Targeting Precision Medicine: Evidence from Prenatal Screening

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

November 29, 2023

Abstract

Medical technologies can target care to patients identified through screening, raising questions of how broadly to screen for potential use. We explore this empirically in the context of a non-invasive prenatal screening, cfDNA, which is used to target a more costly invasive test that elevates miscarriage risk. Using Swedish administrative data on prenatal choices for pregnancies conceived between 2011 and 2019 – a period in which Swedish regions began providing coverage for the new screening – we document that coverage of cfDNA substantially increases cfDNA screening and reduces invasive testing. To assess the impact of counterfactual targeting of cfDNA coverage, we develop and estimate a stylized model of prenatal choices. We find that narrow targeting of cfDNA coverage can improve outcomes and reduce costs, while broader coverage also improves outcomes but with increased 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. We are grateful to Michaela Granfors, Kerstin Petersson, and Jonas Söderling at the Pregnancy Register. We appreciate helpful comments from Rachel Griffith (the Editor), three anonymous referees, Barton Hamilton, and seminar participants at ASU, Berkeley, Chicago, Copenhagen, Duke, IFS, LSE, Mannheim, MIT, TSE, Stanford, Stockholm, Tel Aviv, UCLA, USC, Warwick, the All California Labor Conference, the Cowles Summer Conference on Labor and Public Economics, the NBER Children's Group Spring Meeting, and the NBER Health Care Summer Institute. We thank Sean Gao, Kelsey Moran, Jasmin Moshfegh, Katherine Moulton, Cameron Pfiffer, and especially Gabriel Swagel and Carine You for excellent research assistance, as well as Sarah Bögl and Iliriana Shala for research assistance at the Research Institute of Industrial Economics. We are especially grateful to Heidi Williams, who played a key role in the genesis of this project and made important contributions throughout its early stages. Persson additionally acknowledges funding from the National Science Foundation (NSF) CAREER Award No. 2144072. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation.

1 Introduction

"Big data" has the potential 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 the holy grail of health care: improving patient health and well-being while simultaneously reducing health-care spending (Aspinall and Hamermesh 2007; Armstrong 2012). However, what is less appreciated – and certainly less celebrated – is that personalized medicine requires information that is needed for personalization, and acquiring such information typically entails costs.[2] Ironically, therefore, the availability of precision medicine in turn raises questions about how finely to target the acquisition of information that can be used for further targeting.

In this paper, we provide a quantitative economic framework for analyzing the trade-off between costly information acquisition and the ability to use the acquired information to better target subsequent, more expensive and invasive procedures. Although we estimate and illustrate these issues in one specific context, the medical and policy questions surrounding how broadly to use an inexpensive screening or procedure to better target subsequent medical care are widespread. These policy discussions span a range of medical settings, including screenings in pregnancy and newborns, cancer screening and treatment, and the management of acute illness. Our framework and empirical results highlight, among other things, that the value of information acquisition may be non-monotone, highest for intermediate risk levels.

Our specific setting is 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 (CVS) – have been able to diagnose fetal chromosomal abnormalities, but elevate the risk of miscarriage.

1. See, e.g., Basu and Meltzer (2007) and Basu, Carlson, and Veenstra (2016). An example of this type of precision medicine is targeted therapy for the treatment of some cancers (Aspinall and Hamermesh 2007; Goldman et al. 2013; Berndt, Goldman, and Rowe 2018).

2. Several recent papers have studied targeting of either screening or treatment technologies based on readily – and hence costlessly – observable individual characteristics, such as age (Einav et al. 2020) or family medical history (Persson, Qiu, and Rossin-Slater 2021). Our focus, however, is on information that is costly to acquire. Examples under active discussion include the possibility of 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).

Since the early 2000s, the development of a non-invasive prenatal screening technology – nuchal translucency (NT) – has offered a lower-cost way to assess the likelihood of the most common chromosomal abnormalities without any risk of miscarriage, and thus inform the decision of whether to undertake invasive testing. We analyze the impact of a second-generation form of non-invasive screening – known as cell-free DNA (cfDNA) screening – which substantially increases the informativeness of the non-invasive screening, but also considerably elevates screening costs. We empirically explore the costs and benefits of how to target the availability of this second-generation screening technology through public policy decisions of whether and when to cover its cost. This is currently an active – and evolving – policy debate in both the US and in many European countries (Minear et al. 2015; Gadsbøll et al. 2020; GenomeWeb 2020).

Our analysis draws on detailed Swedish administrative data for pregnancies that received the first-generation screening between 2011 and 2019. This screen produces a risk score for each pregnancy that is a cardinal measure of the probability of chromosomal abnormalities. This risk score is used by parents to make further decisions and by the government to determine coverage for second-generation screening. Crucially, we observe this risk score, as well as subsequent screening and testing decisions.

Throughout our study period, first-generation NT screening and invasive diagnostic tests are available for free for all pregnancies, as are downstream medical procedures. Over our study period, various regions in Sweden introduced coverage for the new cfDNA screening, making different choices about how narrowly to define the set of risk scores eligible for free cfDNA screening. Our data and setting thus provide a rare opportunity to empirically evaluate the impacts of targeting a new screening tool and the implications for optimal policy design. We begin by documenting that the introduction of coverage for cfDNA screening has enormous effects, both on increasing cfDNA screening and on reducing invasive testing for risk scores that receive coverage.

In order to quantitatively assess the impact of alternative cfDNA coverage regimes on pregnancy outcomes and patient welfare, 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 chromosomal abnormalities, or no live birth (due to miscarriage or abortion). All pregnancies receive the initial NT screen and its associated risk score. Patients 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 depend on the risk score, the patient's relative

utilities from the possible pregnancy outcomes, and the out-of-pocket cost they face for cfDNA screening. The socially optimal decision regarding cfDNA screening coverage depends on the patient's expected utility with and without coverage, as well as the impact of coverage on government costs. The impact of coverage on government costs is a-priori ambiguous: the cfDNA screening itself entails costs, but can decrease spending on subsequent invasive testing, which is about twice as expensive.

We restrict our analysis to pregnancies with risk scores that are high enough to have the cfDNA screening covered under the recommendation of the relevant Swedish advisory body; this is also the most expansive coverage regime we observe. While this limits our analysis to about 13% of pregnancies in the data, these pregnancies account (in expectation) for almost all (97%) those 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 then apply the estimates to considering outcomes under alternative coverage regimes. We analyze a variety of outcomes, including screening and testing rates, the three possible pregnancy outcomes, government spending, patient surplus, and the rate of what we term (ex-post) inefficient outcomes. There are two forms of inefficient outcomes: no live birth from a pregnancy that had no chromosomal abnormalities (due to miscarriage resulting from invasive testing), and live births from a pregnancy with chromosomal abnormalities born to patients who would have preferred no live birth (had they known about the chromosomal abnormalities in advance).

The results suggest that with relatively broad coverage, the introduction of the new cfDNA screening would have the standard impact of most new medical technologies: improving patient welfare but also increasing health-care costs. In particular, we estimate that in the absence of the new cfDNA screening technology, about one-quarter of patients opt to do invasive testing, despite the miscarriage risk that comes with it. Spending per pregnancy for this testing is $298 US, and 0.30% of pregnancies result in an (ex-post) inefficient outcome. If the new screening technology covered all studied pregnancies,[3] cfDNA screening rates would be 86%, the rate of invasive testing would fall to 5% (a 79% decline), and the rate of inefficient outcomes would fall to 0.03% (a 90% decline). Consumer surplus would be $267 higher per pregnancy, but government spending on screening and testing per pregnancy would rise by $250 per pregnancy; that is, the increased spending on cfDNA screening would not be fully offset by the decreased spending on invasive testing.

3. Recall that these are the 13% of pregnancies that have the highest risk, and coverage for them is recommended in Sweden.

By contrast, if the government instead introduced more targeted coverage for the new cfDNA screening technology, it could both improve patient welfare and lower health-care costs. For example, if – as is the case in several of the Swedish regions – the screening technology were covered for only the highest one-third of the risk scores in our sample, we estimate that the cfDNA screening rate would be only about 30%, and invasive testing would be 5%. Introducing this more narrowly targeted cfDNA coverage would both increase consumer surplus (by $154 per pregnancy) *and* reduce government spending (by $89 per pregnancy) relative to the case where the new technology is not available.

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 of health care – improving patient well-being while saving money. However, if made more widely available, screening becomes the more typical form of medical technology, raising both patient well-being as well as health-care spending.

Our findings speak to the large literature on the consequences of medical innovation, suggesting that those consequences are not innate, immutable characteristics of the technology, but can be shaped by public policies. In the United States, medical innovation is both widely credited for the dramatic secular improvements in health that have occurred over the last half century, and widely blamed for the equally dramatic rise in health-care spending over the same period (Newhouse 1992; Cutler 2004; Chandra and Skinner 2012). In this context, there is growing interest in the potential for precision medicine to transform technologies that are highly effective for a subset of patients but end up cost-increasing because they are applied too broadly. Better targeting of these technologies would make them both highly health effective and highly cost effective, with little chance of overuse (Goldman et al. 2013; Berndt, Goldman, and Rowe 2018). Using the framework developed by Chandra and Skinner (2012), this has the potential to transform such technologies from "type II" into "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 can depend critically on the targeting design of public policy regarding the prices patients face for screening.

In this sense, our paper complements an existing literature analyzing the optimal pricing of alternative, more expensive, treatment options (Hirth, Chernew, and Orzol 2000; Einav, Finkelstein, and Williams 2016; Hamilton et al. 2018). For therapeutic treatments, the value of the therapy is increasing in the probability that the patient will be found appropriate for it, while for screening and diagnostic tests the value of information may be highest for mid-range risks, where the information is most likely to affect subsequent medical decisions.

Indeed, we see in our setting that the willingness to pay for the new cfDNA screening is hump-shaped with respect to risk.

Further, our paper contributes to a literature at the intersection of medical innovation, family economics, and bioethics concerns over health technologies that offer the promise of learning more about a pregnancy. Such technologies enable patients 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 (Ball 2017; Devlin 2019; 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 more than 95% of fetuses who are prenatally diagnosed with the condition are aborted (Conner et al. 2012; Zhang 2020). Our modeling framework allows us to recover the distribution of parental preferences for terminating pregnancies diagnosed with Down syndrome, and our results suggest that this statistic over-states patient preferences for termination.

The rest of the paper proceeds as follows. Section 2 describes our setting and data, and Section 3 presents the descriptive patterns for screening and testing. Section 4 lays out the model and its estimation. Section 5 describes our results. The final section summarizes our findings and offers a discussion of their implications for other medical settings.

2 Setting and data

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 abnormalities. Amniocentesis, the first procedure to be developed, 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, these diagnostic tests carry a miscarriage risk – a risk that is typically quoted to be around 0.5% in our clinical context during our sample period (Oster 2021).[4]

4. As the procedures have becomes safer over time, this oft-quoted estimate may in practice be over-stated; for example, Akolekar et al. (2015) estimates the procedure-related risks to be 0.22% for amniocentesis and 0.11% for CVS, respectively.

More recently, from the 2000s onward, 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 to 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 (GARD 2020; MedlinePlus 2021).[5]

Non-invasive screening poses no risk of miscarriage and costs less than an invasive diagnostic test, but it is also less informative about the presence of chromosomal abnormalities. In particular, unlike an invasive test, a non-invasive screen only provides a risk assessment, not a definitive diagnosis.[6] Non-invasive screening is therefore typically used to inform decisions about whether or not 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).[7] 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, cfDNA screening is both more expensive and better than its earlier counterpart. 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,

5. More recently, screens have been developed for other, much rarer chromosomal abnormalities (Kliff and Bhatia 2022), but these so-called "third-generation" prenatal screens had not arrived in Sweden by the end of our study period.

6. In addition, invasive testing can diagnose all chromosomal abnormalities, whereas non-invasive screening only provides information about the three most common abnormalities described above, and about very rare abnormalities in sex chromosomes.

7. cfDNA is sometimes also referred to as non-invasive prenatal testing or screening, or NIPT.

8. In our setting, costs run from about 1,500 SEK ($174) for NT screening to 5,000 SEK ($567.60) for cfDNA, to 11,000 SEK ($1,248.50) for an invasive test (SBU 2016; Ingvoldstad-Malmgren et al. 2017).

9. These rates are computed by converting NT risk scores to a positive result if they are greater than or

therefore, there is no consensus on recommended practice for cfDNA coverage.

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 make the NT screening 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 is above some threshold (Flessel and Lorey 2011; Crombag et al. 2014; 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 only for certain NT 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. The Swedish recommendation, however, also takes into account the significant cost difference between NT and cfDNA screening (SFOG 2016). The cost of cfDNA screening relative to the information it provides is therefore 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.[12] There are, however, some differences in covered services across Sweden's 21 regions, and this includes coverage of prenatal screening. At the start of our study period in 2011, many regions provided free NT screening for all pregnancies. Swedish law[13] stipulates that all women with a heightened likelihood of fetal chromosomal abnormalities be offered additional information about diagnostic testing; "heightened risk" is defined as an NT score of $\frac{1}{200}$ or higher (Kublickas, Crossley, and Aitken 2009), which is

equal to $\frac{1}{200}$ 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).

12. 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. Lag (2006:351) om genetisk integritet.

also referred to as a "positive" NT test result in our setting.[14]

With the development of the second-generation non-invasive screen, the SFOG issued a national recommendation in 2016 to offer universal NT screening, followed by cfDNA screening for pregnancies with NT risk scores that fall between $\frac{1}{1,000}$ and $\frac{1}{51}$. It recommended that the highest-risk pregnancies (risk score of $\frac{1}{50}$ or higher) "skip" cfDNA screening and obtain a definitive diagnosis via invasive testing (SFOG 2016). The logic behind not offering cfDNA screening to the lowest-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 screening is lower. The logic behind not offering cfDNA screening for the highest-risk pregnancies is more nuanced. It reflects an assumption that patients receiving this risk score are very likely to undertake 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 would be wasteful (SFOG 2016).

The SFOG recommendations do not map directly into policy regarding cfDNA coverage.[15] In the first few years after its 2016 recommendation, different Swedish 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 $\frac{1}{51}$.[16] Other regions chose to cover all risk scores above $\frac{1}{200}$, all risk scores above $\frac{1}{1,000}$, or (in the case of two regions) to follow the national recommendation and cover risk scores between $\frac{1}{1,000}$ and $\frac{1}{51}$.[17]

2.3 Data and sample construction

We provide a brief overview of our data and variable definitions. Appendix A provides more detail on both.

14. In principle, invasive testing was available for free for all pregnancies, regardless of the NT score or of whether an NT screening was undertaken, but initiating a follow-up discussion with the patient was only recommended for risk scores of $\frac{1}{200}$ or higher.

15. The same is true for the recommendations of medical authorities in 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, for no pregnancies, or for pregnancies with certain risk scores (Gadsbøll et al. 2020).

16. 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 result, i.e., with a risk score of $\frac{1}{200}$ or higher (Ingvoldstad-Malmgren et al. 2017).

17. Appendix Table A1 provides more detail, including information about additional regimes that were adopted in regions that are not included in our data.

Data. The backbone of our data is Sweden's *NT database* of pregnancies from 2011-2019, which is part of the Swedish Pregnancy Register. The NT database contains prenatal testing data reported by clinics that performed 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 risk score.

The database only contains pregnancies that received an NT screening, which is performed during weeks 11-14 of pregnancy. Women who terminate their pregnancy prior to the NT screening do not enter our sample. Given that more than 90% of all abortions are performed by the end of week 12 (Socialstyrelsen 2021), and that later abortions are often performed after the discovery of chromosomal abnormalities (Graviditetsregistret 2020), we assume that the pregnancies that receive an NT screening are all desired pregnancies.

For each pregnancy, we observe the date of the NT screening, gestational age at screening, mother's age at the due date, the number of fetuses, and the region where the clinic is located. Crucially, we also observe the result of the NT screening: a risk score q for each fetus that is a cardinal measure of the probability of chromosomal abnormalities.[18] 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 data from the NT database to population-wide health records from the National Board of Health and Welfare. For each pregnant woman, we observe all pregnancies recorded in the Medical Birth Register (MBR) from 1985 through 2019 and all inpatient and specialist outpatient visits from 2001 through 2019. The MBR contains all pregnancies carried 22[19] weeks or longer. It records the pregnancy outcome (live birth or stillbirth), gestational age at birth, the baby's diagnoses (ICD codes), and whether the baby dies within 28 days of birth. We use the MBR to track the outcomes of each pregnancy in the NT database; we code the pregnancy as terminated if we do not observe it in the MBR (i.e., we do not observe it after week 22 of gestation; we cannot distinguish between miscarriages and abortions). We also use the MBR to characterize the *prior* pregnancy history for each patient with a pregnancy in the NT database; specifically, we measure whether the patient has had a *prior* pregnancy that resulted in a stillbirth, a live birth where the infant died within 28 days of birth, or a pre-term live birth (prior to 37 weeks of gestation). Likewise, the inpatient and specialist outpatient records allow us to measure whether the patient had a prior miscarriage late in a pregnancy.[20]

18. 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, we (as well as physicians) use the approximation $q \approx \max\{q_1, q_2\}$.

19. Prior to July 1, 2008 the MBR contains all pregnancies carried 28 weeks or longer.

20. Many early miscarriages will not make it into these data because they are either handled at home or

Finally, we link these data to information on maternal demographics from several Swedish population registers from 2009-2019. We record the mother's education and marital status in the year prior to the due date, her household income in the two years prior to the due date, her household income rank, and whether she is foreign born.[21]

Sample construction. We focus on pregnancies in the NT database from 2011 through 2019. We limit the sample to the approximately 97% of all pregnancies that are singleton pregnancies. We also limit our sample to the approximately one-half of pregnancies that occur in region-months that provide universal NT screening (see Appendix Table A2). When NT screening is universally covered, we estimate that about three-quarters of pregnancies receive an NT screen (see Appendix A). 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.[22]

After these restrictions, our data contain 234,817 pregnancies, carried by 180,697 unique women. We refer to this as our "full sample." For most analyses, we further restrict attention to a "baseline sample" of the 13% of these pregnancies with a risk score q of $\frac{1}{1{,}000}$ or higher; this is the lowest risk score for which cfDNA screening is covered in any region at any point during our sample period. This baseline sample includes 30,479 pregnancies, carried by 28,512 unique women. Although this sample excludes the majority of pregnancies, the variation in risk score in the excluded 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 almost all (97%) of the pregnancies in the full sample that are associated with chromosomal abnormalities.[23]

through the primary care system.

21. To calculate household income rank we take the maximum of the household income percentile measured one and two years before the (year of the) due date, with percentiles defined relative to other women who give birth in the year of the due date.

22. Appendix Table A3 shows how these restrictions affect the sample composition. Compared to all live births (column (1)), limiting to region-months with universal NT (column (2)) has little impact on the sample composition. Requiring an NT screen within those region-months (column (3)) results in a sample that is slightly older, more educated, and has higher income.

23. To compute this share, we assume that the NT risk score reflects the actual probability of chromosomal abnormalities for each pregnancy; this is realistic as the risk score in our data is produced by an algorithm that is calibrated off of the Swedish pregnancies in our database (Kublickas, Crossley, and Aitken 2009). 97% is then given by the ratio $\sum_{q_i \geq \frac{1}{1{,}000}} q_i / \sum_{q_i \geq \frac{1}{20{,}000}} q_i$.

3 Descriptive results

3.1 Summary statistics

Table 1 presents summary statistics for three samples: all singleton pregnancies, the baseline sample (pregnancies with an NT risk score of $\frac{1}{1,000}$ or higher), and the approximately 80% of pregnancies in the baseline sample that are in region-months that adopt the most common cfDNA coverage regime – which covers cfDNA for NT risk scores between $\frac{1}{200}$ and $\frac{1}{51}$; we will sometimes analyze this subset of the baseline sample separately. In the full sample of singleton pregnancies (column (1)), average maternal age is 32, 38% of the pregnancies are carried by married women, 49% by college graduates, and 23% by women who are foreign-born. About half the sample has had a prior birth, and about one-quarter has had a prior pregnancy issue (a previous miscarriage, stillbirth, death within 28 days of birth, or a pre-term live birth). Two percent has had a prior birth associated with congenital deformation or chromosomal abnormality.

The pregnant women in our baseline sample (column (2)) are older on average than women in the full sample (average maternal age of 35 relative to 32), which is to be expected given the relationship between maternal age and the risk of chromosomal abnormalities. They are also of slightly higher income and educational attainment, and are slightly more likely to have had a previous pregnancy or birth issue. About one third of our baseline sample does some post-NT testing, compared with about 5% in the full sample. Pregnancies in our baseline sample are also more likely to result in no live birth (8% vs. 3%). Despite higher rates of testing and pregnancy termination, women in our baseline sample are more likely to have a live birth with chromosomal abnormalities (0.3% vs 0.1%), as would be expected given the much higher risk of chromosomal abnormalities for pregnancies in this sample.

Figure 1 shows the distribution of NT risk scores and the rate of any post-NT testing (cfDNA screening and/or invasive testing) by risk score. We show this separately for the full sample (panels (a) and (c)) and for the baseline sample (panels (b) and (d)). Low-risk pregnancies that are excluded from our baseline sample (that is, pregnancies with a risk score lower than $\frac{1}{1,000}$) are associated with an extremely low post-NT testing rate. Post-NT testing rates also rise sharply at the risk score of $\frac{1}{200}$; as noted earlier, risk scores of $\frac{1}{200}$ or higher are considered "heightened risk," and (throughout our study period) Swedish law stipulates that all women with a heightened likelihood of chromosomal abnormalities be offered additional information about diagnostic testing.

The risk score from the NT screening is a key predictor of further prenatal testing, but other factors, such as maternal age, income and education, and prior pregnancy experiences

may also play a role. To investigate this, Figure 2 shows post-NT testing rates by various maternal demographics. The top two panels show the results for the full sample, with panel (a) showing results unconditionally, and panel (b) showing results conditional on the NT risk score. Panel (a) shows that post-NT testing rates rise sharply with age, from 3.4% for pregnant women who are 25-35, to 10% for pregnant women over 35. Testing rates also increase with education and income, and are higher for women who have had previous children and for women who had previous pregnancy or birth complications (specifically, a previous miscarriage, stillbirth, death within 28 days of birth, pre-term live birth, or live birth with chromosomal abnormality or congenital deformation). Panel (b) shows that these gradients are substantially attenuated once we condition on NT risk score. For example, the 6.6 percentage point difference in post-NT testing rates for pregnant women over age 35 compared to those aged 25-35 shrinks to only a 0.5 percentage point difference conditional on risk score. This is not surprising since the algorithm determining the NT risk score takes maternal age into account. But the reduction in the other demographic gradients suggests that much of the difference in testing across women of different demographics reflects underlying differences in risk scores, not preferences conditional on risk score. Not surprisingly, therefore, in our baseline sample (which conditions on an NT risk score of $\frac{1}{1,000}$ or higher), post-NT testing rates are fairly uncorrelated with demographics (panel (c)). This is especially true after conditioning on the NT score (panel (d)). For this reason, our analysis below will abstract from these demographics; we show below that our main findings are unaffected by incorporating maternal age directly into our model.

3.2 Testing decisions by risk score and policy regime

Figure 3 plots some initial, descriptive evidence of how coverage of the new cfDNA screening technology affects screening and testing decisions for individuals with different risk scores. We show decisions before and after the introduction of the most common cfDNA screening coverage policy: cfDNA coverage for NT risk scores in $[\frac{1}{200}, \frac{1}{51}]$. Slightly over 80% of our baseline sample is in regions which adopted this policy in 2016 or 2017 (Appendix Table A1).[24] Moving from left to right along the x-axis in each panel, the NT risk score – the probability of chromosomal abnormalities – rises from $\frac{1}{1,000}$ to $\frac{1}{2}$.[25] The gray dots show testing decisions when cfDNA screening is not covered for anyone, while the black dots show behavior after cfDNA coverage is introduced for NT risk scores in $[\frac{1}{200}, \frac{1}{51}]$.

24. Appendix Figures A1, A2, and A3 reproduce the corresponding descriptive figures for the three other cfDNA coverage regimes that we observe in the data.

25. As noted, the risk scores are censored (also in practice and in communication with the patient) so that all those above $\frac{1}{2}$ receive a risk score of $\frac{1}{2}$.

When cfDNA screening is not covered for anyone (gray dots), both cfDNA screening rates (panel (a)) and invasive testing rates (panel (b)) are essentially zero for risk scores below $\frac{1}{200}$, and jump sharply at the $\frac{1}{200}$ threshold; these sharp jumps likely reflect the medical practice of classifying 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. Interestingly, as the risk score increases further above $\frac{1}{200}$, rates of cfDNA screening fall, and reach essentially zero again for the highest risk score bin, while rates of invasive testing rise. These patterns suggest that, as risk scores rise above $\frac{1}{200}$, more and more patients plan to do invasive testing regardless of the cfDNA screening result; they therefore "skip" cfDNA screening (and its associated cost to the patient of about $567.50 in this initial period) and move directly to invasive testing.

When coverage for cfDNA screening is introduced for risk scores between $\frac{1}{200}$ and $\frac{1}{51}$, rates of cfDNA screening jump from about 20% (gray dots) to 90% (black dots) in the covered region, but drop sharply once coverage ends at $\frac{1}{50}$.[26] The pronounced increase in cfDNA screening in the covered range (panel (a)) is mirrored by a pronounced drop in the rate of invasive testing in that range (panel (b)); however, once cfDNA coverage ends at $\frac{1}{50}$, the rate of invasive testing jumps up sharply to over 60% (panel (b)).

Thus, coverage of cfDNA screening in the risk score range of $\frac{1}{200}$ to $\frac{1}{51}$ switches the dominant form of information acquisition in that range from invasive testing to cfDNA screening. It also increases the probability of *any* testing after the NT screening (that is, either cfDNA screening or invasive testing or both) to virtually 100% (panel (c)). The reduction in invasive testing illustrates 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 higher accuracy of the cfDNA screening, many of the pregnancies that were flagged as "positive" for chromosomal abnormalities via the NT screening (i.e. an NT risk score of $\frac{1}{200}$ or greater) now get a more accurate negative prediction from the cfDNA screening; equipped with that additional information, many patients then choose to avoid the invasive test (which is also more costly to the government) and its associated miscarriage risk. However, the net impact on government spending is unclear, since cfDNA screening rises substantially. The model that we develop and estimate below allows us to understand and rationalize these substitution patterns in the data, map demand for cfDNA screening into willingness to pay and patient welfare, and quantify consumer surplus and government spending under observed

26. Rates of cfDNA screening also rise slightly for risk scores outside of the covered range. This reflects a general time trend of increasing use of the new screening technology, which is likely due to growing awareness of and comfort with the technology among patients and physicians. Some of this may be the natural rate of secular diffusion of a new technology, and some may be accelerated by the increased use of the technology in the covered risk score range.

and counterfactual cfDNA coverage policies.

4 An empirical model of prenatal testing choices

All pregnancies in our baseline sample receive an NT screen and the resultant risk score. Patients then face a sequence of decisions about whether to do additional (cfDNA) screening (if it is available), whether to do invasive testing, and whether to terminate the pregnancy. These decisions are therefore interdependent, and in order to account for this interdependence we write down a simple dynamic model that describes the sequence of choices patients face. We then estimate the model using the baseline sample, allowing us to quantify the implications of various counterfactual regimes on birth outcomes, public spending, and consumer surplus.

The key counterfactuals we will examine (in Section 5) are the introduction of cfDNA screening and policies regarding the coverage of the new cfDNA screening technology. That is, we assess whether to offer cfDNA screening for free or to require patients to pay out of pocket, and the extent to which this pricing policy should be targeted to particular patients. All these analyses occur in a setting in which all other medical choices (such as invasive testing and termination) are available for free. The policy choice thus mirrors many similar choices made by public insurance systems regarding the coverage of a new technology in the context of an existing insurance system, taking as given any potential pre-existing price distortions.

Setting and notation. The unit of observation is a pregnancy, denoted by i, which is associated with three possible outcomes: birth of a baby with no chromosomal abnormalities, birth of a baby with chromosomal abnormalities, or no live birth (due to miscarriage or abortion).[27] We normalize the utility of the patient to be zero for a live birth with no chromosomal abnormalities, and denote the (monetized) utility from having a live birth with chromosomal abnormalities by $c_i < 0$ and the (monetized) utility from having no live birth by $a_i < 0$. The assumption that a_i and c_i are negative is made for tractability, and is supported by results from alternative specifications in which a_i and c_i are not restricted. For a_i, the assumption is also natural, as it implies that all the pregnancies in our sample are desired pregnancies, which (as discussed in Section 2) seems likely.

27. This is, of course, a modeling abstraction. Abortion and miscarriage are different. However, our model and data do not allow us to separately identify utility parameters for each. It also seems natural to assume that miscarriage and abortion, both of which result in no live birth, have a more similar impact on the patient than the birth of a baby (with or without chromosomal abnormalities).

We denote by q_i the risk score that all pregnancies receive from the NT screening, and assume that it is known to all patients in our sample. This is a reasonable assumption as clinicians typically discuss the NT results with their patients. We assume that q_i provides an unbiased prediction of the probability that pregnancy i carries a fetus with chromosomal abnormalities.[28] 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 (Kublickas, Crossley, and Aitken 2009). 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.[29] We denote the out-of-pocket cost of cfDNA screening by f_i; this will depend on the NT risk score and on the cfDNA coverage policy regime. For expositional clarity, we omit the i subscripts 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 (binary) 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 to ease the exposition)

$$v_2(p) = \max\{a, pc\}, \tag{1}$$

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 abnormalities is normalized to zero).

28. Recall that q_i is censored from above at $\frac{1}{2}$. We verified that this censoring does not drive any of the results by estimating the model without the small number of pregnancies associated with $q_i = \frac{1}{2}$ and confirming that the results remained essentially the same (not reported).

29. 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% (Oster 2021).

In period 1, the patient chooses whether or not to have an invasive test. Without it, her expected utility is defined by $v_2(p)|_{p=q}$ in equation (1), where q (her NT risk score) denotes her beliefs (absent any additional information). If she does an invasive test, the (definitive) test result implies either $p = 0$ (if negative) or $p = 1$ (if positive). Taken together, and accounting for the miscarriage risk associated with invasive testing (g), expected utility from taking an invasive test is therefore given by

$$v_{1|inv}(p) = ga + (1-g)(pv_2(1) + (1-p)v_2(0)) = ga + (1-g)p\max\{a, c\}, \tag{2}$$

where p is the patient's beliefs about the probability of fetal chromosomal abnormalities.[30] The expected utility from an invasive test thus factors in the risk of miscarriage g and its associated dis-utility a, as well as – in the absence of miscarriage – the utility, $\max\{a, c\}$, from the patient's optimal choice if the fetus is found to have chromosomal abnormalities.

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

$$v_{1|inv}(p) > v_2(p). \tag{3}$$

This optimal decision rule reflects a natural trade-off: the test provides information about the presence of chromosomal abnormalities but entails a risk of losing the pregnancy. The patient 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.

The model yields some intuitive properties. In particular, if $c > a$, so that the patient prefers having a baby with chromosomal abnormalities to losing the baby, it is optimal to never do invasive testing, regardless of p; the test would not affect her decision to keep the baby, so she wants to avoid the miscarriage risk. In contrast, a patient with $c < a$ strictly prefers to abort a fetus with 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. Note that under our assumptions, all women who have an invasive test and detect the presence of chromosomal abnormalities will terminate the pregnancy; this is consistent with the evidence in our setting that 99% of pregnancies are terminated when invasive testing reveals chromosomal abnormalities (Conner et al. 2012).[31] Moreover, no patient will terminate a pregnancy without an

30. In the absence of cfDNA screening, $p = q$. However, we keep the notation general so that it will still apply when we introduce cfDNA screening below.

31. This extremely high termination rate is also consistent with our decision to abstract from potential

invasive test, since the test is free to the patient and – given our modeling assumptions – the miscarriage risk is irrelevant if the patient is otherwise planning to terminate the pregnancy.

Decisions in the presence of cfDNA screening. Once cfDNA screening exists, the patient can choose whether to do cfDNA screening after observing her NT risk score 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 k^{FP} and a false negative rate of k^{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 model (regarding invasive testing) presented above. It is convenient to define $v_1(p) = \max\{v_{1|inv}(p), v_2(p)\}$, which is the period 1 continuation value of the patient, conditional on having (believes of) probability p of fetal chromosomal abnormalities, where $v_{1|inv}(p)$ and $v_2(p)$ were defined in equations (2) and (1), respectively. If the patient chooses no cfDNA screening, then $p = q$ and her expected utility is given by $v_1(p)|_{p=q}$. If she chooses cfDNA screening in period 0, the result provides additional information about the risk of chromosomal abnormalities. We assume Bayesian updating, and denote the patient's posterior (after cfDNA screening) by $p(q,+)$ and $p(q,-)$ for, respectively, positive and negative cfDNA screening results.[32] Naturally, the Bayesian updating depends on the precision of the cfDNA screening as measured by the false positive rate k^{FP} and the false negative rate k^{FN}.[33] Her expected utility from cfDNA screening is therefore given by:

$$v_{0|screen}(q) = \Pr(+ \mid q)(v_1(p)|_{p=p(q,+)}) + \Pr(- \mid q)(v_1(p)|_{p=p(q,-)}) - f, \tag{4}$$

where $\Pr(+ \mid q) = q(1-k^{FN}) + (1-q)k^{FP}$ and $\Pr(- \mid q) = qk^{FN} + (1-q)(1-k^{FP})$ are the ex-ante probabilities (that depend on q) of the two potential outcomes of the cfDNA screening, and the last component, f, is the out-of-pocket cost of cfDNA screening (which depends on q and on the cfDNA coverage regime). Given the expected utility from cfDNA

benefits of knowledge of chromosomal abnormalities even for a patient who wishes to keep the baby (e.g., it may help in preparing for the baby's arrival). Interestingly, in the US, a recent review of studies that reported termination rates for pregnancies with Down syndrome diagnosis showed a substantially lower mean termination rate, 67%, with estimates ranging between 61% and 93% (Natoli et al. 2012).

32. In practice, physicians and genetic counselors help patients interpret screening and test results, and guide them in their decisions. This may be one reason why, as we will see below, the model fits the data remarkably well despite this strong assumption on patients' ability to process information.

33. That is, $p(q,+) = \frac{q(1-k^{FN})}{q(1-k^{FN})+(1-q)k^{FP}}$ and $p(q,-) = \frac{qk^{FN}}{qk^{FN}+(1-q)(1-k^{FP})}$.

screening in equation (4), the patient chooses cfDNA screening in period 0 if and only if

$$v_{0|screen}(q) > v_1(q). \tag{5}$$

Once again, the model yields intuitive properties. In particular, cfDNA screening can only be an optimal choice if the result can affect the invasive testing decision. Moreover, conditional on cfDNA being consequential, one can show that a positive cfDNA test leads to invasive testing and a negative cfDNA test causes the patient to forego invasive testing, which implies a live birth.[34] The logic 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 abnormalities without harming the fetus, but obtaining this information may not be free due to the associated out-of-pocket cost f. We also 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, $v_{0|screen}(q) = v_1(p)|_{p=q}$). We make the (tie-breaking) assumption that patients do not do cfDNA screening when it is inconsequential.[35] 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.

Econometric specification. The key estimable object of interest is the joint distribution of a_i and c_i, which we assume is drawn from a bivariate lognormal distribution. That is,

$$\begin{pmatrix} \log(-a_i) \\ \log(-c_i) \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_a \\ \beta_c \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_c \\ \rho\sigma_a\sigma_c & \sigma_c^2 \end{pmatrix} \right). \tag{6}$$

where the βs, σs, 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 $g = 0.5\%$,[36] that the false positive rate from cfDNA screening k^{FP} and false negative rate from cfDNA screening k^{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

34. Recall that one of the implications of the model is that no patient would terminate the pregnancy without invasive testing.

35. This assumption is consistent with the data. As seen in Figure 3(a), black circles, approximately 10% of pregnancies in the $[\frac{1}{200}, \frac{1}{51}]$ policy regime with risk in the covered range still choose not to do cfDNA screening even though it is free.

36. As discussed in footnote 4, although there have been lower estimates in recent years, it is still the commonly-quoted estimate by physicians and patients in our clinical context. Below we also report results that retain these common beliefs about miscarriage risk, but allow the actual risk to be lower.

screening at $1,248.50 and $567.50, respectively. Costs of invasive testing are always borne by the government. cfDNA screening costs are either borne 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 $567.50.[37]

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, patients who have 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 3(b), there is a large – approximately 40 percentage point – jump in the propensity of invasive testing around an NT score of $\frac{1}{200}$ prior to coverage of cfDNA screening (gray circles). This occurs even though invasive testing is free for everyone in the sample, both below and above $q = \frac{1}{200}$, and presumably reflects the impact of the additional consultation offered to those with a risk score higher than $\frac{1}{200}$. 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 $q_i \geq \frac{1}{200}$, their belief about the probability that the fetus has chromosomal abnormalities is q_i^{ψ} instead of q_i.[38] In other words, we model the impact of the recommendation as equivalent to the patient revising upwards (toward 1) their post-NT screening prior about the probability of chromosomal abnormalities. This helps us fit the jump in invasive testing rates at $q = \frac{1}{200}$ prior to the introduction of cfDNA coverage. After cfDNA coverage is introduced, the sharp jump at $\frac{1}{200}$ is a combination of the same consultation effect and the fact that cfDNA coverage (often) changes discontinuously at $q = \frac{1}{200}$.[39]

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 coverage and by different coverage policy regimes (which vary across regions). We bin the data by risk score into 20 bins, each of width 50.[40,41]

37. In addition to the government's cost of testing, the universal health insurance system also pays for the cost of childbirth and follow-up care for all live births. We abstract from these costs in our calculation. Similarly, we abstract from any benefits of a live birth that are not internalized by the patient.

38. In the robustness analysis below, we also report results from a "reverse" tweak, which assumes that beliefs are correct, at q_i, after the consultation (for $q_i \geq \frac{1}{200}$), but are q_i^{ψ} (for $\psi > 1$ in this case) otherwise.

39. 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.

40. In the robustness analysis below (Section 5.5), we show that results are essentially the same when we use alternative bin widths of 25 and 100.

41. Although we can also observe pregnancy outcomes directly (Table 1), we do not use them as moments in the estimation. Given our sample size, and the fact that the vast majority of pregnancies result in a live birth with no chromosomal abnormalities, the risk score q provides a more accurate estimate of pregnancy

As seen in Appendix Table A1, there are four distinct cfDNA screening policy regimes in our analysis sample. Figure 3 illustrates the moments we match (using a finer bin width of 25) for the most common cfDNA coverage policy, which covers cfDNA coverage for $q \in [\frac{1}{200}, \frac{1}{51}]$. About 80% of our baseline sample is in regions which adopt this coverage policy. The remaining sample is roughly evenly split across regions that adopt three other policy regimes that we observe: covering cfDNA when $q \geq \frac{1}{200}$, when $q \in [\frac{1}{1{,}000}, \frac{1}{51}]$, and when $q \geq \frac{1}{1{,}000}$.

For each policy regime separately, we match the share of pregnancies that received cfDNA screening once cfDNA screening became covered, by NT risk bin; and 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 period.[42]

Thus, overall, we have 180 distinct testing moments that we try to match: 20 bins by 2 outcomes by 4 policy regimes, in addition to 20 bins of the rate of invasive testing prior to the introduction of cfDNA coverage. To do so, we simulate testing decisions using the model and a given set of parameter values, and search for the parameters that 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 associated with each moment. The optimization is done in two steps: first we run a global search, and then a local search. Appendix B.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 sharp jump in the propensity to do 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 chromosomal abnormalities. Patients who want the baby regardless (that is, $c_i > a_i$) prefer to avoid the miscarriage risk, while parents who get a large disutility from having a baby with chromosomal abnormalities relative to not having the baby at all (that is, c_i much smaller than a_i) do invasive testing for sure; the propensity to do invasive testing thus identifies the *relative* importance of a_i and c_i. The extent to which cfDNA coverage (which lowers the cost to the patient from \$567.50 to \$0) increases cfDNA

outcomes than our data. For example, in our baseline sample of more than 30,000 pregnancies, only 88 pregnancies (that is, 29 basis points) result in a live birth with chromosomal abnormalities. To further illustrate this point, Appendix Figure A4 shows the analog of Figure 3 for pregnancy outcomes. The rates of live births with chromosomal abnormalities (panel (b)) or no live birth (panel (c)) are too small to detect any meaningful changes associated with the coverage regime.

42. For estimation purposes, we assume that cfDNA screening rates are zero in the pre-coverage period, when we assume that the cfDNA screening technology is not available. In practice, as seen in Figure 3(a), there is some cfDNA screening that occurs in the pre-coverage period, but we abstract from this in estimation as rates are low and the screening is not widely available.

screening identifies the (monetized) magnitude of these preferences, and the extents to which cfDNA screening and invasive testing rates change with q_i informs the dispersion parameters, the σs and ρ. The individual estimates of the dispersion parameters – the two σs and ρ – are identified by the (lognormal) parametric assumptions. Below we report results from several alternative distributional assumptions and verify that the key conclusions are not overly sensitive to this parametrization.

5 Results

5.1 Model fit and parameter estimates

Figure 5 shows the fit of our model. It plots various testing rates both in the data and as predicted by the model, using the estimated parameters. Specifically, it plots (as a function of the NT risk score q) the probability of an invasive test, both before (panel (a)) and after (panel (b)) the introduction of cfDNA coverage, and the probability of cfDNA screening after it is covered (panel (c)). We note that while we match the moments separately by cfDNA coverage policy regime, for expositional clarity Figure 5 aggregates the estimation moments across the different regimes.[43] The model fits the data remarkably well, especially given its parsimonious parametrization.[44]

Table 2 presents the parameter estimates. They imply that the average monetized dis-utility of losing the baby is approximately \$153,000, and that the average dis-utility of having a baby with chromosomal abnormalities is about \$244,000, or roughly 60% greater (in absolute value). We estimate a moderate level of heterogeneity in both of these (σ_a and σ_c of approximately 0.5 and 0.6, respectively) and a positive correlation between them (ρ of 0.83). We estimate that $c_i > a_i$ for almost 11% of the sample, implying that under perfect information about one-tenth of pregnancies in our sample are carried by patients who would prefer a live birth with chromosomal abnormalities over losing the baby.[45] Finally, we estimate ψ, which creates the impact of the consultation that is triggered by "positive" ($q \geq \frac{1}{200}$) NT results, to be 0.93. This implies, for example, that a patient with an NT risk score of $\frac{1}{200}$ (which is when the consultation kicks in) behaves as if her score is in fact $\frac{1}{138}$.

43. Appendix Figures A5, A6, A7, and A8 present the model fit separately for each cfDNA coverage policy regime.

44. We also note that Figure 5 plots the data and model predictions by bins of risk score with a width of 25, while we estimate the model by matching moments defined by bins of width 50. That is, the prediction of the model for adjacent width-25 bins (within each bin of 50) is not a targeted moment, and this fit is quite reassuring.

45. Note that this estimate applies to the approximately three-quarters of pregnancies that undergo the initial non-invasive NT screening when it is freely available. Our estimates cannot speak to the preferences of the remaining one-quarter of the sample that do not do so, and thus do not receive a risk score.

To provide more intuition for the parameter estimates, the top panel of Figure 6 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 between expected utility conditional on cfDNA screening and expected utility conditional on not receiving cfDNA screening. The figure shows that willingness to pay for cfDNA screening is hump-shaped in q_i. This non-monotone value of information as a function of the appropriateness for the "treatment" (invasive testing) is, as we emphasized in the Introduction, quite different from therapeutic treatments where the value of information is increasing in the probability that the patient will be found appropriate for treatment. To see why the value of information from screening is non-monotonic, recall that for low-risk pregnancies ($q_i < \frac{1}{200}$), most patients would not do invasive testing without the cfDNA screening (see Figure 3(b), gray dots); the main value of cfDNA screening for these patients therefore comes from detecting fetuses that very likely have chromosomal abnormalities, so it is natural that willingness to pay is monotonically increasing in q_i. In contrast, for higher-risk pregnancies, the same figure shows that most patients would do invasive testing absent cfDNA screening. For these patients, therefore, the value of cfDNA screening stems from detecting pregnancies that do not require invasive testing (due to a negative cfDNA result); this is (ex ante) more likely for lower risk pregnancies, which is why the willingness to pay is decreasing in q_i in this range.[46]

The bottom panel of Figure 6 shows how the average willingness to pay for cfDNA screening varies with a_i (holding everything else fixed). For low values of q_i, greater disutility from losing the baby (that is, lower a_i) makes the patient more inclined to avoid invasive testing, which in turn makes the value of cfDNA lower. However, for higher-risk pregnancies many patients expect invasive testing, so the result is reversed, with lower disutility from losing the baby (that is, higher a_i) reducing the value of cfDNA; in this case the value of cfDNA arises from the potential for a negative cfDNA result which would let the patient avoid the invasive test with its miscarriage risk.

With these parameter estimates, we can now use the model to simulate screening and testing decisions and the distribution of pregnancy outcomes – live birth without chromosomal abnormalities, live birth with chromosomal abnormalities, and no live birth – for different NT risk scores q. We will do so under various counterfactual environments regarding the existence of cfDNA and the nature of its coverage. Importantly, in all counterfactuals, the simulated outcome is drawn based on the true probability of a chromosomal abnormality

46. Appendix Figure A9 may be instructive here. It presents the Bayesian updating of patients in response to the cfDNA screening result, as a function of their priors (q_i). A positive result causes quantitatively important updating throughout the distribution of prior beliefs (top panel). By contrast, a negative cfDNA screening result causes quantitatively meaningful updating only for the highest-risk pregnancies (bottom panel).

for a given pregnancy, while the patient's testing decision is made based on her expected utility (i.e., taking into account the "distortion" that arises because of the parameter ψ).[47] Appendix B.2 provides more details about these counterfactual calculations.

5.2 The value of (information) technology

Our first set of counterfactual exercises explores 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-case benchmarks, respectively. The key comparison is between a world with only invasive testing (column (3)) and a world that has cfDNA screening available as well (column (4)); for both we assume that any available technologies are covered and free to the patient.

We present a variety of outcomes in the rows, including testing rates, pregnancy outcomes, government spending, and consumer surplus. While we will discuss several specific findings of interest, our main focus is on how the various counterfactuals affect government spending and consumer surplus – i.e. social welfare – and how they affect the rate of (ex-post) inefficient outcomes. There are two types of such inefficient outcomes, which we shade in gray in the table: a live birth of a pregnancy with chromosomal abnormalities to a patient who would have preferred no live birth, and a termination of a pregnancy with no chromosomal abnormality.

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 risk score q and their preferences. In this hypothetical "worst case," we estimate that 1.10 out of 100 pregnancies would be terminated due to the risk of having a baby with chromosomal abnormalities, thus generating 98.90 live births out of 100 pregnancies.[48] All of the terminations occur among patients who prefer no live birth to a birth with chromosomal abnormalities (i.e. $a_i > c_i$). Half of these terminations (that is, 0.55) 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

47. Recall that patients with a "positive" ($q \geq \frac{1}{200}$) NT result overestimate the true probability of a chromosomal abnormality. For example, our estimate implies that a patient with an NT risk score of $\frac{1}{200}$ (which is when the consultation kicks in) behaves as if her score is in fact $\frac{1}{138}$.

48. This baseline number is artificially much higher than the typical "industry figure" (of, for example, 62 live births per 100 pregnancies (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 spontaneous miscarriages that are not induced by invasive testing (see footnote 29).

is the live birth of babies with chromosomal abnormalities born to patients who would have preferred to terminate the pregnancy had they known (i.e. $a_i > c_i$); we estimate that there are 2.27 such occurrences per 100 pregnancies, accounting for 2.3% of live births and 87% 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 chromosomal abnormalities 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). Therefore, no testing is necessary, and patients only terminate the pregnancy if the fetus has chromosomal abnormalities *and* they prefer no live birth to a baby with chromosomal abnormalities (i.e. $a_i > c_i$). We estimate that 2.82 out of 100 pregnancies end in termination, and 0.33 out of 100 pregnancies result in a live birth with chromosomal abnormalities. By design, in this "first best" scenario there are no inefficient outcomes, but there are fewer live births (97.18 compared to 98.90 in the no post-NT testing scenario in column (1)). The avoidance of inefficient outcomes implies a large increase in consumer surplus relative to the no post-NT testing scenario in column (1), of \$2,670 per pregnancy, or approximately \$95,000 per "affected" pregnancy (i.e. the 2.82 (=2.27+0.55) per 100 pregnancies in column (1) that result in an inefficient outcome).[49]

In column (3) we return to the imperfect information world but 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 \$1,248.50 per test). From the patient perspective, the invasive test provides a definitive diagnosis regarding fetal chromosomal abnormalities, but elevates the risk of miscarriage. We estimate that 23.8% of patients elect to do an invasive test, with the probability of invasive testing increasing in NT 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)). In particular, of the 0.55 fetuses per 100 that do not have chromosomal abnormalities and are "mistakenly" aborted in the absence of post-NT testing (see column (1)), 80% of them now have a live birth; the remaining 0.11 (20%) have no live birth because of the miscarriage risk that accompanies invasive testing. Similarly, invasive testing allows most patients who would prefer to terminate a pregnancy associated with chromosomal abnormalities to do so: of the 2.27 per 100 pregnancies (column (1)) with chromosomal abnormalities that are carried to term even though the patient would

49. Note that our estimate that 3.15 (=2.82+0.33) out of 100 pregnancies have chromosomal abnormalities is occurring in the set of pregnancies with risk scores above $\frac{1}{1,000}$, which, as already discussed in Section 2, consists of only 13% of pregnancies but account for almost all (97%) of pregnancies with chromosomal abnormalities.

have preferred to have no live birth, 92% are now aborted (after they are identified as having chromosomal abnormalities by invasive testing), and only 0.19 per 100 (8%) of them result in a live birth (because the patient preferred to avoid the invasive testing miscarriage risk and so could not know that the fetus had chromosomal abnormalities). Overall, invasive testing eliminates 89% of the inefficient outcomes from the "no post-NT testing" scenario (2.52 out of 2.82) and generates 89% ($2,375 out of $2,670) of the potential consumer surplus from the first best, at a cost of $298 (per pregnancy) to the government.[50]

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 $567.50 per test) and is therefore available to the patient for free; we will 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 (86%) use it.[51] Out of the 86 (per 100) pregnancies that do cfDNA screening, 2.41 have a positive result and therefore follow up with invasive testing (in addition to the 2.52 per 100 who go immediately to invasive testing without cfDNA screening). We thus estimate that the introduction of (covered) cfDNA screening into a world in which cfDNA does not exist reduces the rate of invasive testing from 23.8% (column (3)) to about 5% (column (4)). This decline in invasive testing in turn reduces by 91% (from 0.11 to 0.01 per 100 pregnancies) the rate of inefficient miscarriages from invasive testing (fetuses with no chromosomal abnormalities that are miscarried). Similarly, the improved targeting of the invasive testing (due to the information obtained from the cfDNA screening) reduces the rate of inefficient births (pregnancies with chromosomal abnormalities carried to term by patients who preferred no live birth) by more than 89%, from 0.19 per 100 pregnancies to 0.02. As it turns out, the quantitative impact of both effects is broadly similar, so the rate of live births is barely affected (97.27% in column (3) compared to 97.19% in column (4)).

Our results also imply that introducing covered cfDNA screening reduces the rate of live births with chromosomal abnormalities by 33% (from 0.52 in column (3) to 0.35 in

50. Consistent with these results, invasive testing is widely seen to have revolutionized prenatal care when it was introduced in the 1970s. Indeed, it likely had a bigger impact than we estimate here since it was introduced in a world where, unlike in our column (1), NT screening did not yet exist.

51. The 14% of patients who do not do the cfDNA screening even when it is offered for free are predominantly patients who prefer having a baby with chromosomal abnormalities to no baby ($c_i > a_i$). These patients would always avoid the miscarriage risk associated with invasive testing regardless of the cfDNA result, so they have no value from the cfDNA screening. In addition to these patients, there are two other cases that lead patients to avoid (free) cfDNA. One is when the NT risk score q is sufficiently high to make the patient 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). A second case is when the patient has 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 a false positive).

column (4)). This is interesting in light of public discussion about the possible "end of Down Syndrome" due to the arrival of prenatal screening technologies (Zhang 2020). We estimate a reduction in live births with chromosomal abnormalities far below the 100% envisioned in the ethical debates. While cfDNA allows those who prefer no live birth over a live birth with chromosomal abnormalities to detect and terminate such pregnancies, recall that our estimates also imply that a sizeable share of our sample (almost 11%) would prefer a live birth with chromosomal abnormalities over no live birth.[52] Thus, our results suggest that making cfDNA widely available for free would not eradicate live births with detectable chromosomal abnormalities, although it would (nearly) eradicate *undesired* live births with such abnormalities. Indeed, column (4) shows that under expansive coverage of cfDNA, the vast majority (94%) of live births with chromosomal abnormalities are born to patients who prefer this outcome to no birth.

Taken together, the introduction of covered cfDNA screening into a world with invasive testing (i.e. column (4) vs. column (3)) eliminates 90% of the inefficient outcomes (0.27 out of 0.30) and generates 91% ($\$267 = 2,642 - 2,375$ out of $\$295 = 2,670 - 2,375$) of the potential (remaining) consumer surplus. This comes at an incremental cost of \$250 ($= \$548 - \$298$) (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 cost. We now turn to examining whether more targeted coverage of cfDNA screening can improve on this outcome.

5.3 Optimal targeting

To provide some initial intuition for optimal targeting of cfDNA coverage, Figure 7 plots, by NT risk score, the cfDNA screening rate (top panel) and the invasive testing rate (bottom panel) when that risk score is not covered for cfDNA screening (gray line) and when it is (black line). 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). At low levels of risk, cfDNA screening increases dramatically with coverage, but invasive testing goes up only slightly; this slight increase is due to false positives from cfDNA screening for pregnancies that otherwise would not have bothered with invasive testing. For higher risks, however, coverage of cfDNA screening creates a fair amount of substitution from invasive testing to cfDNA; these represent pregnancies where the risk of chromosomal abnormality is sufficiently high that in the absence

52. Moreover, this estimate only reflects the decisions of the approximately three-quarters of parents who choose to undergo NT screening (see Appendix A), so that the impact of cfDNA screening on the share of live births with chromosomal abnormalities may be even lower in the entire population.

of cfDNA coverage the patient prefers the miscarriage risk associated with invasive testing to paying out of pocket for cfDNA screening, but when cfDNA screening is free she gets the screen and – in most cases – gets a negative result and thus declines the invasive test.

Table 4 explores the impacts of alternative targeting of cfDNA coverage.[53] The first two columns reproduce the results from columns (3) and (4) in Table 3, which provide benchmarks for the extremes of no cfDNA technology (column (1)) and full coverage of cfDNA (column (2)). As already seen, relative to no cfDNA technology, full coverage for cfDNA screening raises consumer welfare but also increases government spending. The remaining columns therefore explore more targeted coverage policies.

We first consider introducing cfDNA technology but without any coverage for it, so that individuals must pay the full cost out of pocket ($567.50). The results are shown in column (3) and a comparison to column (1) – where cfDNA technology does not exist – indicates that the introduction of cfDNA without any insurance coverage is able to both increase consumer surplus (by $48 per pregnancy) and to reduce government spending (by $123 per pregnancy). The increase in consumer surplus comes primarily from the 45% reduction in the inefficient outcome of miscarriages among babies with no chromosomal abnormalities, which falls from 0.11 when there is no cfDNA (column 1) to 0.06 when cfDNA is introduced without coverage, because the cfDNA screening reduces the invasive testing rate (from 24% to 14%). However, compared to full cfDNA coverage (column (2)), the vast majority (0.17) of the 0.19 (per 100 pregnancies) inefficient live births that occur without the cfDNA technology (in column (1)) are not identified and therefore continue to occur when cfDNA is not covered.

The last two columns therefore consider the impact of more targeted coverage of cfDNA screening rather than no coverage (as shown in column (3)) or full coverage (shown in column (2)).[54] In these scenarios, patients whose risk score falls within the covered range have access to cfDNA screening for free; other patients can still access cfDNA but must pay its full cost out of pocket.

Column (4) shows the results of providing cfDNA coverage to patients whose risk score is $\frac{1}{200}$ or higher, which is the risk score at which an NT score is labeled "positive." Compared to no coverage of cfDNA screening (column (3)), there is an increase in cfDNA screening (from 15% to 30%) and a decrease in invasive testing (from 14% to 5%). This decreases both inefficient outcomes. It reduces the rate of live births with chromosomal abnormalities to parents who prefer no birth (from 0.17 to 0.13 per 100 pregnancies) because more patients are

53. 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.

54. Recall that given our construction of the baseline, "full coverage" in practice means covering all pregnancies with NT risk score of $\frac{1}{1,000}$ and above.

doing cfDNA screening and detecting fetuses with chromosomal abnormalities. At the same time, it also reduces the risk of miscarriage of babies without chromosomal abnormalities (from 0.06 to 0.01 per 100 pregnancies) because there is less invasive testing. As a result of both of these effects, consumer surplus increases by \$106 per pregnancy (from \$2,423 to \$2,529) but so too does government spending per pregnancy (from \$175 to \$209), because the decreased spending on invasive testing is not enough to offset the increased spending on (covered) 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 much less common – "type I" technology. 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 \$106 more (per pregnancy) while raising government cost by only \$34, so for any reasonable assumption about the social cost of public funds, total welfare is higher with targeted coverage than with no coverage.

Finally, in column (5) we consider even narrower 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 80% of pregnancies (see Appendix Table A1), but our results indicate that it produces lower consumer surplus and higher government spending than covering any $q_i \geq \frac{1}{200}$ (column (4)). Recall that the rationale for not covering cfDNA screening in the highest risk range ($q_i \geq \frac{1}{50}$) is the assumption that for these highest-risk pregnancies, patients are likely to undertake an invasive test regardless of the result of the cfDNA screening, so that incurring the (non-trivial) cost associated with cfDNA screening would be wasteful. In practice, however, we find that this assumption is flawed. cfDNA coverage of the 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 accidental miscarriages of pregnancies without chromosomal abnormalities (0.01 per 100 pregnancies instead of 0.03) and higher consumer surplus (by \$38 per pregnancy), while saving the government money (\$3 per pregnancy). Consistent with this, the bottom panel of Figure 7 shows that the invasive testing rate is far below 100% even in the top NT risk score bin when all patients are covered for cfDNA.

Figure 8 summarizes these results visually, by representing the five scenarios presented in Table 4 in terms of their impacts on consumer surplus (y-axis) and government spending

(x-axis). Total welfare is higher to the north-west (higher consumer surplus and lower government spending). Relative to cfDNA not existing, 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 of $\frac{1}{200}$ and higher dominates (on both these dimensions) covering risk scores in $[\frac{1}{200}, \frac{1}{51}]$, which is by far the most common policy during our study period (see Appendix Table A1). To evaluate alternative policies where there is no clear dominance (i.e. higher consumer surplus and lower government spending), the figure also shows iso-social welfare curves. We define social welfare per pregnancy as consumer surplus minus $(1 + \lambda)G$, where G denotes government spending and λ denotes the marginal cost of public funds; we use 0.3 as the (standard estimate of) marginal cost of public funds (Poterba 1996). Covering risk scores of $\frac{1}{200}$ and higher leads to greater social 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 $q_i \geq x$, where x varies from $\frac{1}{1{,}000}$ (full coverage) to 1 (no coverage).[55] We report both consumer surplus and government spending, and also a measure of total welfare, which assumes (as earlier) a 0.3 marginal cost of public funds (Poterba 1996).

The results in Figure 9 indicate that targeting is quite important. Total welfare is increasing in coverage until q_i is $\frac{1}{200}$, and then starts declining, indicating that cfDNA coverage for $q_i \geq \frac{1}{200}$ is optimal. At the same time, offering no coverage at all is highly inefficient: many of the highest-risk patients would substitute to invasive testing, thus increasing government cost and reducing consumer surplus. As shown in the bottom panel of Figure 9, the reduction in consumer surplus associated with more stringent cfDNA coverage comes almost entirely from an increase in live births with chromosomal abnormalities to parents who would prefer no birth. In contrast, the low rates of invasive testing (except for the very highest NT risk scores) make the second type of inefficiency – miscarriage (due to invasive testing) of babies with no chromosomal abnormalities – quantitatively small, and not very sensitive to cfDNA targeting except for very high-risk pregnancies.

55. We also explored (not reported) targeting policies that cover cfDNA screening for all pregnancies with $q_i \in [x, y]$, but we consistently found that covering the highest-risk pregnancies (that is, $y = 1$) was always optimal.

5.4 The impact of entirely removing NT screening

Thus far, we have considered the consequences of introducing cfDNA screening and cfDNA coverage in regimes in which NT screening is universally offered as the first (free) step in the screening process. As discussed in Section 2, this type of two-step policy is common in Europe. However, the American College of Obstetricians and Gynaecologists (as of recently) recommends skipping NT screening entirely, and instead making cfDNA screening universally available. In this section, we therefore analyze the implications of this American recommendation.

To do so, we must now analyze decisions and outcomes for the full sample of singleton pregnancies (column (1) in Table 1) rather than our baseline sample, which is selected on the basis of the NT score (since now the NT score is not known).[56] In addition, to consider a policy in which cfDNA screening replaces NT screening, we must model the patient's belief of the probability her fetus has a chromosomal abnormality in the absence of an NT score. It seems likely that patients have considerable uncertainty about the risk of chromosomal abnormalities, both in the overall population and for their particular risk profile. It is also possible that their beliefs are biased. To proceed, however, we make the (strong) assumption that the individual's prior – even in the absence of NT screening – is her NT risk score (i.e. we continue to assume $p = q$ as in Section 4). This assumption presents the most optimistic case for the benefits of the US-style one-step policy relative to the European-style two-step policy.

The results suggest that the US-style policy of full cfDNA screening without any NT screening performs poorly relative to a European-style two-step policy of free NT screening followed by cfDNA coverage targeted at certain NT risk scores. As shown in Appendix Table A7, under the US-style policy, half of the population opts to get cfDNA testing (column (2)), considerably more than the 4% who do so under the optimal European-style policy (column (4)). Because cfDNA is about three times more expensive than NT screening, the increased government spending on cfDNA screening swamps the government savings from foregoing the cost associated with NT testing. As a result, although the additional information provided by the much greater use of cfDNA screening in the US-style policy generates an additional \$24 in consumer surplus per pregnancy relative to the European-style policy, this is more than offset by the increase in government spending of \$272 per pregnancy, thus illustrating the value of using the less expensive NT technology to target the use of the more expensive

56. For completeness, Appendix Tables A6 and A7 replicate the analyses in Table 3 and Table 4 for the full sample. Compared to the (higher-risk) baseline sample, expanding to the full sample unsurprisingly produces much lower testing rates and higher rates of live births, but the comparative statics across different technologies and different cfDNA coverage policies remain qualitatively similar.

cfDNA technology.[57]

5.5 Robustness

In order to investigate the robustness of our results, we explore a set of alternative specifications, which we summarize here. Appendix C provides more details. Each alternative specification described below modifies an assumption in the baseline version. Across all these specifications, the key empirical results are remarkably stable: a full cfDNA coverage increases consumer surplus but also raises government cost, while a targeted coverage increases consumer surplus (by less) and at the same time reduces cost.

Appendix Table A9 reports the corresponding implied estimates of cost and surplus associated with the key cfDNA coverage regimes we consider.[58] First, while our baseline specification aggregates observations into risk score bins of width 50 for estimation, we now also present results for 25-width and 100-width bins. The results are very similar, suggesting that our (arbitrary) assumption on how to aggregate moments is not consequential.

Second, as we discuss in the end of Section 4, the identification of the dispersion parameters of the model (the two σs and ρ) is likely driven by the parametric (lognormal) distributional assumptions regarding the bivariate distribution of a_i and c_i. We therefore report results from three alternative bivariate distributions. Relying on parametric assumption should surely make us apply caution in interpreting these dispersion estimates, but the robustness of the main results to the distributional assumptions suggest that the key conclusions are not overly sensitive to it.

Finally, we perform three additional checks. One specification allows the patient age to shift the mean of a_i and c_i in order to account for the possibility that, since pregnancy chances decline with age, there may be differences across patients by age in the desire to avoid a miscarriage. This has only a small effect on the results (and the age coefficients are statistically insignificant; see Appendix Table A8), which is consistent with out discussion in the end of Section 3.1. A second specification "reverses" the impact of the medical consultation that is triggered by an NT score of $\frac{1}{200}$. In our baseline model, we assume that patients interpret the NT risk score correctly for lower scores, and over-estimate the risk for "positive" NT result (when $q_i \geq \frac{1}{200}$). In the alternative specification, we assume that patients under-estimate the risk for lower NT scores, and interpret it correctly in response to consultation (when $q_i \geq \frac{1}{200}$). The results are not affected much. The third specification

57. Recall that this represents the most optimistic scenario for the US-style policy, given that we assumed that patients know their NT risk score even without going through the NT screening.

58. Appendix Table A8 reports the parameter estimates for many of these alternative specifications. Estimated parameters for the remaining specifications are reported in Appendix C.

retains the model estimates, which are based on the beliefs that invasive testing is associated with a miscarriage probability of $g = 0.005$, but allows the actual outcomes to be driven by lower miscarriage risk of $g = 0.00125$. The impact on the key estimates is small.

6 Summary and implications for other settings

We develop and estimate a simple model of decision making and use it to analyze the welfare gains from coverage of a new (and costly) diagnostic technology that can improve targeting of downstream procedures. Empirically, coverage of the new screening technology for fetal chromosomal abnormalities substantially increases its use, and substantially decreases the use of subsequent invasive testing, which is twice as costly and elevates the risk of miscarriage. The model estimates illustrate that the value of the new technology is largest in the middle of the risk range, where the screening result is most likely to influence decisions regarding subsequent invasive testing. Our counterfactual analyses suggest that narrow targeting of coverage for the screening has the potential to improve patient well-being and reduce government health-care cost, while broader coverage creates the familiar pattern of a new technology increasing both patient well-being and government cost.

A frequent focus of health-care policy – in both the US and other countries – is whether and when to provide coverage for new medical technologies.[59] We study a specific case of a more general policy problem: how to target – i.e., whether and when to cover or recommend – a cheap or non-invasive upfront technology, which can be used to guide the application of a subsequent, more invasive and costly procedure.

The type of problem we study is emblematic of a range of policy questions that all have the same structure. Most closely related to our context are other applications within obstetric and prenatal care. For example, the Oral Glucose Tolerance test for pregnant women (often referred to as the 1-hour glucose challenge test) screens for the risk for gestational diabetes mellitus and can be used to eliminate the need for a more invasive, 3-hour test (NIH, n.d.). However the screen has a high false positive rate of about 30% (Temming et al. 2016), and there is an active policy discussion on whether to skip screening and go straight to diagnostic testing, possibly for high-risk groups only (Benhalima et al. 2016). Likewise for newborns, it is common to administer an invasive diagnostic test for hyperbilirubinemia, a condition that can impair the central nervous system. In recent years, several non-invasive (and thus

59. Different countries employ different decision-making frameworks. Some, like England, try to employ a strict cost-effectiveness criteria, while others, like Sweden and the Netherlands, also explicitly 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).

painless) screening technologies have emerged. Though non-invasive bilirubin screening is not yet routine, its proponents emphasize a sequential use of technologies: widely used non-invasive bilirubin screening, followed by an invasive diagnostic test if the screening result is positive (Hulzebos et al. 2021).

These same trade-offs also arise in the management of acute illness. For example, for hospitalized patients, a pulse oximetry – a probe placed on the finger – is routinely used for patient triage and monitoring (Andrist et al. 2022). It is cheap and non-invasive, but also noisier than more precise, more invasive measures of oxygen saturation such as an Arterial Blood Gas (ABG) test (Sjoding et al. 2020) or readings from an esophageal probe (Vicenzi et al. 2000). The initial choice between these technologies thus reflects a trade-off between cost and patient discomfort on the one hand, and precision of information on the other.

Perhaps the most salient applications of our framework are to cancer, the second-largest cause of death in the United States (Siegel et al. 2022). The trade-offs involved in the design of cancer screening programs have the same structure as in the design of prenatal testing programs. Mammogram screening, for example, is currently recommended based on age, a predictor of breast cancer risk that is noisy, but readily available at no cost (Einav et al. 2020). An active policy discussion concerns the possibility of (also) basing mammogram screening recommendations on breast density, a more precise predictor of risk (Trentham-Dietz et al. 2016). This would require upfront spending on breast density screening, but could eliminate unnecessary mammogram screenings and false positives. Likewise, colonoscopies are currently recommended for all adults in a certain age range to screen for colorectal cancer (CDC, n.d.), but recent years have seen the arrival of a range of non-invasive colorectal cancer screening options – such as stool samples and the use of blood samples to detect cells associated with colorectal tumors in the bloodstream (much like cfDNA) – that can be used to guide the application of the more precise, but also more invasive, colonoscopy. As in the case of prenatal testing, several different colon cancer screening technologies can be used sequentially (Mead et al. 2011; Wu et al. 2013). Yet another example is screening for the BRCA genes that carry an elevated risk of breast and ovarian cancer (King, Levy-Lahad, and Lahad 2014; Long and Ganz 2015). BRCA gene screening is typically not recommended for all women, but instead targeted to women who are at a high risk (close relatives of a known carrier); a universal BRCA screening program would elevate costs upfront, but potentially reduce the number of discoveries of late-stage breast cancer in patients carrying the BRCA gene.[60]

Beyond cancer screening, a similar set of issues arise in the design of cancer treatments.

60. See https://www.cdc.gov/genomics/disease/breast_ovarian_cancer/testing.htm for more details about BRCA gene testing.

The arrival of precision medicine has brought an increased ability to tailor cancer treatment based on biomarkers and genetic testing (Banerjee et al. 2020); each case presents an example where upfront collection of information – at a cost – has the potential to improve downstream treatment decisions. One example is the 21-gene recurrence score, a genetic test that helps identify patients that stand to gain more – or less – from adjuvant chemotherapy to prevent recurrence of breast cancer (Epstein et al. 2015; Wang et al. 2018; Moshfegh 2023).[61]

In all of these types of applications, our findings suggest that appropriate coverage decisions have the potential to improve the cost effectiveness of a technology's use, and even to transform some technologies from the more-common, cost-increasing and health-improving type described by Chandra and Skinner (2012), to their more elusive cost-decreasing and health-improving type. Our analysis also underscores the point that the development of "precision medicine" – which offers the possibility of targeting medical care to the most appropriate patients – in turn creates important policy questions regarding how finely or broadly to screen patients for appropriateness, thus kicking the "precision can" further down the road.

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.

Andrist, Erica, Mark Nuppnau, Ryan P. Barbaro, Thomas S. Valley, and Michael W. Sjoding. 2022. "Association of Race With Pulse Oximetry Accuracy in Hospitalized Children." *JAMA Network Open* 5 (3): e224584.

Armstrong, Katrina. 2012. "Genomics and Health Care Disparities: The Role of Statistical Discrimination." *JAMA* 308 (19): 1979–1980.

61. Not all these technologies produce test results that are binary in nature. The 21-gene recurrence score, for example, ranges from 0 to 100, with a higher score reflecting a higher likelihood of cancer recurrence in the absence of chemotherapy treatment. However, in clinical settings, the test result is often "converted" to a binary outcome. For example, the current guidelines from the American Society of Clinical Oncology denotes patients with a recurrence score of 26 or below as "low risk" – they can "safely" forgo chemotherapy – whereas patients with a score above 26 are denoted "high risk," who are likely to benefit from chemotherapy (Moshfegh, 2023).

Aspinall, Mara G., and Richard G. Hamermesh. 2007. "Realizing the Promise of Personalized Medicine." Section: Competitive strategy, *Harvard Business Review.*

Ball, Philip. 2017. "Designer babies: an ethical horror waiting to happen?" *The Guardian.*

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 (9).

Basu, Anirban, Josh J. Carlson, and David L. Veenstra. 2016. "A Framework for Prioritizing Research Investments in Precision Medicine." *Medical Decision Making* 36 (5): 567–580.

Basu, Anirban, and David Meltzer. 2007. "Value of Information on Preference Heterogeneity and Individualized Care." *Medical Decision Making* 27 (2): 112–127.

Benhalima, Katrien, Peter Damm, André Van Assche, Chantal Mathieu, Roland Devlieger, Tahir Mahmood, and Fidelma Dunne. 2016. "Screening for gestational diabetes in Europe: where do we stand and how to move forward?: A scientific paper commissioned by the European Board & College of Obstetrics and Gynaecology (EBCOG)." Publisher: Elsevier, *European Journal of Obstetrics and Gynecology and Reproductive Biology* 201:192–196.

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.

CDC. 1999. *1999 Fact Sheet - Trends in Pregnancies and Pregnancy Rates.*

———. n.d. *Colorectal Cancer Screening Tests.* Accessed November 16, 2023. https://www.cdc.gov/cancer/colorectal/basic_info/screening/tests.htm.

Chandra, Amitabh, Anupam B. Jena, and Jonathan S. Skinner. 2011. "The Pragmatist's Guide to Comparative Effectiveness Research." *Journal of Economic Perspectives* 25 (2): 27–46.

Chandra, Amitabh, and Jonathan Skinner. 2012. "Technology Growth and Expenditure Growth in Health Care." *Journal of Economic Literature* 50 (3): 645–80.

Conner, Peter, and Peter Malcus. 2017. "Obstetriskt ultraljud och prenatal diagnostik i första trimestern." Läkartidningen.

Conner, Peter, Magnus Westgren, Anna Marsk, Sven Gustafsson, and Marius Kublickas. 2012. "Combined ultrasound and biochemistry for risk evaluation in the first trimester." *Acta Obstetricia et Gynecologica Scandinavica* 91 (1): 34–38.

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.

Devlin, Hannah. 2019. "IVF couples could be able to choose the 'smartest' embryo." *The Guardian.*

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 (1): 52–79.

Epstein, Andrew J., Yu-Ning Wong, Nandita Mitra, Anil Vachani, Sakhena Hin, Lin Yang, Aaron Smith-McLallen, Katrina Armstrong, and Peter W. Groeneveld. 2015. "Adjuvant Chemotherapy Use and Health Care Costs After Introduction of Genomic Testing in Breast Cancer." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 33 (36): 4259–4267.

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." *Acta Obstetricia et Gynecologica Scandinavica* 99 (6): 722–730.

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.

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.

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.

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

Graviditetsregistret. 2020. *Graviditetsregistrets Årsrapport 2020.* Technical report. https://www.medscinet.com/GR/uploads/hemsida/dokumentarkiv/GR%20%C3%85rsrapport%202020%203.0.pdf. Graviditetsregistret.

Hamilton, Barton H., Emily Jungheim, Brian McManus, and Juan Pantano. 2018. "Health Care Access, Costs, and Treatment Dynamics: Evidence from *In Vitro* Fertilization." *American Economic Review* 108 (12): 3725–77.

Hercher, Laura. 2021. "A New Era of Designer Babies May Be Based on Overhyped Science." *Scientific American.*

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." *Inquiry* 37 (3): 282–294.

Hulzebos, Christian V., Libor Vitek, Carlos D. Coda Zabetta, Aleš Dvořák, Paul Schenk, Eline A. E. van der Hagen, Christa Cobbaert, and Claudio Tiribelli. 2021. "Screening methods for neonatal hyperbilirubinemia: benefits, limitations, requirements, and novel developments." *Pediatric Research* 90 (2): 272–276.

Ingvoldstad-Malmgren, Charlotta, Erik Iwarsson, Niklas Juth, and Peter Lindgren. 2017. "SFOG ger nationella riktlinjer för fosterdiagnostik med NIPT." *Läkartidningen.*

Jones, D. R., C. D. Perttunen, and B. E. Stuckman. 1993. "Lipschitzian optimization without the Lipschitz constant." *Journal of Optimization Theory and Applications* 79 (1): 157–181.

King, Mary-Claire, Ephrat Levy-Lahad, and Amnon Lahad. 2014. "Population-based screening for BRCA1 and BRCA2: 2014 Lasker Award." *JAMA* 312 (11): 1091–1092.

Kliff, Sarah, and Aatish Bhatia. 2022. "When They Warn of Rare Disorders, These Prenatal Tests Are Usually Wrong." *The New York Times.*

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." *Acta Obstetricia et Gynecologica Scandinavica* 88 (6): 635–638.

Long, Elisa F, and Patricia A Ganz. 2015. "Cost-effectiveness of universal BRCA1/2 screening: evidence-based decision making." *JAMA oncology* 1 (9): 1217–1218.

Mai, Jan-Frederk, and Matthais Scherer. 2017. *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications.* Hackensack, New Jersey: World Scientific Publishing.

Mead, R., M. Duku, P. Bhandari, and I. A. Cree. 2011. "Circulating tumour markers can define patients with normal colons, benign polyps, and cancers." *British Journal of Cancer* 105 (2): 239–245.

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." *Prenatal diagnosis* 35 (10): 959–967.

Moshfegh, Jasmin. 2023. "Who Gains When Medicine Becomes More Precise? Evidence from Genomic Testing In Breast Cancer." *Stanford University mimeo.*

Natoli, Jaime L., Deborah L. Ackerman, Suzanne McDermott, and Janice G. Edwards. 2012. "Prenatal diagnosis of Down syndrome: a systematic review of termination rates (1995–2011)." *Prenatal Diagnosis* 32 (2): 142–153.

Newhouse, Joseph P. 1992. "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives* 6 (3): 3–21.

NIH. n.d. *Tests & Diagnosis for Gestational Diabetes.* Accessed November 16, 2023. https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/gestational/tests-diagnosis.

Oster, Emily. 2021. *Expecting Better: Why the Conventional Pregnancy Wisdom Is Wrong–and What You Really Need to Know.* Penguin Publishing Group.

Persson, Petra, Xinyao Qiu, and Maya Rossin-Slater. 2021. "Family Spillover Effects of Marginal Diagnoses: The Case of ADHD." *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 (3): 251–257.

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?" *Health Affairs* 37 (5): 710–716.

Poterba, James M. 1996. "Government Intervention in the Markets for Education and Health Care: How and Why?" In *Individual and Social Responsibility: Child Care, Education, Medical Care, and Long-Term Care in America,* 277–308. University of Chicago Press.

Rowan, Thomas Harvey. 1990. "Functional Stability Analysis of Numerical Algorithms." UMI Order No. GAX90-31702. PhD diss.

Sabik, Lindsay M, and Reidar K Lie. 2008. "Priority setting in health care: Lessons from the experiences of eight countries." *International journal for equity in health* 7:4–4.

SBU. 2016. *Fosterdiagnostik med mikroarray för utökad analys av kromosomer.* SBU-rapport 246. Stockholm: Statens beredning för medicinsk och social utvärdering (SBU).

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.*

Siegel, Rebecca L., Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. 2022. "Cancer statistics, 2022." *CA: a cancer journal for clinicians* 72 (1): 7–33.

Sjoding, Michael W., Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, and Thomas S. Valley. 2020. "Racial Bias in Pulse Oximetry Measurement." *New England Journal of Medicine* 383 (25): 2477–2478.

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.* https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/ovrigt/halsodataregister-information-om-nyttan-med-register.pdf.

———. 2021. *Statistik om aborter 2020.* Technical report. Socialstyrelsen.

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." *Acta Obstetricia et Gynecologica Scandinavica* 97 (4): 466–476.

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

Temming, L. A., M. G. Tuuli, M. J. Stout, G. A. Macones, and A. G. Cahill. 2016. "Diagnostic ability of elevated 1-h glucose challenge test." *Journal of Perinatology: Official Journal of the California Perinatal Association* 36 (5): 342–346.

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." *Annals of Internal Medicine* 165 (10): 700–712.

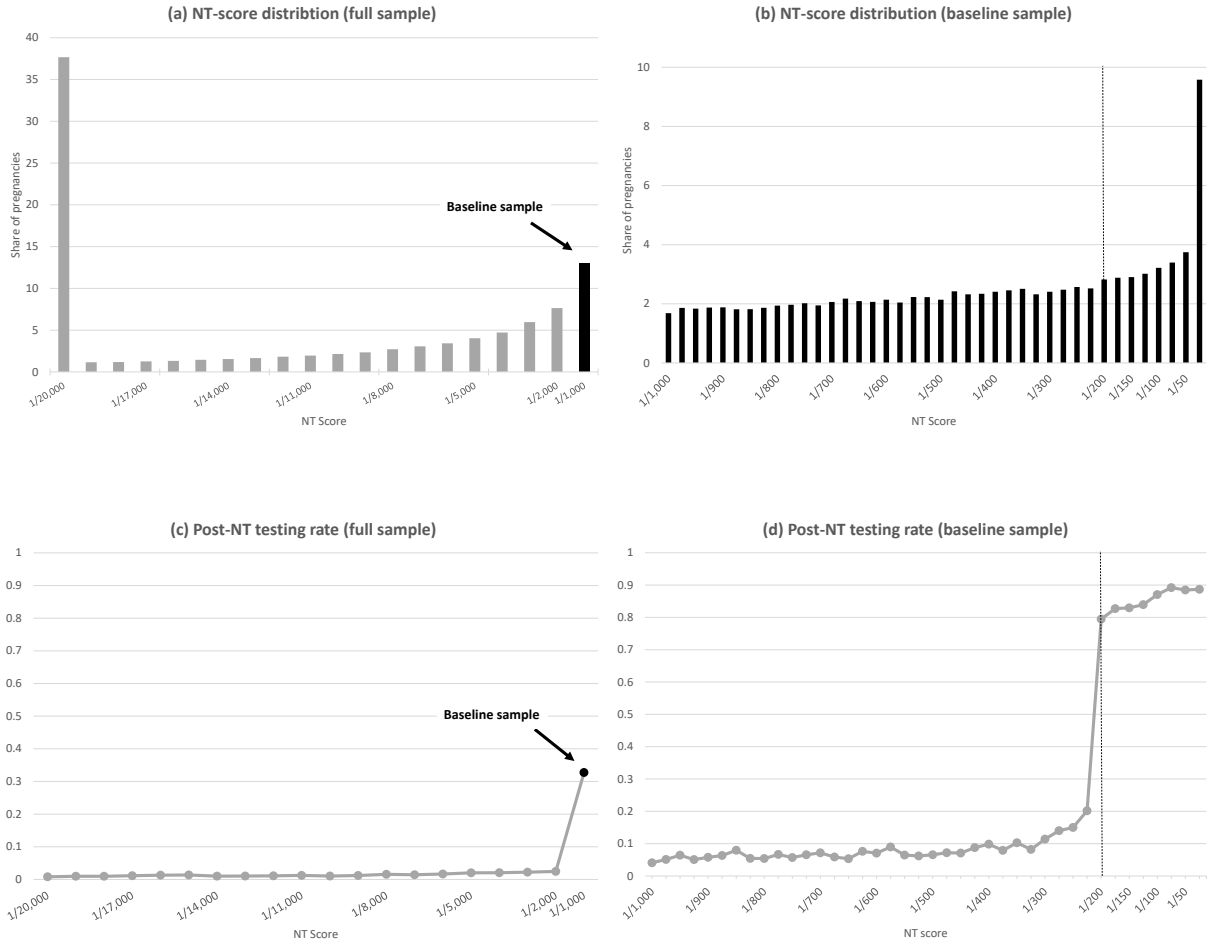
Vicenzi, M. N., H. Gombotz, H. Krenn, C. Dorn, and P. Rehak. 2000. "Transesophageal versus surface pulse oximetry in intensive care unit patients." *Critical Care Medicine* 28 (7): 2268–2270.

Wang, Shi-Yi, Weixiong Dang, Ilana Richman, Sarah S. Mougalian, Suzanne B. Evans, and Cary P. Gross. 2018. "Cost-Effectiveness Analyses of the 21-Gene Assay in Breast Cancer: Systematic Review and Critical Appraisal." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 36 (16): 1619–1627.

Wu, Jen-Chia, Po-Yuan Tseng, Wen-Sy Tsai, Mei-Ying Liao, Si-Hong Lu, Curtis W. Frank, Jinn-Shiun Chen, Han-Chung Wu, and Ying-Chih Chang. 2013. "Antibody conjugated supported lipid bilayer for capturing and purification of viable tumor cells in blood for subsequent cell culture." *Biomaterials* 34 (21): 5191–5199.

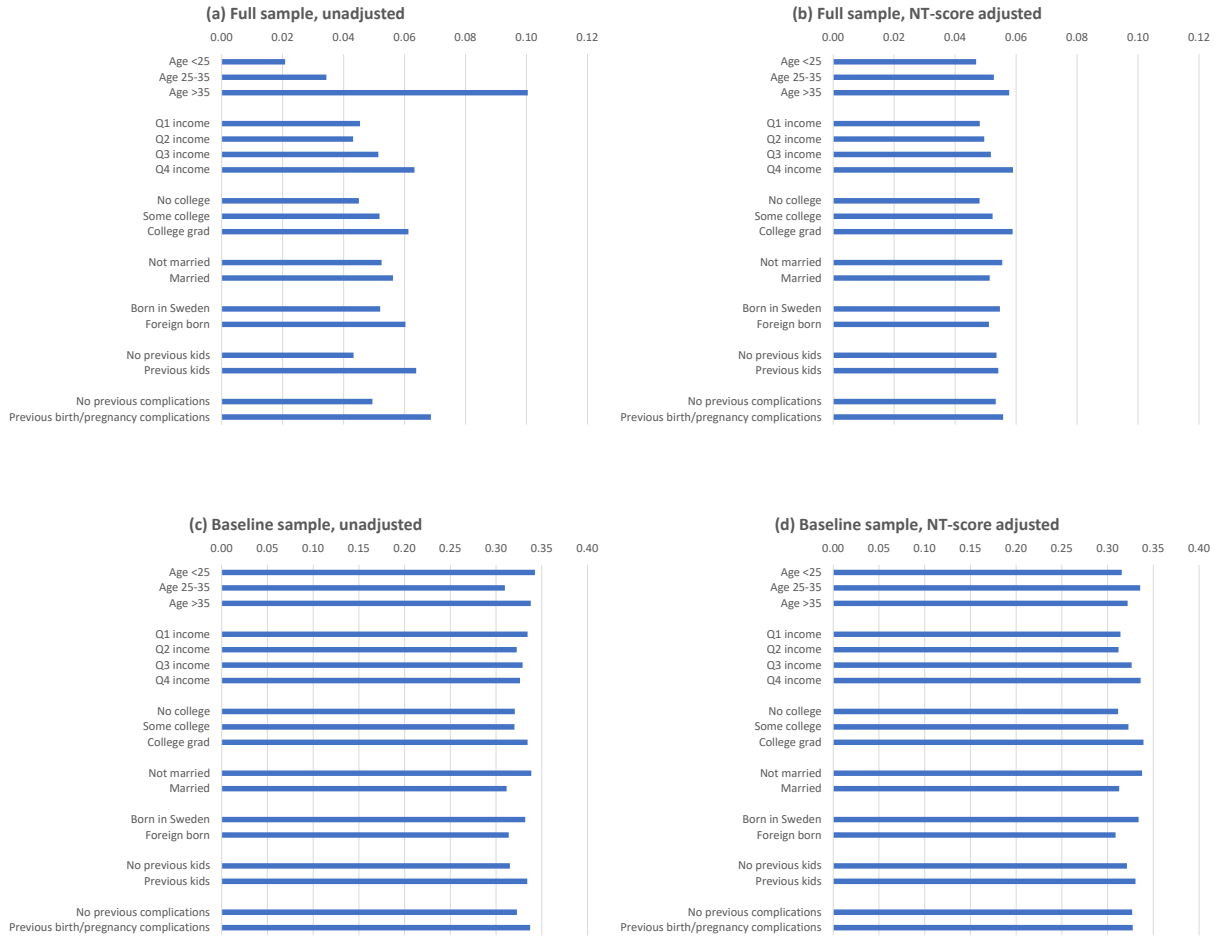
Zhang, Sarah. 2020. "The Last Children of Down Syndrome." *The Atlantic* (December 2020).

Figure 1: NT risk-score distribution and post-NT testing rates



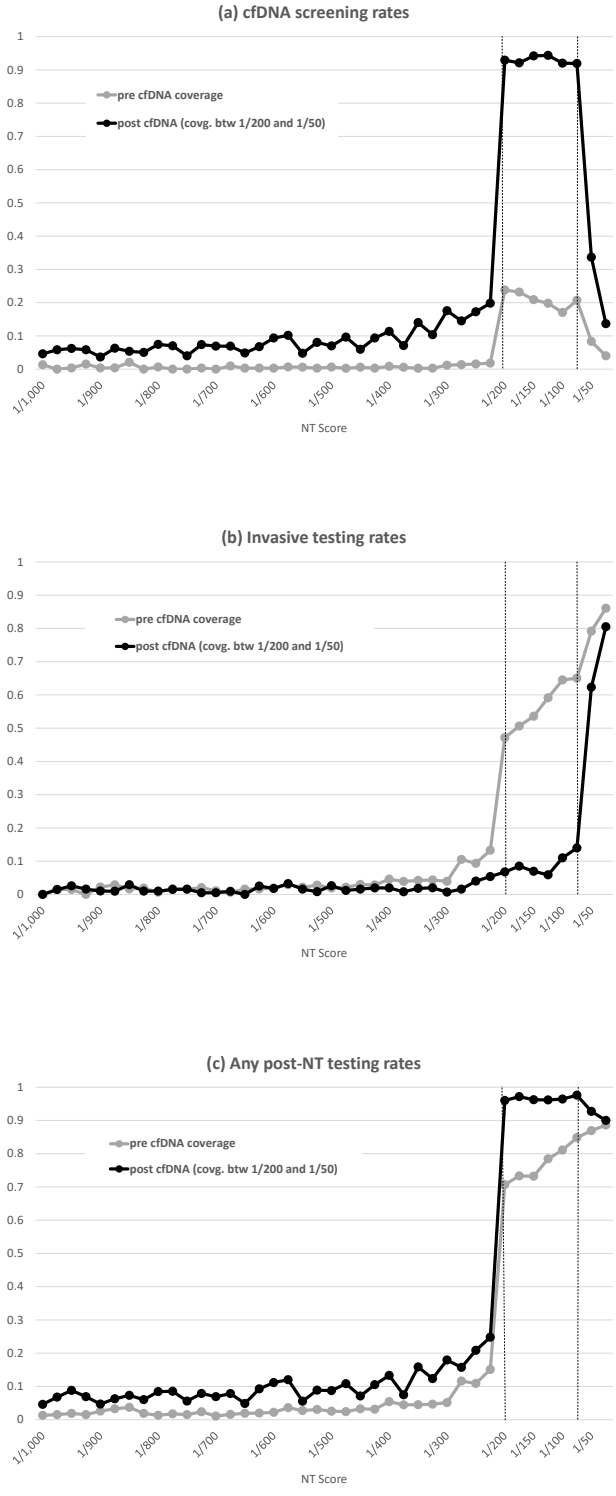
Note: Figure shows the distribution of NT risk scores and post-NT testing rates for the full sample (all singleton pregnancies in universal NT coverage region-months) in panels (a) and (c), and for the baseline sample (all pregnancies in the full sample with NT risk scores of $\frac{1}{1{,}000}$ and higher) in 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 each bin. The vertical line in panels (b) and (d) denotes the risk score above which the NT screening result is considered "positive" and Sweden recommends that the patient be offered the opportunity to discuss follow-up testing. The full sample includes 234,817 pregnancies; the baseline sample includes 30,479 pregnancies.

Figure 2: Post-NT testing rates by demographics



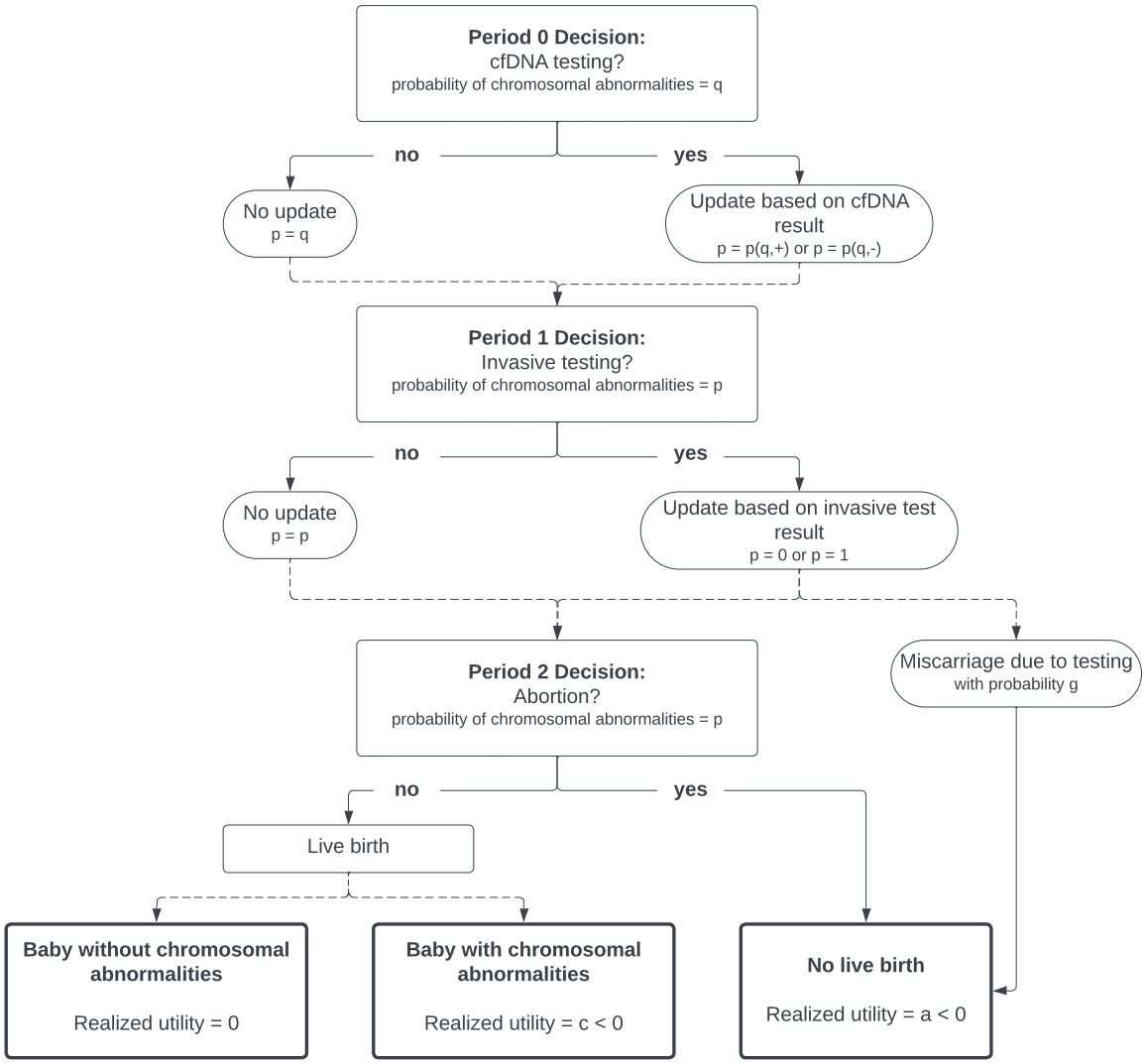
Note: Figure shows post-NT testing rate by various demographics. Panels (a) and (b) use the full sample (N = 234,817), while panels (c) and (d) use the baseline sample (N = 30,479). The two left panels (panels (a) and (c)) report "raw" post-NT testing rates. The two right panels (panels (b) and (d)) show rates conditional on the NT risk score; to do this conditioning, we regress testing rates on indicators for 40 NT risk score bins (of width 50), and then report the average (by demographic characteristics) residual, adding back the overall sample mean.

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



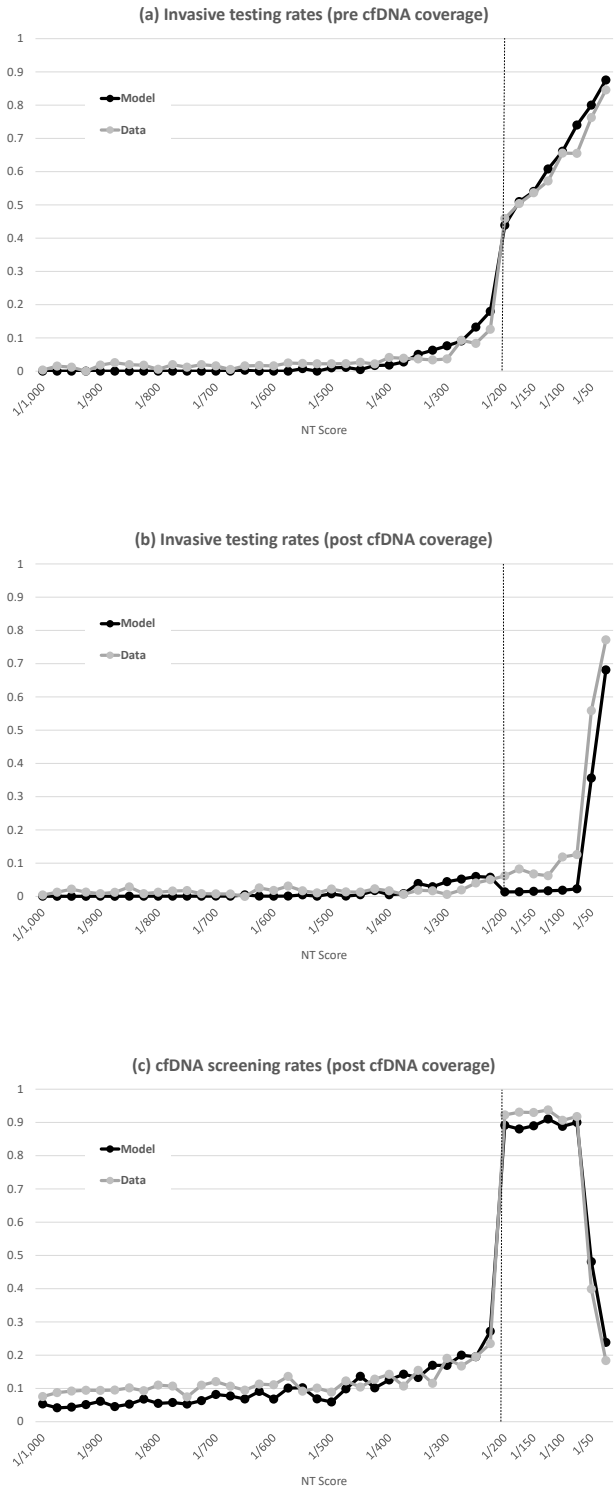
Note: Figure shows testing rates by NT risk score, separately before and after the introduction of coverage for cfDNA, for the subset of the baseline sample that is in regions where the modal cfDNA policy regime is introduced (Table 1, column (3)); this regime covered cfDNA for NT risk scores in $[\frac{1}{200}, \frac{1}{51}]$. 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. $N = 14,010$ pre-cfDNA coverage, $N = 10,722$ post-cfDNA coverage.

Figure 4: Decision tree



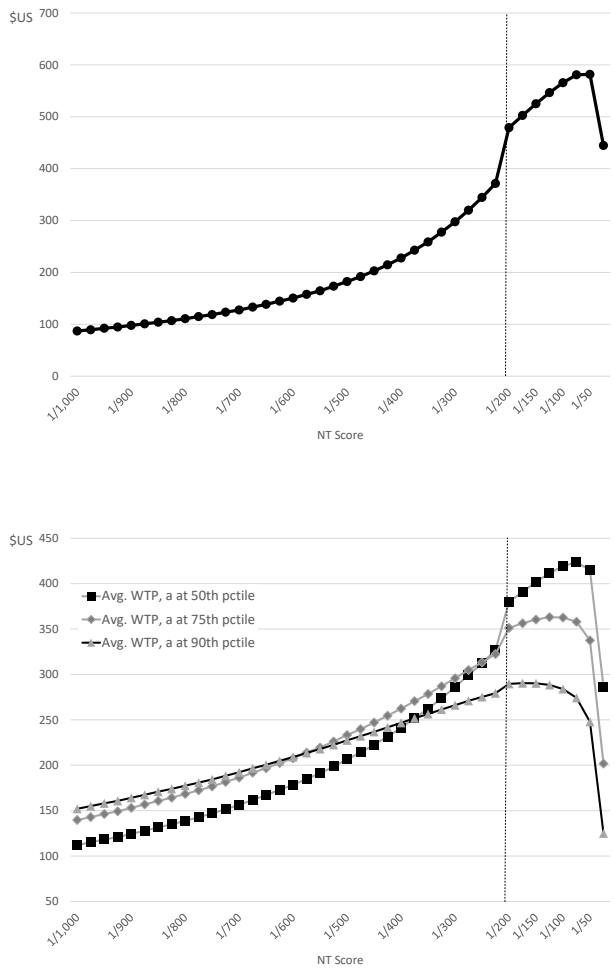
Note: Figure shows the decision tree associated with the model of prenatal testing choices described in Section 4. The bold rectangles indicate terminal outcomes, while the non-bold rectangles indicates individual decisions.

Figure 5: Model fit, pooled across regimes



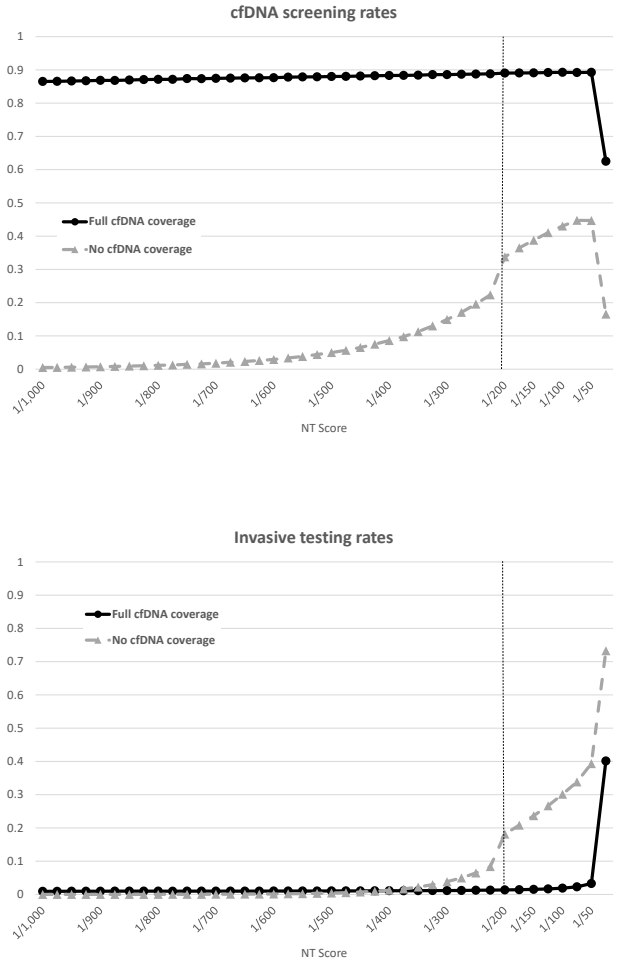
Note: Figure shows testing rates by NT risk score in the data and as predicted by the model, using the estimated parameters from Table 2. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical line denotes the risk score above which the NT screen is 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 Figures A5, A6, A7, and A8 for the fit of the moments separately, by coverage regime). Sample: Baseline sample, N = 30,479.

Figure 6: Willingness to pay for cfDNA



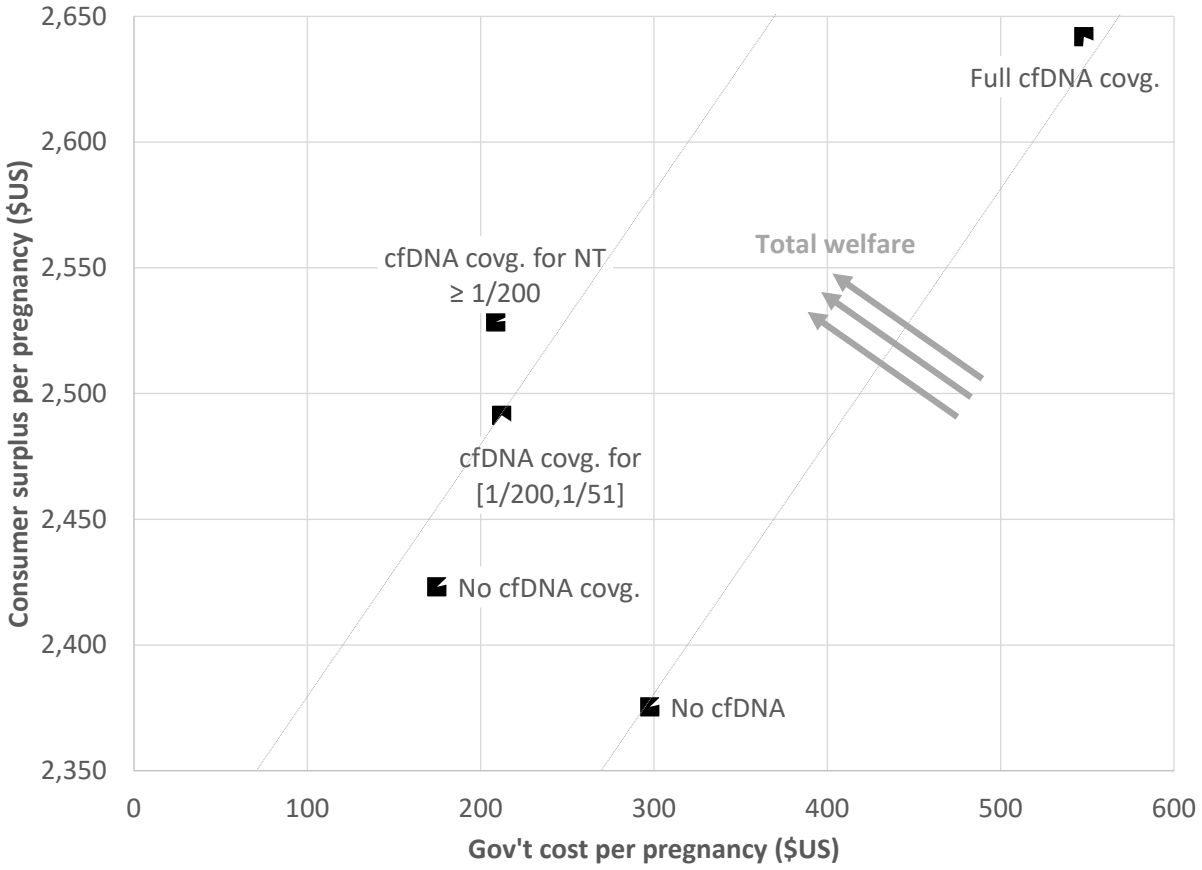
Note: Using the estimated parameters from Table 2, figure shows the average willingness to pay for cfDNA by NT screening risk (top panel), and comparative statics in this average willingness to pay after setting the a_i of all pregnancies to the 50th, 75th, and 90th percentile of the estimated distribution of a_i (bottom panel); a_i is the utility from no live birth.

Figure 7: Counterfactual screening and testing rates with and without cfDNA coverage



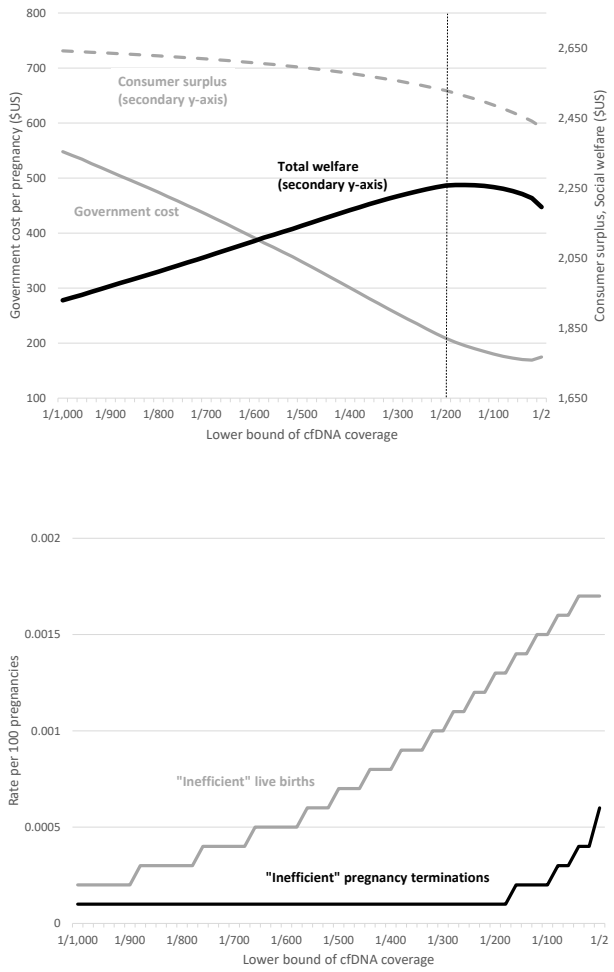
Note: Using the estimated parameters from Table 2, figure shows counterfactual cfDNA screening (top panel) and invasive testing rates (bottom panel) by NT risk under two different cfDNA coverage regimes: no cfDNA coverage and full cfDNA coverage (i.e coverage for all NT risk scores of $\frac{1}{1,000}$ and higher).

Figure 8: Visualizing trade-offs for counterfactual cfDNA coverage policies



Note: Figure plots estimated average (per pregnancy) consumer surplus and government costs of prenatal testing (in $US) under counterfactual cfDNA insurance coverage policies for pregnancies in the baseline sample (i.e. risk score $\geq \frac{1}{1,000}$). "Full cfDNA coverage" denotes coverage for all pregnancies with risk score $\geq \frac{1}{1,000}$, while other scenarios show cfDNA coverage for no risk scores ("No cfDNA coverage,") coverage for risk scores $\geq \frac{1}{200}$ or coverage for risk scores between $\frac{1}{200}$ and $\frac{1}{51}$ (inclusive). In all of these scenarios, patients whose cfDNA screening is not covered have the option to pay for it out of pocket. By contrast, the "no cfDNA" scenario assumes that cfDNA screening is unavailable. Average 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)). The gray lines represent iso-social welfare curves, where social welfare is defined as consumer surplus minus 1.3 times the government cost (thus assuming that the marginal cost of public funds is 0.3).

Figure 9: Optimal cfDNA coverage



Note: Top panel plots the counterfactual government cost of prenatal testing per pregnancy (left y-axis), average consumer surplus per pregnancy (right y-axis), and total welfare per pregnancy (right y-axis), under different potential lower bounds for the NT score at which cfDNA coverage begins. Total welfare is defined as consumer surplus minus 1.3 times the government cost (thus assuming that the marginal cost of public funds is 0.3). The dotted vertical line shows the welfare-maximizing lower bound of the NT score that qualifies for cfDNA coverage. Average 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)). Bottom panel shows counterfactual shares of two types of "inefficient" pregnancy outcomes under the same set of exercises performed in the top panel. "Inefficient" live births are live births with chromosomal abnormalities born to patients 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.

Table 1: Summary Statistics

	Full sample (all singleton pregnancies) (1)	Baseline sample (= NT score ≥ 1/1,000) (2)	cfDNA covg. for NT in [1/200,1/51] (and NT score ≥ 1/1,000) (3)
Number of pregnancies	234,817	30,479	24,732
Demographics:			
Married	0.381	0.425	0.438
Foreign-born	0.226	0.271	0.279
Maternal age:			
Average age	32.0	35.1	35.5
<25	0.063	0.024	0.016
25-35	0.628	0.389	0.368
>35	0.309	0.587	0.616
Household income:			
Average income ($US)	75,960	81,371	86,162
Lowest quartile	0.121	0.108	0.094
Second quartile	0.217	0.179	0.171
Third quartile	0.237	0.234	0.223
Highest quartile	0.424	0.478	0.511
Missing	0.001	0.001	0.001
Education:			
No college	0.354	0.322	0.300
Some college	0.145	0.144	0.146
College graduate	0.488	0.521	0.541
Missing	0.014	0.013	0.013
Any previous children	0.517	0.626	0.618
Any previous pregnancy or birth complications	0.231	0.293	0.292
Miscarriage, stillbirth, pre-term, death w/in 28 days	0.217	0.274	0.274
Congenital deformation or chromsomal abnormalities	0.022	0.031	0.031
Testing:			
Any post-NT testing	0.054	0.327	0.315
cfDNA testing	0.027	0.152	0.141
Invasive testing	0.028	0.184	0.183
Pregnancy outcomes:			
Live birth	0.969	0.918	0.920
Live birth, w/o chrom. abnormalities	0.969	0.915	0.917
Live birth, w/ chrom. abnormalities	0.001	0.003	0.003
No live birth	0.031	0.082	0.080
Ended before 22 weeks	0.028	0.078	0.077
Stillbirth	0.003	0.003	0.003

Note: Table shows summary statistics (means) for maternal characteristics, testing rates, and birth outcomes for the full sample (column (1)), the baseline sample (column (2)), and the sub-sample of pregnancies in the baseline sample that are in regions that ultimately introduce coverage for cfDNA screening for NT scores between $\frac{1}{200}$ and $\frac{1}{51}$ (inclusive, column (3)); Appendix Table A1 describes which regions are associated with each policy regime. Household income is calculated as the average household income across the two years prior to the year of the pregnancy's expected birth date, in SEK CPI-adjusted to 2012. SEK are then converted to USD using a 1 USD to 8.81 SEK exchange rate. Household income quartiles are defined relative to other mothers who give birth in the year of the due date.

Table 2: Parameter estimates

	Coeff.	Std. Err.
β_a	11.799	0.039
β_c	12.220	0.040
σ_a	0.522	0.055
σ_c	0.606	0.023
ρ	0.831	0.013
ψ	0.926	0.009

Note: Table shows the parameter estimates associated with the model described in Section 4. Standard errors are computed using 1,000 bootstrap samples.

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	--	23.83	88.23
cfDNA only	0	--	0	83.29
Invasive only	0	--	23.83	2.52
Both	0	--	0	2.41
Pregnancy outcomes (per 100 pregnancies):				
Live birth	98.90	97.18	97.27	97.19
No chrom. Abnormalities	96.30	96.85	96.74	96.84
Chrom. Abnormalities, a > c	2.27	0.00	0.19	0.02
Chrom. Abnormalities, a < c	0.33	0.33	0.33	0.33
No live birth	1.10	2.82	2.73	2.81
No chrom. Abnormalities	0.55	0.00	0.11	0.01
Chrom. Abnormalities (& a>c)	0.55	2.82	2.62	2.80
Cost and surplus ($US per pregnancy):				
Total government cost	0	--	298	548
cfDNA cost	0	--	0	486
Invasive testing cost	0	--	298	62
Consumer surplus	normalized to 0	2,670	2,375	2,642

Notes: Table shows counterfactual testing decisions and pregnancy outcomes (per 100 pregnancies), and average government spending and consumer surplus (per pregnancy) under alternative assumptions about available technology and information for pregnancies in the baseline sample (i.e. NT risk score $\geq \frac{1}{1,000}$). Consumer surplus and government spending are in $US. Consumer surplus is normalized to zero for the counterfactual with no testing (column 1). Government spending includes invasive testing ($1,248.50 per test) and cfDNA screening ($567.50 per screen). Outcomes that are (ex-post) inefficient are shaded in gray. For ease of exposition, standard errors are reported in Appendix Table A4.

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 [1/200,1/51] (5)
Testing (per 100 pregnancies):					
Any testing	23.83	88.23	28.29	33.10	32.72
cfDNA only	0	83.29	14.30	28.01	23.16
Invasive only	23.83	2.52	13.57	3.35	8.97
Both	0	2.41	0.42	1.74	0.59
Pregnancy outcomes (per 100 pregnancies):					
Live birth	97.27	97.19	97.30	97.30	97.29
No chrom. Abnormalities	96.74	96.84	96.79	96.84	96.81
Chrom. Abnormalities, a > c	0.19	0.02	0.17	0.13	0.14
Chrom. Abnormalities, a < c	0.33	0.33	0.33	0.33	0.33
No live birth	2.73	2.81	2.70	2.70	2.71
No chrom. Abnormalities	0.11	0.01	0.06	0.01	0.03
Chrom. Abnormalities (& a>c)	2.62	2.80	2.64	2.69	2.68
Cost and surplus ($US per pregnancy):					
Total government cost	298	548	175	209	212
cfDNA cost	0	486	0	145	93
Invasive testing cost	298	62	175	64	119
Consumer surplus	2,375	2,642	2,423	2,529	2,491

Notes: Table shows counterfactual testing decisions and pregnancy outcomes (per 100 pregnancies), and average government spending and consumer surplus (per pregnancy) under alternative assumptions about cfDNA coverage for pregnancies in the baseline sample (i.e. NT risk score $\geq \frac{1}{1,000}$). Throughout, invasive testing is assumed to be available for free. Consumer surplus and government spending are in $US. 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 screening ($567.50 per screen). Outcomes that are (ex-post) inefficient are shaded in gray. For ease of exposition, standard errors are reported in Appendix Table A5.

Online Appendix

A Data and variable definitions

A.1 Data

Our data use agreement allows us to observe women (and their children) who were registered in the Swedish Population Register (Skatteverket) 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 out of Sweden before 2000. All linkages across data sets are performed using the mother's individual identifier (which is observed in all data sources).[62]

The backbone of our data is the NT database of pregnancies from 2011 through 2019; it is part of the Swedish Pregnancy Register (Stephansson et al. 2018). It is compiled for the subset of clinics in Sweden which use the more common algorithm to compute NT scores (Graviditetsregistret 2020); see Kublickas, Crossley, and Aitken (2009) for more detailed information about this algorithm, which is calibrated to the Swedish population. The NT database covers about 80% of all NT screenings carried out in Sweden (Graviditetsregistret 2019).[63] Because health care in Sweden is provided by the region (county) in which a woman lives, this clinic-based restriction effectively translates into our observing of NT screening data for the sub-sample of patients who live close to the clinics covered by this database.

We link each mother in the NT database to the Medical Birth Records (MBR) from the National Board of Health and Welfare (Socialstyrelsen 2019) for the years 1985 through 2019. The MBR contains the universe of births in Sweden (both live births and stillbirths) for pregnancies carried 22 weeks or longer (28 weeks or longer for births prior to July 1, 2008). We also link every mother in the NT database to her records on inpatient and specialist outpatient medical care, also obtained from the National Board of Health and Welfare (Socialstyrelsen 2019), from 2001-2019, and to additional information in Swedish administrative data on maternal demographics. Specifically, 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, household income, marital status, and whether she is foreign born from the Statistics Sweden's 2009-2019 longitudinal database of individual administrative records (LISA) (Statistics Sweden, n.d.).

In addition to these existing administrative data, we complied our own data to determine each region's coverage policies for NT and cfDNA over our sample period. To do so, we engaged in e-mail exchanges with representatives of each health-care region during November 2020. We complemented (and cross checked) the information we obtained with information from www.1177.se, a website operated by Sweden's health-care regions that provides information about health-care coverage, and with information from the health-care regions' individual websites and news articles. Appendix Table A1 and Appendix Table A2 show the regions, years, and maternal ages for which cfDNA and NT were covered, respectively. We

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

63. The precise estimate is for 2019; similar estimates are not available for prior years.

determine which prenatal screenings were covered for the mother based on maternal age, the information in the NT database on the clinic's region, and the date of NT screening.

We use the MBR to track the outcomes of each pregnancy in the NT database; we code the pregnancy as terminated if we do not observe it in the MBR (i.e., after week 22 of gestation; we cannot distinguish between miscarriages and abortions). To implement this we check, for each (singleton) pregnancy that receives a screening in the NT database, whether it subsequently appears as a birth in the MBR.[64] Specifically, we start by defining a match (between a screening in the NT database and a subsequent birth in MBR) if the NT screening date is no more than 250 days prior to the date of birth imputed from MBR.[65] The NT screening is generally performed between week 11 (which begins on day 77,[66] 203 days until term) and week 14 (which begins on day 98, 182 days until term) of gestation, but we allow for a longer link period to account for the potential for earlier screenings and post-term births. We confirm that this procedure identifies a maximum of one correct match for each screening. For pregnancies in the NT database that do not have a match in the MBR using this procedure, we perform an additional step. Specifically, if the birth date is within 250 to 270 days of the screening date (i.e. close to our 250-day cutoff but not identified as a correct match in the primary matching step), and the 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.

A.2 Selection into the NT database

Since the point of entry into the NT database is an NT screen, all pregnancies in the data have an NT screen and an NT score. Of course, not all pregnancies receive NT 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 pregnancies that do not. Specifically, we can observe all pregnancies that survive until 22 weeks of gestation or longer in MBR. However, we have no record in the NT database of pregnancies that are terminated or miscarry before 22 weeks if they do not receive NT screening.

In our baseline sample, we limit our analyses to pregnancies in the NT database that are in region-months that provide universal NT coverage (see Appendix Table A2). To get

64. We limit to the approximately 97% of pregnancies that are singleton pregnancies because we cannot distinguish the different fetuses in a multi-fetal pregnancy when linking to the MBR. There are also 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.

65. We only observe the month and year of birth in MBR, not the exact date of birth. However, in MBR we also observe the exact pregnancy due date and the exact gestational age at birth. Using the fact that the due date is calculated as day 280 of gestational age, we impute the exact date of birth using the following formula: due date minus 280 plus gestational age at birth. This method is possible for 95.21% of pregnancies in the MBR. If either the due date or the gestational age at birth is missing in the MBR, we impute the date of birth as the 15th of the (observed) month and year of birth. Using the gestational age at birth method, only 0.54% of the pregnancies in the MBR have imputed birth dates that are not within the reported birth month. In these cases, we still use the birth date calculated from the gestational age.

66. By convention, the first week of pregnancy is denoted as "week 0," the second week as "week 1", and so on.

a rough sense of what share of pregnancies in our region-months receive NT screening, we use our data for 2019 – where we have an estimate that 80% of the NT screens performed in Sweden are in the NT database (Graviditetsregistret 2019) – and inflate the number of NT screens we observe in each region in 2019 by 1.25 to reflect the fact that we only observe 80% of them. We then compare the resulting (inflated) number of linked NT-MBR pregnancies with the total number of pregnancies observed in MBR. We estimate that about 72% of pregnancies in our universal NT coverage sample receive NT screening. This rate is naturally lower if NT coverage is limited by maternal age, which is why we limit our analysis to the universal NT coverage sample.

A.3 Variable definitions

Here we provide more detail on the construction of some of the specific variables in our analysis.

Insurance rules. We assign the pregnancy to a universal NT region-month and to the relevant cfDNA policy regime based on the regime in place for the pregnancy's region at the beginning of the 10th week of gestation, defined as 210 days prior to the pregnancy due date as calculated below.

Pregnancy region. Assigned based on the region of the NT screening clinic.

Pregnancy due date. We calculate a due date for all pregnancies. We start by calculating the due date from the NT database, defined as 280 days after the first day of last menstrual period. This accounts for 99.75% of the pregnancies in our baseline sample.[67] If this date is missing, we assign the due date recorded in MBR (if the pregnancy was carried at least 22 weeks); if the pregnancy is not in MBR, we assign the due date as the NT screening date plus 196 days (which assumes that the NT test occurs at the beginning of the 12th week of pregnancy). The former accounts for 0.21% of our baseline sample, while the latter accounts for 0.04% of our baseline sample.

cfDNA screening. In the NT database, we observe the cfDNA screening if it is done at the same clinic as the NT screen. We do not observe the result (positive or negative) of the cfDNA screening.

Invasive test. In the NT database, we observe if the pregnancy has an invasive test (i.e. CVS or amniocentesis) if the test occurs at the same clinic as the NT screen.

Maternal age. In the NT database, we observe maternal age at the due date.

67. For IVF pregnancies, which are separately flagged in the data, the due date is based on the date of the egg transfer.

Chromosomal abnormality diagnoses. The MBR contains (ICD) diagnosis codes determined at birth. We determine that a child has a chromosomal abnormality if the ICD-10 diagnosis codes contain any of Q90-Q99.

Previous miscarriage/stillbirth/death within 28 days of birth/pre-term live birth. Indicator that is equal to 1 if the mother has had a prior pregnancy that resulted in a miscarriage, stillbirth, death within 28 days of birth, or a pre-term live birth (< 37 weeks). We determine that a mother has had a previous miscarriage if there is a miscarriage recorded (ICD-10 code O00-O03) in either the inpatient or specialist outpatient registers prior to the date of conception of the pregnancy; that is, if there are diagnoses associated with a visit from 2001 (when our patient registry data starts) to the date of conception (either the secondary diagnoses or the main diagnosis) that includes a miscarriage ICD code. We further determine that a mother has had a previous pregnancy that resulted in a stillbirth, death within 28 days of birth, or a pre-term live birth (< 37 weeks) if there is a birth recorded with any of these characteristics in the MBR from 1985 (when our MBR data starts) to the date of conception.

Previous pregnancy with a congenital deformation or chromosomal abnormality. Indicator that is equal to 1 if the mother has a prior pregnancy in MBR with a diagnosed congenital deformation or chromosomal abnormality (any ICD-10 code starting with Q).

Mother's education. Mother's education level (no college, some college, or completed college) measured in the year before the year of the due date. Source: LISA (Statistics Sweden, n.d.).

Income quartile. Maximum of mother's household income quartile in year $t-1$ and $t-2$. Income percentiles are defined relative to other mothers who give birth in year t. For 2020 births, we use 2019 as year t for calculating income (as we only have tax data through 2019. q1 is the lowest quartile, q4 is the highest.[68] Source: LISA (Statistics Sweden, n.d.).

Married. Indicator that is equal to 1 if the mother is married in the year before the year of the due date. Source: LISA (Statistics Sweden, n.d.).

Foreign-born. Indicator that is equal to 1 for mothers that were born outside of Sweden. Source: Statistics Sweden Population Registry (Swedish Research Council 2020).

68. Note that the distribution of income quartiles is skewed towards the highest income quartile because we take the higher percentile across $t-1$ and $t-2$. Defining a mother's income percentile in year $t-k$ as $ptile_{t-k}$, there are approximately 25% of pregnancies in each quartile of the $ptile_{t-k}$ distribution. This is not exactly 25% as a few mothers have two pregnancies in the same calendar year and we take percentiles over mothers, not over pregnancies. However, once we take the maximum quartile a mother was in across more than one year, the distribution of quartiles across mothers will skew higher so long as some women changed quartiles across the years.

Previous children. Indicator that is equal to 1 for pregnancies in which the mother had at least one previous child. A previous child is defined as a live birth in the MBR prior to the current pregnancy, measured from 1985 onwards.

B Model and estimation details

B.1 Estimation details

We estimate the model using method of moments. We assume that (the negative of) a_i and c_i are jointly distributed via a bivariate lognormal distribution, so that

$$\begin{pmatrix} \log(-a_i) \\ \log(-c_i) \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_a \\ \beta_c \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_c \\ \rho\sigma_a\sigma_c & \sigma_c^2 \end{pmatrix} \right). \tag{7}$$

Let $\theta = (\beta_a, \beta_c, \sigma_a, \sigma_c, \rho, \psi)$, where $\psi \in (0, 1]$ is the parameter that captures the impact of the recommendation to take invasive testing following a "positive" NT screening ($q_i \geq \frac{1}{200}$). Let $\aleph$ be the set of relevant observables in the data (q_i, policy regime, pre/post-cfDNA coverage introduced, patient characteristics).

We match four sets of observed moments, $m(\aleph)$. We bin the data by NT risk score into 20 bins, each of width 50 (i.e. pregnancies with $q_i \in [\frac{1}{250}, \frac{1}{201}]$ comprise one of the bins). In the pre-cfDNA coverage period, we match the share of pregnancies that do invasive testing by NT risk bin. In the post-cfDNA coverage period, we match two sets of moments separately for each of the four coverage policy regimes (these regimes vary across regions, see Appendix Table A1 for details on these four regimes): the share who do cfDNA screening, by NT risk bin, and the share who do invasive testing conditional on *not* doing cfDNA, by NT risk bin. Thus, we have 180 distinct testing moments that we try to match: 20 bins by 2 outcomes by 4 policy regimes, in addition to 20 bins of the rate of invasive testing prior to the introduction of cfDNA coverage.

Given a fixed θ, we create corresponding model-simulated moments, $\widehat{m(\aleph \mid \theta)}$, as follows. First, we set the out-of-pocket cost (f_i) of cfDNA for each pregnancy i according to the policy regime in that pregnancy's region. In the period with cfDNA coverage, for those with cfDNA covered, $f_i = \$0$, otherwise $f_i = \$567.50$. If the region had not introduced coverage for cfDNA by the 10th week of the pregnancy, we set the out-of-pocket cost of cfDNA to infinity, imposing that a patient cannot do cfDNA in the pre-coverage period.[69] Second, for each pregnancy i, we draw (a_i, c_i) from the bivariate lognormal distribution defined by θ. We then apply our model to simulate testing decisions for each pregnancy. The model predicts a binary decision of whether or not to receive cfDNA screening and (if not) an invasive testing decision. Finally, we aggregate these testing decisions to testing shares at the NT risk bin by policy regime to calculate the simulated moments $\widehat{m(\aleph \mid \theta)}$.

We use a two-stage non-linear optimization procedure to search for the parameter set $\hat{\theta}$ that minimizes the distance between the observed and simulated moments. In the first stage, we use a global search algorithm, the unscaled Dividing Rectangles method (Jones,

69. As seen in Figure 3(a), there is some cfDNA screening that occurs in the pre-coverage period. We abstract from it in our estimation, as rates are low and the screening is not widely available.

Perttunen, and Stuckman 1993), with a search range of $\beta_a \in [0, 15]$, $\beta_c \in [0, 15]$, $\sigma_a \in [0.1, 10]$, $\sigma_c \in [0.1, 10]$, $\rho \in [-0.9, 0.9]$, and $\psi \in [0, 1]$. In the second stage, we run an implementation of the Subplex algorithm (a variant of the Nelder-Mead algorithm) beginning from the parameter set that achieved the best fit in the first-stage (Rowan 1990). We use a squared distance objective function and weight moments by the number of pregnancies associated with each moment. More formally,

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \quad ||m(\aleph \mid \theta) - m(\aleph)||$$

where

$$||m(\aleph \mid \theta) - m(\aleph)|| \equiv (m(\aleph \mid \theta) - m(\aleph))' W (m(\aleph \mid \theta) - m(\aleph)),$$

where W is the weighting matrix with the number of observations associated with each moment. Standard errors are computed via 1000 bootstrapped samples for the baseline model; to speed up computation, we use 100 bootstrapped samples for all alternative specifications we report.

B.2 Counterfactual simulations

We generate counterfactual policy simulations for our analysis sample using our model of prenatal choices and the estimated parameters $\hat{\theta}$. To reduce simulation error, we create 1000 duplicate observations ($j = 1, 2, ..., 1000$) for each pregnancy i in the analysis sample. For each observation ij we simulate using binomial draws whether that observation has chromosomal abnormalities based on the NT risk q_i. We use this simulated chromosomal abnormality status and the false positive and false negative rates of cfDNA screening to simulate (again, using binomial draws) what the result of a cfDNA screening would be for each observation (if it were to receive cfDNA screening). Finally, we simulate (with a probability of 0.5%) whether an observation would result in a miscarriage if it were to receive an invasive test.

Next, for each observation ij, we draw (a_{ij}, c_{ij}) from the bivariate lognormal distribution defined by $\hat{\theta}$. Combining this draw, its corresponding values of q_i and f_i, and the simulated cfDNA screening result, we use the model to infer optimal choices. With these ij observations in hand, we simply adjust the availability and cost of testing (if the policy precludes a form of testing we set its cost to infinity) and apply our model to simulate the counterfactual policy exercises reported in the paper. If an observation has cfDNA covered, $f_i = \$0$; if cfDNA is available but not covered, $f_i = \$567.50$, and if it is not available we assume that f_i is equal to infinity. With prenatal choices given by the model and the known (simulated) outcomes associated with each duplicate pregnancy, we can read the counterfactual results directly off the (simulated) data.[70]

For each counterfactual, we aggregate these simulated testing decisions and pregnancy

70. The one exception is the "first best (full information)" counterfactual. In this scenario, all observations receive no testing, as the patient already knows the chromosomal abnormality status of the fetus. The patient thus chooses to terminate the pregnancy if and only if $a_{ij} > c_{ij}$ *and* the fetus has chromosomal abnormalities.

outcomes to calculate the estimates in Table 3, Table 4, Figure 7, Figure 8, and Figure 9. To calculate government cost, we assume that the government pays $567.50 for each cfDNA screening received by a covered pregnancy, and $1,248.50 for each invasive test (invasive tests are covered for all pregnancies under all counterfactuals we consider). To calculate consumer surplus, we take the (monetized) utility of each pregnancy outcome (0 if observation ij resulted in a live birth without chromosomal abnormalities, a_{ij} if it ended in a termination, and c_{ij} if it resulted in a live birth with chromosomal abnormalities), subtract any out-of-pocket cost associated with cfDNA screening, and take the (weighted) sum across observations.

C Model robustness exercises

The following section describes several modifications we make to probe the robustness of the key results to certain modeling and estimation assumptions. Appendix Table A8 reports the parameter estimates for many of these exercises (and for others we report the parameter estimates in the text below).[71] Appendix Table A9 reports the implications for the key results associated with the consumer surplus and government cost under various cfDNA coverage regimes.

C.1 Bin sizes for moments

The baseline model estimation described in Section B.1 matches moments in 20 bins of NT risk scores, each of width 50. To investigate whether the choice of bin size does not drive our results, we also estimate the model by matching moments in bins of width 25 (by splitting the bins of width 50 at their midpoints) and width 100 (by aggregating consecutive pairs of bins of width 50).

C.2 Alternative distributions for (a_i, c_i)

In our baseline model, we assume that a_i and c_i have a joint lognormal distribution as specified by equation (6). We can alternatively parametrize the model with the following joint distributions for (a_i, c_i):

1. Joint normal with correlation ρ:

$$\begin{pmatrix} a_i \\ c_i \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_c \\ \rho\sigma_a\sigma_c & \sigma_c^2 \end{pmatrix} \middle| \, a_i < 0 \right). \tag{8}$$

   We impose bounds $\beta_a \in [-300000, 0]$, $\beta_c \in [-300000, 0]$, $\sigma_a \in [1, 100000]$, $\sigma_c \in [1, 100000]$, $\rho \in [-0.9, 0.9]$, and $\psi \in [0.5, 1]$ during estimation. The estimated parameters are $\hat{\beta}_a = -136,205$, $\hat{\beta}_c = -175,681$, $\hat{\sigma}_a = 31,549$, $\hat{\sigma}_c = 11,662$, $\hat{\rho} = 0.188$, and $\hat{\psi} = 0.807$.

71. The very last exercise reported below uses the baseline parameter estimates so is not associated with alternative parameter estimates. Only the counterfactual results change in this case.

2. Exponential a_i and normal c_i with approximate correlation ρ:

$$\begin{aligned} -a_i &\sim \text{Exp}\left(\beta_a^{-1}\right) \\ c_i &\sim N(\beta_c, \sigma_c^2) \\ \text{Corr}(a_i, c_i) &\approx \rho. \end{aligned} \tag{9}$$

We impose bounds $\beta_a \in [1, 900000]$, $\beta_c \in [-900000, 0]$, $\sigma_c \in [1, 900000]$, $\rho \in [-0.99, 0.99]$, and $\psi \in [0.5, 1]$ during estimation. The estimated parameters are $\hat{\beta}_a = 193,855$, $\hat{\beta}_c = -250,491$, $\hat{\sigma}_c = 191,366$, $\hat{\rho} = 0.989$, and $\hat{\psi} = 0.871$.

3. Chi-squared a_i and normal c_i with approximate correlation ρ:

$$\begin{aligned} -a_i &\sim \chi^2(\beta_a) \\ c_i &\sim N(\beta_c, \sigma_c^2) \\ \text{Corr}(a_i, c_i) &\approx \rho \end{aligned} \tag{10}$$

We impose bounds $\beta_a \in [1, 300000]$, $\beta_c \in [-300000, 0]$, $\sigma_c \in [1, 100000]$, $\rho \in [-0.9, 0.9]$, and $\psi \in [0.5, 1]$ during estimation. The estimated parameters are $\hat{\beta}_a = 129,634$, $\hat{\beta}_c = -160,293$, $\hat{\sigma}_c = 22,938$, $\hat{\rho} = -0.010$, and $\hat{\psi} = 0.767$.

We note that specifications (9) and (10) require simulating a_i and c_i with a correlation of approximately ρ from different marginal distributions. To do so, we follow Mai and Scherer (2017) and generate variables x_i and y_i from a standard bivariate normal with correlation ρ. We then transform x_i and y_i into uniform random variables with the normal CDF: $u_i = \Phi(x_i), v_i = \Phi(y_i)$. The joint distribution (u_i, v_i) is a Gaussian copula. We then let the marginal CDFs of a_i and c_i be $F_a(\cdot)$ and $F_c(\cdot)$, and we transform (the uniformly distributed) u_i and v_i using the marginal inverse CDFs of interest: $a_i = F_a^{-1}(u_i), c_i = F_c^{-1}(v_i)$. It is important to note that in doing so, the linear (Pearson) correlation between a_i and c_i will not be exactly ρ because we apply nonlinear transformations to x_i and y_i, but in most cases the correlation will be close to ρ.

C.3 Parametrizing the distribution of (a_i, c_i) with age

In our baseline specification in equation (6), the joint distribution of (a_i, c_i) is not parametrized by any patient characteristics. However, since pregnancy chances decline with age, there may be differences across patients by age in the desire to avoid a miscarriage. To account for this possibility, we can instead estimate a model where (a_i, c_i) varies with age. Specifically, we assume

$$\begin{pmatrix} \log(-a_i) \\ \log(-c_i) \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_a + \gamma_a \cdot \widetilde{\text{age}}_i \\ \beta_c + \gamma_c \cdot \widetilde{\text{age}}_i \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_c \\ \rho\sigma_a\sigma_c & \sigma_c^2 \end{pmatrix} \right), \tag{11}$$

where $\widetilde{\text{age}}_i$ is patient age normalized by subtracting the sample mean (35.12) and dividing by the sample standard deviation (4.84).

C.4 Alternative model for the discontinuity in testing demand

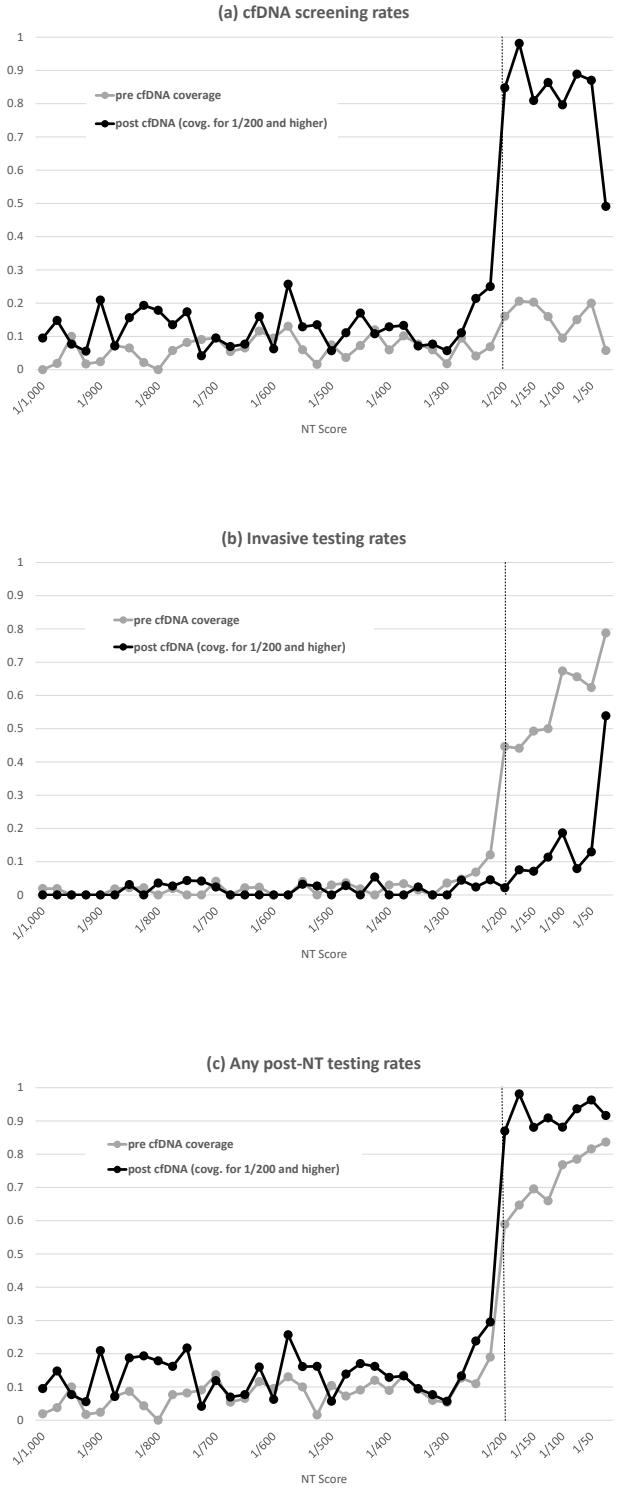
We currently assume that patients overestimate their risks at NT scores of 1/200 and above. Specifically, we model the discontinuous jump in testing demand around the $q = 1/200$ risk threshold by assuming that patients perceive a risk of q^{ψ} for $q \geq 1/200$ with $\psi < 1$.

An alternative modeling choice we can make is to assume that patients with risk scores under 1/200 underestimate their true risks relative to those over 1/200. That is, we assume that patients perceive a risk of $q^{\psi'}$ for $q < 1/200$ with $\psi' > 1$. This alternative assumption is potentially consistent with the fact that patients with positive NT scores receive more counseling from the health care system, and also consistent with a world in which riskier patients are better-informed.

C.5 Alternative miscarriage rate associated with invasive testing

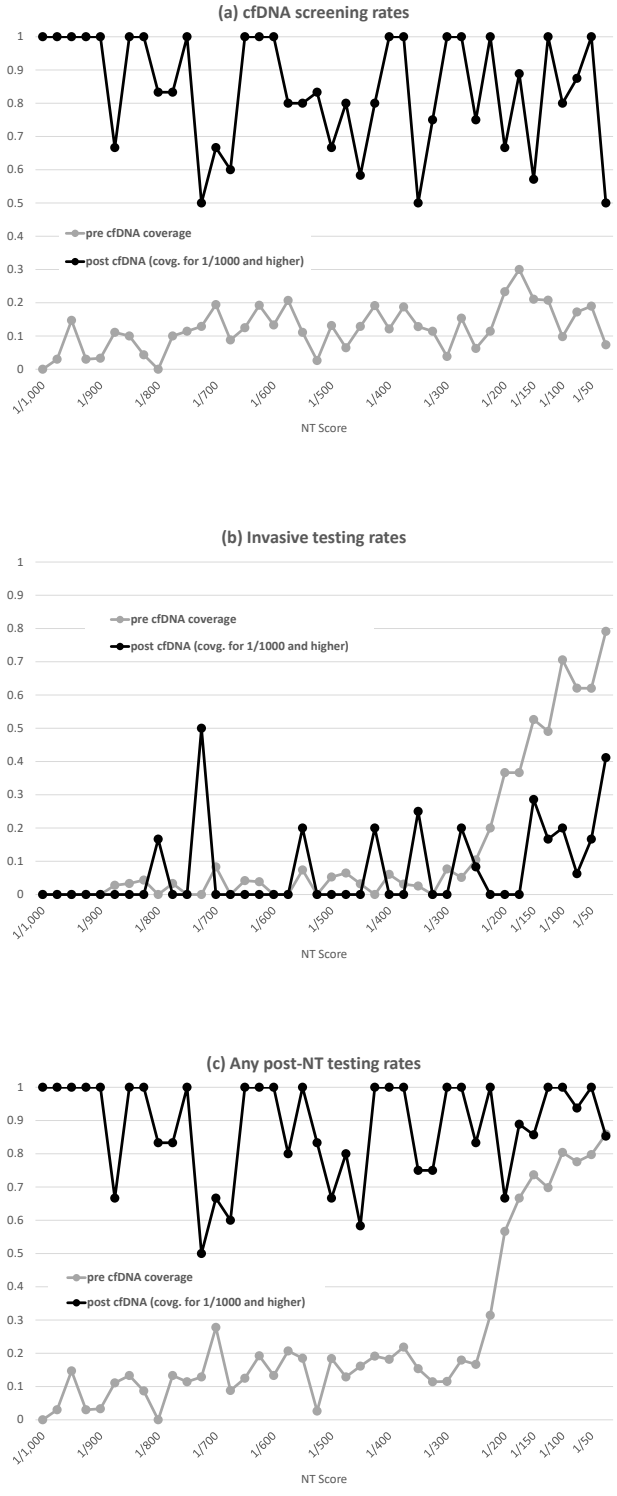
In our baseline estimation, we assume that invasive testing is associated with a known miscarriage rate of $g = 0.005$. However, patients may not be well-informed about the miscarriage risk, and recent estimates for this miscarriage risk are often lower. We can instead assume that the true miscarriage rate (determining pregnancy outcomes) is $g' \neq g$, while patients continue to make decisions based g. Thus, the model estimates remain the same as in the baseline model, but the counterfactual results are generated using $g' = 1/800 = 0.00125$.

Appendix Figure A1: Changes in testing before and after adoption of cfDNA coverage, under cfDNA coverage for $\frac{1}{200}$ and higher



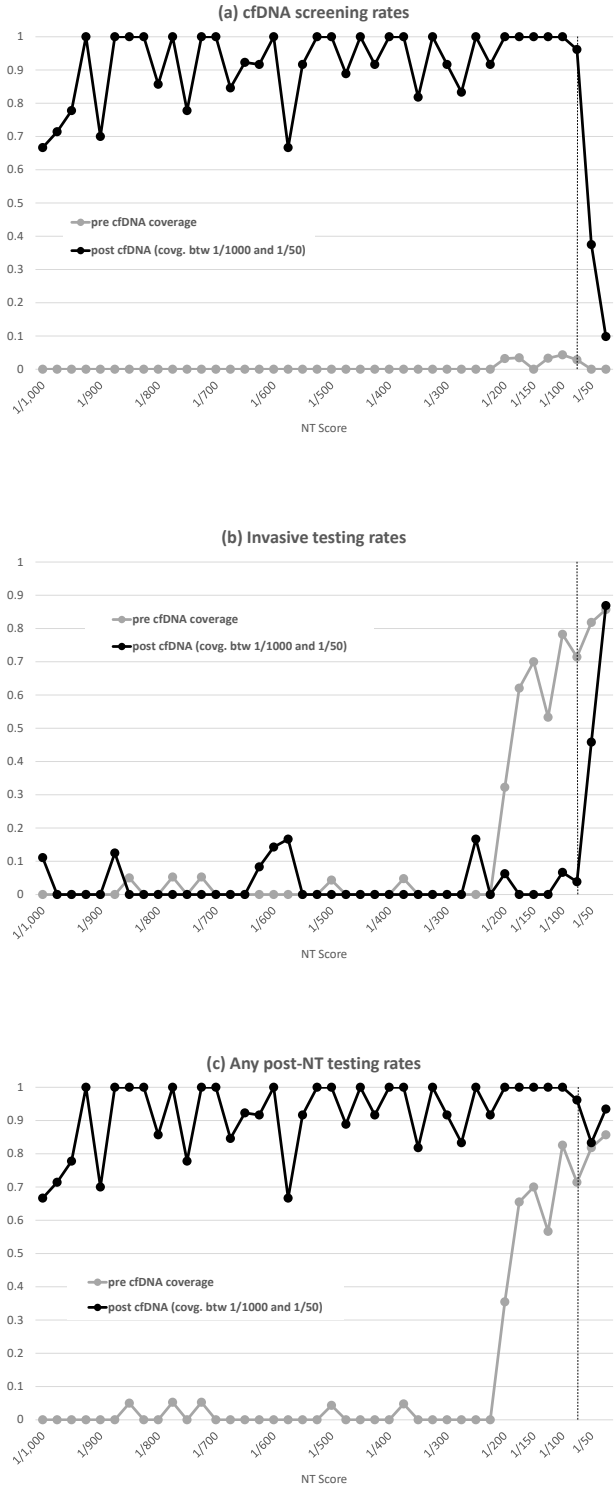
Note: Figure shows testing rates by NT risk score separately before and after the introduction of coverage for cfDNA. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical lines denote the range of risk scores for which coverage of cfDNA screening is introduced. Sample: the subset of the baseline sample that is in regions where the cfDNA policy regime is introduced, covering cfDNA for NT risk score of $[\frac{1}{200}]$ and higher. $N = 2708$ pre-cfDNA coverage, $N = 1649$ post-cfDNA coverage.

Appendix Figure A2: Changes in testing before and after adoption of cfDNA coverage, under cfDNA coverage for $\frac{1}{1{,}000}$ and higher



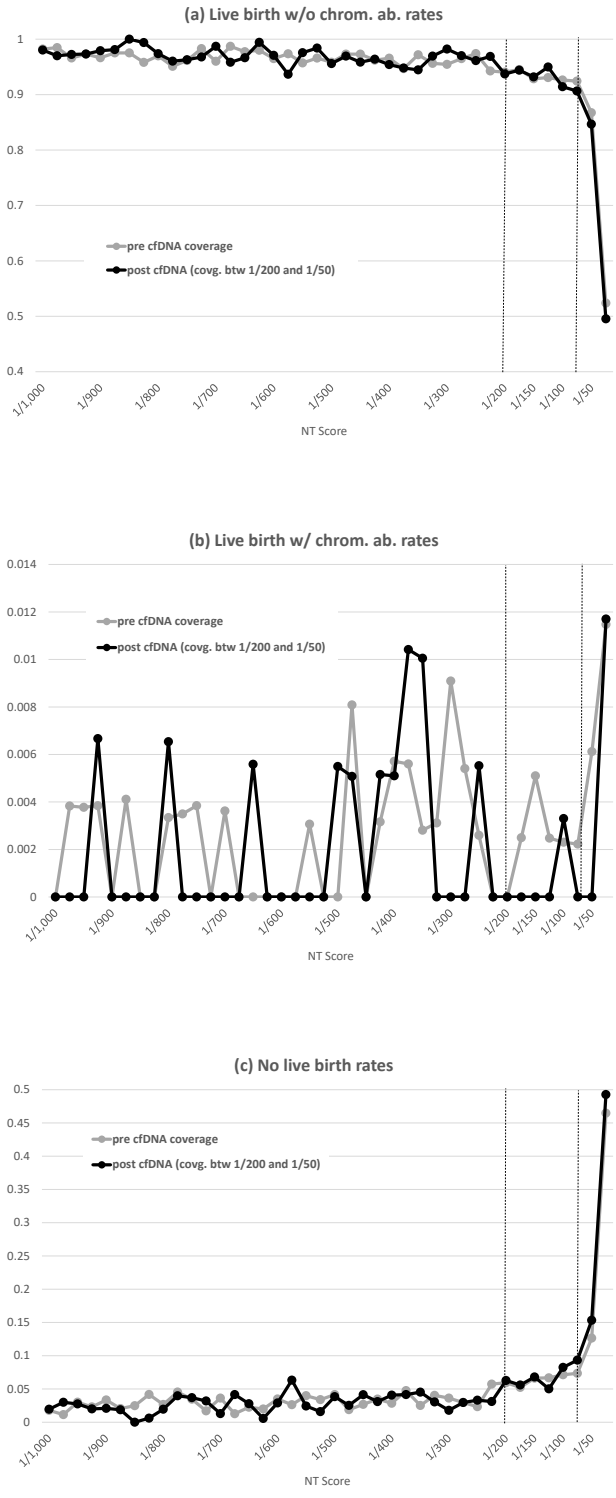
Note: Figure shows testing rates by NT risk score separately before and after the introduction of coverage for cfDNA. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical lines denote the range of risk scores for which coverage of cfDNA screening is introduced. Sample: the subset of the baseline sample that is in regions where the cfDNA policy regime is introduced, covering cfDNA for NT risk score of $[\frac{1}{1{,}000}]$ and higher. $N = 1540$ pre-cfDNA coverage, $N = 241$ post-cfDNA coverage.

Appendix Figure A3: Changes in testing before and after adoption of cfDNA coverage, under cfDNA coverage in $[\frac{1}{1{,}000}, \frac{1}{51}]$



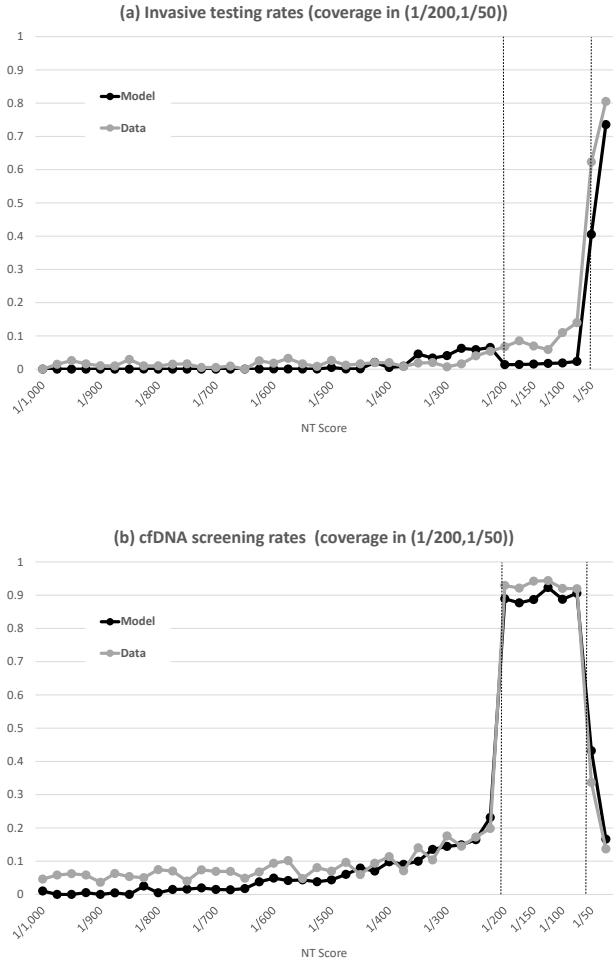
Note: Figure shows testing rates by NT risk score separately before and after the introduction of coverage for cfDNA. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical lines denote the range of risk scores for which coverage of cfDNA screening is introduced. Sample: the subset of the baseline sample that is in regions where the cfDNA policy regime is introduced, covering cfDNA for NT risk score in $[\frac{1}{1{,}000}, \frac{1}{51}]$. $N = 919$ pre-cfDNA coverage, $N = 471$ post-cfDNA coverage.

Appendix Figure A4: Changes in pregnancy outcomes, before and after adoption of cfDNA coverage



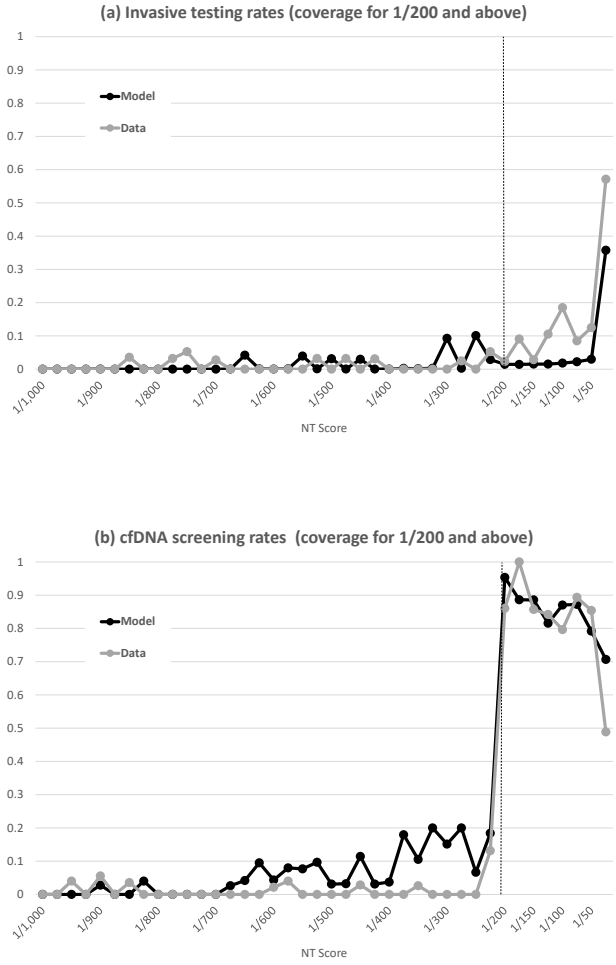
Note: Figure shows pregnancy outcomes by NT risk score separately before and after the introduction of coverage for cfDNA. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Vertical lines denote the range of risk scores for which coverage of cfDNA screening 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 = 24,732.

Appendix Figure A5: Model fit, cfDNA coverage in $[\frac{1}{200}, \frac{1}{51}]$



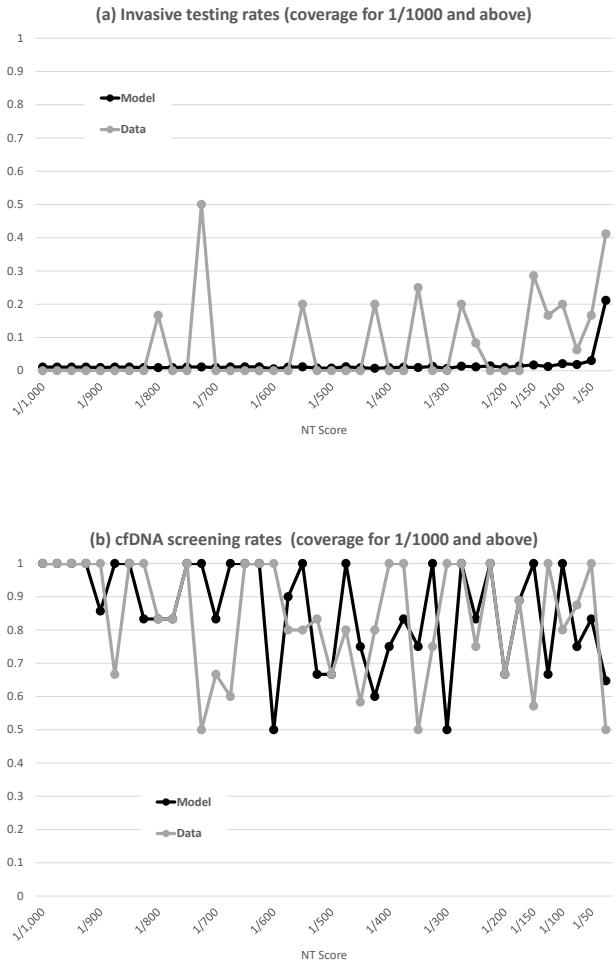
Note: Figure shows testing rates by NT risk score in the data and as predicted by the model, using the estimated parameters from Table 2, for observations that are in the $[\frac{1}{200}, \frac{1}{51}]$ cfDNA coverage regime. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Panel (a) shows invasive testing rates and panel (b) shows cfDNA screening rates. N = 10,722.

Appendix Figure A6: Model fit, cfDNA coverage for $\frac{1}{200}$ and above



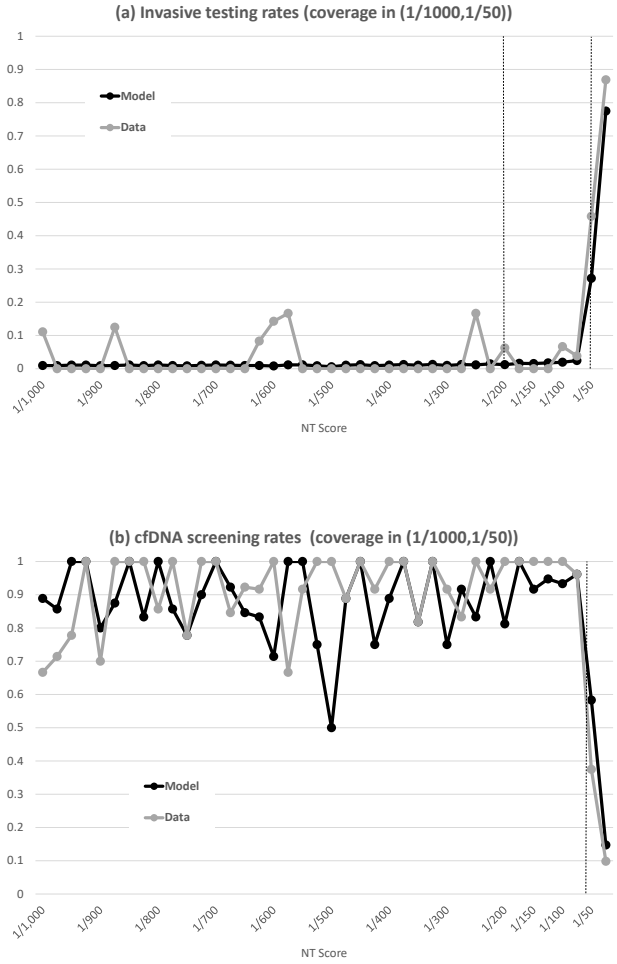
Note: Figure shows testing rates by NT risk score in the data and as predicted by the model, using the estimated parameters from Table 2, for observations that are in the "$\frac{1}{200}$ and above" cfDNA coverage regime. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Panel (a) shows invasive testing rates and panel (b) shows cfDNA screening rates. $N = 1{,}408$.

Appendix Figure A7: Model fit, cfDNA coverage for $\frac{1}{1{,}000}$ and above



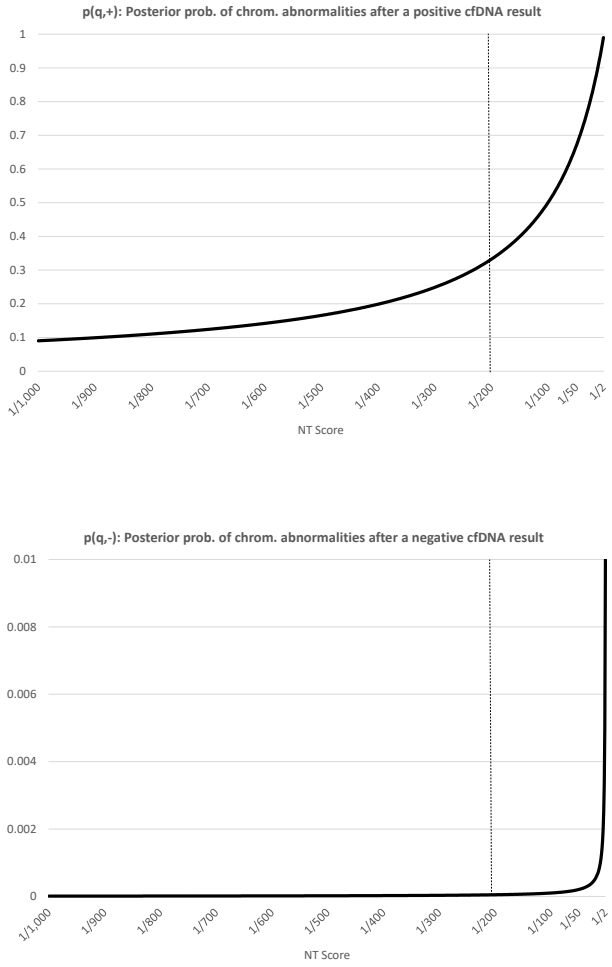
Note: Figure shows testing rates by NT risk score in the data and as predicted by the model, using the estimated parameters from Table 2, for observations that are in the "$\frac{1}{1{,}000}$ and above" cfDNA coverage regime. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Panel (a) shows invasive testing rates and panel (b) shows cfDNA screening rates. $N = 241$.

Appendix Figure A8: Model fit, cfDNA coverage in $[\frac{1}{1{,}000}, \frac{1}{51}]$



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, for observations that are in the $[\frac{1}{1{,}000}, \frac{1}{51}]$ cfDNA coverage regime. Each bin has a risk score width of 25, with the x-axis labels showing the lower end of that bin. Panel (a) shows invasive testing rates and panel (b) shows cfDNA screening rates. $N = 471$.

Appendix Figure A9: Posterior beliefs after cfDNA screening results



Note: Figure shows the posterior beliefs (see footnote 33) about the probability of chromosomal abnormalities, by NT risk score, for both positive cfDNA results (top panel) and negative ones (bottom panel).

Appendix Table A1: cfDNA coverage regimes

cfDNA coverage regime	Region	Month introduced	Share of baseline sample
[1/200,1/51]			0.811
	Stockholm	10/2016	
	Sörmland	09/2016	
	Gotland	02/2016	
	Halland	05/2017	
	Västra Götaland	02/2017	
	Västmanland	02/2017	
	Dalarna	11/2017	
	Jämtland	09/2017	
1/200 and above			0.085
	Uppsala	09/2016	
	Värmland	04/2018	
	Gävleborg	02/2017	
	Västernorrland	09/2017	
[1/1000,1/51]			0.046
	Kronoberg	05/2017	
	Skane	05/2017	
1/1000 and above			0.058
	Örebro	10/2018	
[1/300,1/51]			0.000*
	Östergötland	01/2016	
	Jönköping	06/2018	
	Kalmar	09/2017	
Age 32 and older			0.000*
	Blekinge	06/2016	
No cfDNA coverage			0.000
	Västerbotten	NA	
	Norrbotten	NA	

Note: This table shows the cfDNA coverage regimes in Sweden. For each regime, we list the health-care regions that adopted the regime, the date of introduction of cfDNA coverage in each of these regions, and the share of our baseline sample that is accounted for by pregnancies in the regions that adopted the regime.
*No pregnancies in the three regions that adopted cfDNA coverage in the range $[\frac{1}{51}, \frac{1}{300}]$ are included in the NT database; similarly, no pregnancy in the region that adopted cfDNA coverage for patients aged 32 or older is included in the NT database. (As per the discussion in Appendix A, this stems from the fact that the NT clinics in these regions do not use the dominant algorithm to calculate the NT score.) Consequently, no pregnancies from these regions enter our baseline sample.

Appendix Table A2: NT coverage policies

NT coverage regime	Region	Months (within 2011-19)	Share of pregnancies in 2011-19	Share of pregnancies in 2019
Universal NT coverage			0.477	0.684
	Stockholm	01/2011 - 12/2019		
	Uppsala	01/2016 - 12/2019		
	Östergötland	01/2011 - 12/2019		
	Jönköping	01/2011 - 12/2019		
	Kronoberg	03/2012 - 12/2019		
	Kalmar	01/2012 - 12/2019		
	Skane	03/2018 - 12/2019		
	Dalarna	11/2017 - 12/2019		
	Halland	05/2017 - 12/2019		
	Värmland	01/2011 - 12/2019		
	Örebro	01/2011 - 12/2019		
	Västmanland	02/2017 - 12/2017		
	Västernorrland	01/2015 - 12/2019		
	Jämtland	04/2016 - 12/2019		
Age 35 and higher			0.428	0.281
	Uppsala	01/2011 - 12/2015		
	Sörmland	01/2012 - 12/2019		
	Gotland	01/2011 - 12/2019		
	Skane	01/2011 - 02/2018		
	Västra Götaland	01/2011 - 12/2019		
	Västmanland	01/2018 - 12/2019		
	Dalarna	01/2011 - 10/2017		
	Gävleborg	01/2011 - 12/2019		
	Västernorrland	03/2012 - 12/2014		
	Jämtland	01/2011 - 03/2016		
	Västerbotten	01/2011 - 12/2019		
Age 38 and higher			0.009	0.000
	Blekinge	01/2011 - 06/2016		
No NT coverage			0.086	0.035
	Sörmland	01/2011 - 12/2011		
	Kronoberg	01/2011 - 02/2012		
	Kalmar	01/2011 - 12/2011		
	Halland	01/2011 - 04/2017		
	Västmanland	01/2011 - 01/2017		
	Västernorrland	01/2011 - 02/2012		
	Blekinge	07/2016 - 12/2019		
	Norrbotten	01/2011 - 12/2019		

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

Appendix Table A3: The impact of sample restrictions

	All live births (1)	All live births in region-months w/ universal NT coverage (2)	NT screened pregnancies in region-months w/ universal NT coverage (3)
Number of pregnancies	973,673	458,054	234,817
Demographics:			
Married	0.415	0.431	0.381
Foreign-born	0.265	0.276	0.226
Maternal age:			
Average age	30.4	30.9	32.0
<25	0.128	0.109	0.063
25-35	0.652	0.649	0.628
>35	0.220	0.242	0.309
Household income:			
Average income ($US)	58,654	65,475	75,960
Lowest quartile	0.191	0.167	0.121
Second quartile	0.229	0.223	0.217
Third quartile	0.246	0.241	0.237
Highest quartile	0.310	0.348	0.424
Missing	0.024	0.021	0.001
Education:			
No college	0.455	0.427	0.354
Some college	0.131	0.135	0.145
College graduate	0.379	0.406	0.488
Missing	0.035	0.032	0.014
Any previous children	0.530	0.525	0.517
Any previous pregnancy or birth complications	0.220	0.222	0.231
Miscarriage, stillbirth, pre-term, death w/in 28 days	0.204	0.206	0.217
Congenital deformation or chromsomal abnormalities	0.025	0.024	0.022

Note: Table shows how the sample restrictions affect sample composition. Column (3) presents summary statistics for our full sample, replicating column (1) of Table 1 in the main text. Column (1) and (2) in the above table use the Medical Birth Records to report statistics on all (singleton) live births in Sweden (column (1)) and then on those limited to the region-months with universal NT coverage (column (2)). Note that some pregnancies in column (3) will not appear in column (2) since not all pregnancies with an NT screen result in a live birth.

Appendix Table A4: Outcomes under counterfactual information and technology with standard errors

	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	--	23.83	88.23
	--	--	(0.42)	(0.77)
cfDNA only	0	--	0	83.29
	--	--	--	(0.74)
Invasive only	0	--	23.83	2.52
	--	--	(0.42)	(0.08)
Both	0	--	0	2.41
	--	--	--	(0.04)
Pregnancy outcomes (per 100 pregnancies):				
Live birth	98.90	97.18	97.27	97.19
	(0.06)	(0.06)	(0.06)	(0.06)
No chrom. Abnormalities	96.30	96.85	96.74	96.84
	(0.08)	(0.06)	(0.06)	(0.06)
Chrom. Abnormalities, a > c	2.27	0.00	0.19	0.02
	(0.06)	(0.00)	(0.01)	(0.00)
Chrom. Abnormalities, a < c	0.33	0.33	0.33	0.33
	(0.02)	(0.02)	(0.02)	(0.02)
No live birth	1.10	2.82	2.73	2.81
	(0.06)	(0.06)	(0.06)	(0.06)
No chrom. Abnormalities	0.55	0.00	0.11	0.01
	(0.03)	(0.00)	(0.00)	(0.00)
Chrom. Abnormalities (& a>c)	0.55	2.82	2.62	2.80
	(0.03)	(0.06)	(0.06)	(0.06)
Cost and surplus ($US per pregnancy):				
Total government cost	0	--	298	548
	--	--	(5)	(5)
cfDNA cost	0	--	0	486
	--	--	--	(4)
Invasive testing cost	0	--	298	62
	--	--	(5)	(1)
Consumer surplus	normalized to 0	2,670	2,375	2,642
	--	(262)	(108)	(117)

Notes: Table shows counterfactual testing decisions and pregnancy outcomes (per 100 pregnancies), and average government spending and consumer surplus (per pregnancy) under alternative assumptions about available technology and information for pregnancies in the baseline sample (i.e. NT risk score $\geq \frac{1}{1,000}$). Consumer surplus and government spending are in $US. Consumer surplus is normalized to zero for the counterfactual with no testing (column 1). Government spending includes invasive testing ($1,248.50 per test) and cfDNA screening ($567.50 per screen). Outcomes that are (ex-post) inefficient are shaded in gray. Standard errors (in parentheses) are computed using 1,000 bootstrap samples.

Appendix Table A5: Outcomes under counterfactual coverage of cfDNA with standard errors

	No cfDNA	Full cfDNA covg.	No cfDNA covg.	cfDNA covg. for NT ≥ 1/200	cfDNA covg. for NT in [1/200,1/51]
	(1)	(2)	(3)	(4)	(5)
Testing (per 100 pregnancies):					
Any testing	23.83	88.23	28.29	33.10	32.72
	(0.42)	(0.77)	(0.60)	(0.49)	(0.50)
cfDNA only	0	83.29	14.30	28.01	23.16
	--	(0.74)	(0.96)	(0.55)	(0.62)
Invasive only	23.83	2.52	13.57	3.35	8.97
	(0.42)	(0.08)	(0.74)	(0.24)	(0.31)
Both	0	2.41	0.42	1.74	0.59
	--	(0.04)	(0.03)	(0.03)	(0.03)
Pregnancy outcomes (per 100 pregnancies):					
Live birth	97.27	97.19	97.30	97.30	97.29
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
No chrom. Abnormalities	96.74	96.84	96.79	96.84	96.81
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Chrom. Abnormalities, a > c	0.19	0.02	0.17	0.13	0.14
	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
Chrom. Abnormalities, a < c	0.33	0.33	0.33	0.33	0.33
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No live birth	2.73	2.81	2.70	2.70	2.71
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
No chrom. Abnormalities	0.11	0.01	0.06	0.01	0.03
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Chrom. Abnormalities (& a>c)	2.62	2.80	2.64	2.69	2.68
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Cost and surplus ($US per pregnancy):					
Total government cost	298	548	175	209	212
	(5)	(5)	(9)	(3)	(4)
cfDNA cost	0	486	0	145	93
	--	(4)	--	(2)	(1)
Invasive testing cost	298	62	175	64	119
	(5)	(1)	(9)	(3)	(4)
Consumer surplus	2,375	2,642	2,423	2,529	2,491
	(109)	(117)	(112)	(113)	(113)

Notes: Table shows counterfactual testing decisions and pregnancy outcomes (per 100 pregnancies), and average government spending and consumer surplus (per pregnancy) under alternative assumptions about cfDNA coverage for pregnancies in the baseline sample (i.e. NT risk score $\geq \frac{1}{1,000}$). Throughout, invasive testing is assumed to be available for free. Consumer surplus and government spending are in $US. 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 screening ($567.50 per screen). Outcomes that are (ex-post) inefficient are shaded in gray. Standard errors (in parentheses) are computed using 1,000 bootstrap samples.

Appendix Table A6: Outcomes under counterfactual information and technology, Full sample

	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	--	3.09	50.77
cfDNA only	0	--	0	49.73
Invasive only	0	--	3.09	0.32
Both	0	--	0	0.72
Pregnancy outcomes (per 100 pregnancies):				
Live birth	99.86	99.62	99.65	99.62
No chrom. Abnormalities	99.51	99.58	99.56	9.57
Chrom. Abnormalities, a > c	0.31	0.00	0.04	0.01
Chrom. Abnormalities, a < c	0.04	0.04	0.04	0.04
No live birth	0.14	0.38	0.35	0.38
No chrom. Abnormalities	0.07	0.00	0.01	0.00
Chrom. Abnormalities (& a>c)	0.07	0.38	0.34	0.37
Cost and surplus ($US per pregnancy):				
Total government cost	0	--	39	299
cfDNA cost	0	--	0	286
Invasive testing cost	0	--	39	13
Consumer surplus	normalized to 0	358	306	350

Note: Table reproduces Table 3 of the main text, but using the full sample (instead of the baseline sample), which also includes many low-risk pregnancies (see Table 1).

Appendix Table A7: Outcomes under counterfactual information and technology, Full sample

	No cfDNA (1)	Full cfDNA covg.* (2)	No cfDNA covg. (3)	cfDNA covg. For NT ≥ 1/200 (4)
Testing (per 100 pregnancies):				
Any testing	3.09	50.77	3.68	4.31
cfDNA only	0	49.73	1.87	3.65
Invasive only	3.09	0.32	1.76	0.43
Both	0	0.72	0.06	0.23
Pregnancy outcomes (per 100 pregnancies):				
Live birth	99.65	99.62	99.65	99.65
No chrom. Abnormalities	99.56	9.57	99.57	99.58
Chrom. Abnormalities, a > c	0.04	0.01	0.04	0.03
Chrom. Abnormalities, a < c	0.04	0.04	0.04	0.04
No live birth	0.35	0.38	0.35	0.35
No chrom. Abnormalities	0.01	0.00	0.01	0.00
Chrom. Abnormalities (& a>c)	0.34	0.37	0.34	0.35
Cost and surplus ($US per pregnancy):				
Total government cost	39	299	23	27
NT saving*	0	-174	0	0
cfDNA cost	0	286	0	19
Invasive testing cost	39	13	23	8
Consumer surplus	306	350	312	326

Note: Table reproduces Table 4 of the main text, but using the full sample (instead of the baseline sample), which also includes many low-risk pregnancies (see Table 1).
*In the case of full cfDNA coverage for the full sample, an NT screening is redundant, so one of the benefits of full coverage that we account for is the saving of the cost ($174 per pregnancy) of the NT screening.

Appendix Table A8: Parameter estimates for model robustness exercises

	Baseline	Bin width 25	Bin width 100	Age shifters	"Reversed" ψ
β_a	11.799	11.676	11.803	11.828	11.623
	(0.042)	(0.079)	(0.035)	(0.035)	(0.083)
γ_a				0.056	
				(0.091)	
β_c	12.220	12.070	12.224	12.261	12.180
	(0.042)	(0.075)	(0.035)	(0.039)	(0.091)
γ_c				0.067	
				(0.095)	
σ_a	0.522	0.635	0.445	0.445	0.616
	(0.058)	(0.046)	(0.051)	(0.054)	(0.083)
σ_c	0.606	0.709	0.555	0.574	0.849
	(0.021)	(0.029)	(0.024)	(0.026)	(0.046)
ρ	0.831	0.888	0.819	0.821	0.873
	(0.012)	(0.005)	(0.014)	(0.014)	(0.018)
ψ	0.926	0.910	0.937	0.918	1.084
	(0.009)	(0.010)	(0.007)	(0.012)	(0.014)

Note: Table shows the parameter estimates associated with the alternative model specifications described in Appendix C. Standard errors (in parentheses) are computed using 100 bootstrap samples.

Appendix Table A9: Counterfactual outcomes for model robustness exercises

	No cfDNA		Full cfDNA covg.		cfDNA covg. for NT ≥ 1/200	
	CS	Gov. Cost	CS	Gov. Cost	CS	Gov. Cost
Baseline	2,375	298	2,642	548	2,529	209
Alternative bin width for estimation moments (instead of 50 in the baseline):						
Bin width of 25 (instead of 50)	2,132	290	2,376	543	2,277	206
Bin width of 100 (instead of 50)	2,349	293	2,612	494	2,499	208
Alternative bivariate distribution of (a,c) (instead of lognormal in the baseline):						
Bivariate normal	1,095	278	1,264	542	1,209	204
a exponential, c normal	1,955	298	2,201	502	2,011	200
a chi-squared, c normal	819	278	984	551	943	200
Allowing age to shift the mean of a and c	2,697	320	2,997	568	2,868	213
"Reverse" consultation impact at q_i=1/200	3,138	295	3,459	551	3,309	207
True miscarriage rate of 1/800 due to invasive testing	2,495	298	2,654	548	2,540	209

Note: Table shows average government cost and consumer surplus (CS) per pregnancy under alternative assumptions about cfDNA coverage for pregnancies in the baseline sample, for each of the alternative model specifications described in Appendix C.