	Dalarna	11/2017	
	Jamtland	09/2017	
$[\geq \frac{1}{200}]$	Uppsala	09/2016	10.4
	Varmland	04/2018	
	Gavleborg	02/2017	
	Vasternorrland	09/2017	
$[\frac{1}{51}, \frac{1}{1{,}000}]$ coverage	Kronoberg	05/2017	13.3
	Skane	05/2017	
$[\geq \frac{1}{1{,}000}]$ coverage	Orebro	10/2018	14.4
No cfDNA coverage	Vasterbotten	NA	0.3
Age 32+ coverage	Blekinge	06/2016	0.0

Note: Table shows the cfDNA coverage regimes observed in our baseline sample. For each, we list the regions included in each regime, the date the region introduced cfDNA coverage, and the share of pregnancies in our baseline sample in the regions included in each regime.

C Model and estimation details

C.1 Details in estimation of model

TBA

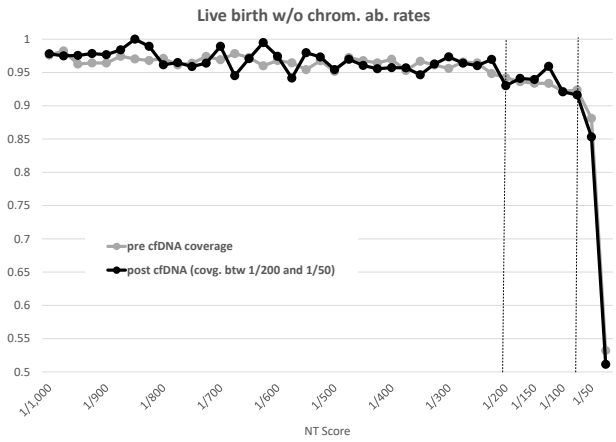
C.2 Details in counterfactual simulations of model

TBA

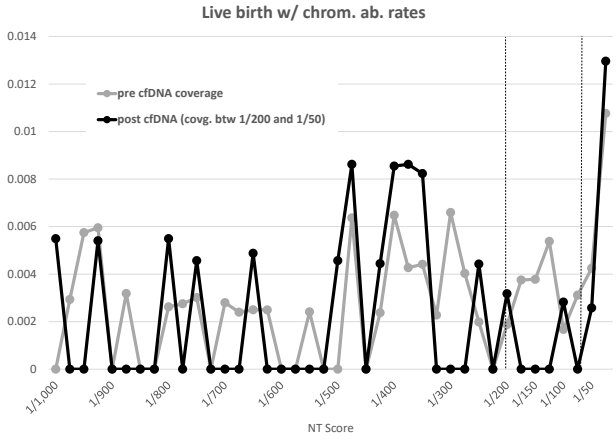
D Additional figures and tables

Figure D.1: Changes in pregnancy outcomes, before and after adoption of cfDNA coverage

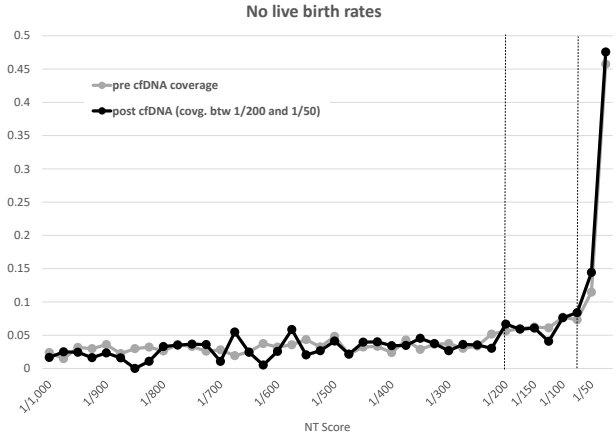
(A)



(B)



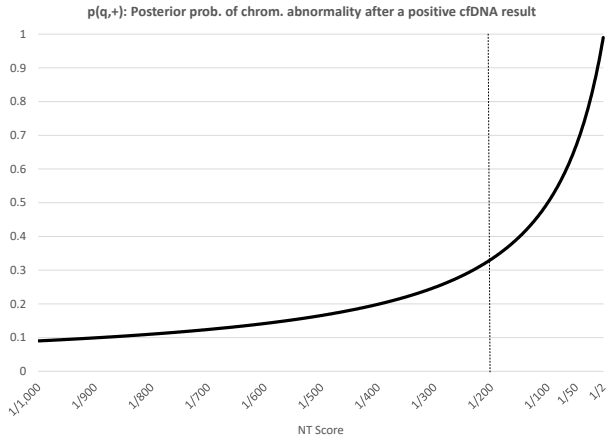
(C)



Note: Figure shows pregnancy outcomes by NT risk score separately before and after the introduction of coverage for cfDNA. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical lines denote the range of risk scores for which coverage of cfDNA screening is introduced. Sample: Subset of baseline sample that is in regions where cfDNA coverage is introduced for $q \in [\frac{1}{200}, \frac{1}{51}]$ (see Table 1 column 3.) N = 28,851.

Figure D.2: Posterior beliefs after cfDNA screening results

(A)



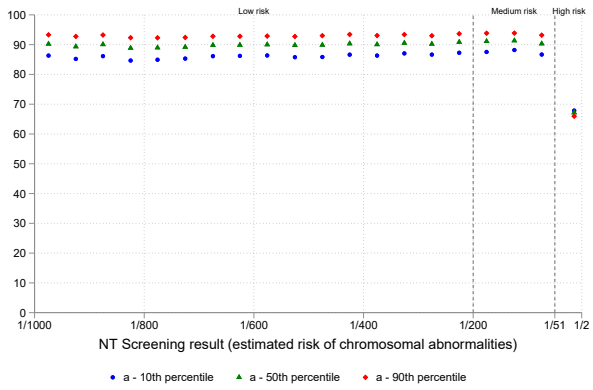
(B)



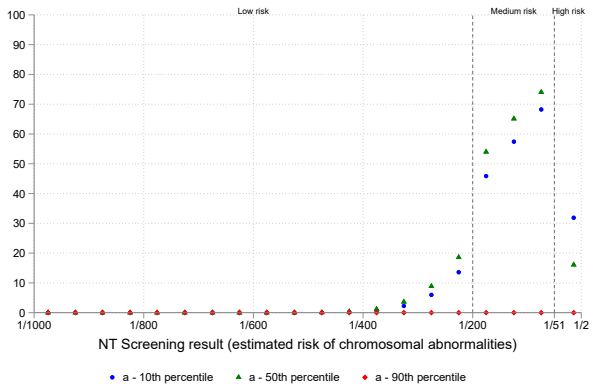
Note: Figure shows the posterior beliefs about the probability of a chromsomal abnormality, by NT risk score, for both positive cfDNA results (panel a) and negative ones (panel b).

Figure D.3: Willingness to pay for cfDNA, comparative statics

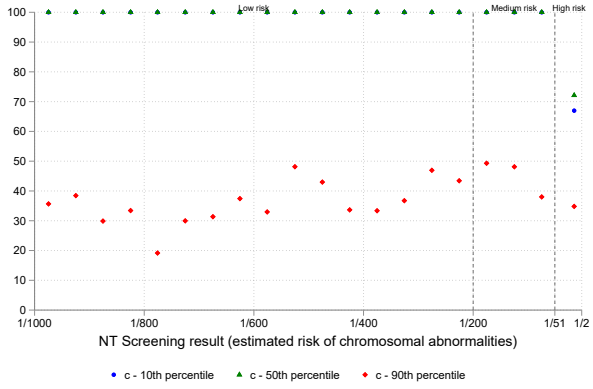
(A) Share above zero, comparative statics over a_i



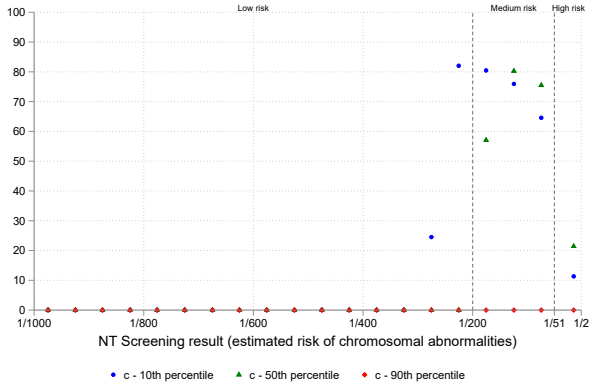
(B) Share above cost, comparative statics over a_i



(C) Share above zero, comparative statics over c_i



(D) Share above cost, comparative statics over c_i



Note: ADD notes here.