# On the Value of Birth Weight*

DAMIAN CLARKE,†,‡  SONIA OREFFICE‡,§,¶  and
CLIMENT QUINTANA-DOMEQUE‡,§,¶,††  iD

†Department of Economics, Universidad de Chile, Santiago, Chile (e-mail:
dclarke@fen.uchile.cl)
‡IZA, Bonn, Germany
§University of Exeter, Exeter, UK (e-mail: s.oreffice@exeter.ac.uk;
c.quintana-domeque@exeter.ac.uk)
¶HCEO, Chicago, IL, USA
††GLO, Essen, Germany

## Abstract

A large body of evidence documents the educational and labour market returns to birth weight, which are reflected in investments in large social safety net programmes targeting birth weight and early life health. However, there is no direct evidence on the *private* valuation of birth weight. In this paper, we estimate the willingness to pay for birth weight in the United States, using a series of discrete choice experiments. Within the normal birth weight range (2,500–4,000 g), we find that individuals are, on average, willing to pay $1.47 (95% CI: [$1.24, $1.70]) for each additional gram of birth weight when the value of birth weight is estimated linearly, or $2.40 (95% CI: [$2.03, $2.77]) when the value of birth weight is estimated non-parametrically.

## I.   Introduction

The weight of a newborn is a well-known measure of the initial endowment or stock of human capital early in life (Almond and Currie, 2011; Almond, Currie and Duque, 2018). The importance of the fetal period as a predictor of health throughout the life course has been recognized in a series of influential papers by Barker and coauthors on the fetal origins of disease (Barker *et al*., 1989; Barker, 1990, 1995), with considerable and ever-growing evidence that insults to fetal health have enduring and significant

costs throughout life (Case, Fertig and Paxson, 2005; Almond, 2006; Currie and Moretti, 2007; Black, Devereux and Salvanes, 2007; Almond, Edlund and Palme, 2009). These findings justify sizeable welfare programmes targeted at babies with poor endowments early in life, such as those focusing on low birth weight infants (Almond, Chay and Lee, 2005; Bharadwaj, Løken and Neilson, 2013) and pre-natal nutrition programmes, such as the Special Supplemental Nutrition Program for Women Infants and Children (WIC).

Despite a large body of evidence on the importance of birth weight and considerable public investment, little is known regarding the private valuation of this birth outcome, or other newborn measures. Knowing the value which people place on birth weight and other birth characteristics is of public concern and a fundamental policy issue, in particular as a key ingredient to policies focused on parental behaviour *prior* to and *during* gestation. To the degree that a wide range of (costly) parental behaviours can positively impact birth weight (Rosenzweig and Schultz, 1983; Sexton and Hebel, 1984; Chevalier and O'Sullivan, 2007), the perceived importance of birth weight to parents may have significant effects on these behaviours.

In this paper, we estimate the importance of birth weight to individuals, as measured by their willingness to pay (WTP) for birth weight. In order to do so, we conducted a series of discrete choice experiments on Amazon's Mechanical Turk (MTurk), an online labour market platform. This is increasingly used in social science research (Kuziemko *et al.*, 2015; Jordan *et al.*, 2016) and, in particular, a recent study has relied on this platform to estimate the value of life before and after birth (Jamison, 2016). We conducted these experiments with approximately 2,000 respondents, half of them interviewed in 2016, and half of them in 2018. Respondents were asked to consider seven pairs of birth scenarios sequentially, amounting to around 28,000 different birth scenarios with a number of different characteristics. These characteristics were each orthogonally varied both within and between experimental subjects. Specifically, we performed conjoint analysis (CA), a method first described by Lancaster (1966).

These experiments allow respondents to reveal their preferences (or lack thereof) over a range of birth characteristics. In particular, we randomise a baby's birth weight, monetary costs of birth, gender and birth timing. Birth weight was randomised within the normal range of 2,500–4,000 g. We restrict our analysis to the 'normal' range for two reasons. First, not only do continuous measures of birth weight have greater explanatory power for a larger range of variables than a low birth weight (LBW, or weights less than 2,500 g) indicator (Black *et al.*, 2007), but recent evidence also suggests that marginal increases in birth weight within the normal weight range are particularly important for well-being. Royer (2009) suggests that given this fact, babies born in the normal range of weights should receive *more* research attention.[1] Indeed, Maruyama and Heinesen (2020) show that health effects of LBW exist also in the

---

[1]In full, Royer (2009) reports (p. 52): 'I find that the effects of birth weight on long-run outcomes are nonlinear and for educational attainment, in particular, are largest above 2,500 g, the cutoff for defining low birth weight. These findings suggest that babies with birth weights outside the lower tail of the distribution (i.e. outside the range of low birth weight) should receive more attention'.

normal range. Second, from a purely practical standpoint, we focus on the normal range of birth weights to avoid priming effects (i.e. respondents linking low birth weight with other health conditions or complications at birth), thus confounding our estimates for the WTP of birth weight alone.

We find that a baby weighing 3,400 g (7 lbs, 8 oz) is 18 percentage points (pp) more likely to be preferred than one weighing 2,500 g (5 lbs, 8 oz). We estimate that over the normal range of birth weights, experimental participants would be willing to pay on average $1.47 (95% CI: [$1.24, $1.70]) for each additional gram of weight when the value of birth weight is estimated linearly, or $2.40 (95% CI: [$2.03, $2.77]) when the value of birth weight is estimated non-parametrically.

Our experimental estimates are consistent with studies showing that individuals make fertility-related decisions based on monetary costs and non-pecuniary birth characteristics. Dickert-Conlin and Chandra (1999) and LaLumia, Sallee and Turner (2015) report that in the US parents may move expected January births backwards to December to gain tax benefits, and Schulkind and Shapiro (2014) clearly show that parents are willing to anticipate the births of their children to gain tax benefits despite impacts of this on their child's birth outcomes. They find that a $1,000 increase in tax benefits translates into an approximately 0.37 pp increase in the probability of a December birth. Moreover, they show that shifts in birth timing owing to the additional $1,000 in tax benefits causes between a 0.07% and 0.19% decrease in average birth weight, or between 2.41 and 6.37 g. Combining these two findings, we obtain an estimated WTP for birth weight of $1.17–$3.09 per gram, a range of non-experimental magnitudes which matches with our experimental estimates.

In what follows, we describe the MTurk data and experimental set-up in section II and the methodology for estimating WTP in section III. In section IV, we present our experimental estimates of the WTP for birth weight. In section V, we assess the validity of our experimental findings, and place them in broader context. In section VI, we check the robustness of our findings to preference heterogeneity. Section VII concludes.

## II.   Data description

We collected data on preferences over birth characteristics by running discrete choice experiments on Amazon's MTurk online platform. This platform is a market place which provides access to a pool of US MTurk workers who are paid per completed Human Intelligence Task (HIT). We posted a HIT request to recruit respondents to complete a series of discrete choice experiments (described further below) as well as a series of demographic questions. These demographic questions were asked *after* the completion of the experiments to avoid any framing or experimenter demand effects (Zizzo, 2010), and the survey was advertised as a general demographic survey. MTurk respondents have been documented to have desirable characteristics, and be more similar to the US population than other frequently used subject pools such as college student samples (Berinsky, Huber and Lenz, 2012). While MTurk samples are increasingly used in social science research, the pool of subjects consists of individuals who sign up to participate in MTurk tasks, and so is self-selected. Nonehtheless,

MTurk is a valuable research tool to collect data and use CA in the context of research in health (Mortensen and Hughes, 2018). In their review, Mortensen and Hughes (2018) highlighted several strengths of MTurk, including its reliability and the high quality of the information provided by the participants, with comparability to responses in high-quality data samples. In addition, MTurk allows researchers to collect large national pools of data, including quantitative and qualitative data about patients' knowledge and experience with events such as miscarriage (Bardos *et al.*, 2016), while also documenting comparability among responses in MTurk samples, and university-wide communities (Wu *et al.*, 2017).

We published two HITs containing an identical experiment at different time periods. The first of these was published on a Monday in September 2016, and the second of these was published on a Monday in May 2018,[2] requesting the completion of a short survey. In the second HIT, we prevented any previous respondents from completing the survey to avoid priming effects. Workers were paid $1.10 for a 6-minute experimental survey (average length) in the first wave, resulting in an effective hourly pay rate of approximately $11. This payment was increased to $1.20 to correct for inflation in the second wave. The survey needed to be completed in order to be able to receive payment, and it was impossible to move forward if the question on the screen was not answered. We required that respondents must be from the United States,[3] and in order to maximise the likelihood that workers were based in the United States at the time of completing the survey, this was launched at 9:00 AM East Coast Time in both cases. In both cases, by approximately 2:00 PM of the same day 1,002 and 1,003 valid responses were collected. We also required that workers had completed at least 100 tasks on MTurk in the past, and had achieved an approval rating of greater than 95% on these tasks. These restrictions are common in MTurk research (Berinsky *et al.*, 2012; Francis-Tan and Mialon, 2015). Of the 2,005 valid responses completed, we removed a small number based on a set of pre-defined consistency checks. These were: (a) workers whose geographical IP address placed them outside of the United States at the time of survey (72 respondents, or 3.6%), any respondents who failed a consistency check where a question was repeated at the beginning and end of the demographic portion of the survey (26 respondents), and any respondents completing the entire exercise in under 2 minutes (16 respondents). The final sample consists of 1,894 respondents.[4]

Summary statistics of the respondents are provided in Table 1. Slightly more than half of all respondents are female (53%), and the average age of our respondents is 36.5 (with a standard deviation of 11.7 years); 82% are white and 8% are Hispanic. Approximately half of the respondents are parents (50%), and of those who are non-parents, 47% intend to have children or are already pregnant (implying that 27% of

---

[2]We were interested in examining robustness over time and in different seasons of the year. Results from only the first HIT are presented in a previous version of this working paper (Clarke, Oreffice and Quintana-Domeque, 2017), and are qualitatively identical to those we document below from the larger two-wave experiment.

[3]Workers on Mechanical Turk are required to have a US Social Security Number.

[4]The categories of removed cases are not mutually exclusive.

TABLE 1

*Summary statistics of respondents*

|  | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Female | 1,894 | 0.53 | 0.50 | 0.00 | 1.00 |
| Age | 1,894 | 36.55 | 11.68 | 18.00 | 82.00 |
| Black | 1,894 | 0.08 | 0.27 | 0.00 | 1.00 |
| White | 1,894 | 0.82 | 0.38 | 0.00 | 1.00 |
| Hispanic | 1,894 | 0.08 | 0.27 | 0.00 | 1.00 |
| Parent | 1,894 | 0.50 | 0.50 | 0.00 | 1.00 |
| Non-parent intending children | 955 | 0.47 | 0.50 | 0.00 | 1.00 |
| Number of children | 1,894 | 1.01 | 1.27 | 0.00 | 6.00 |
| Married | 1,894 | 0.45 | 0.50 | 0.00 | 1.00 |
| Employed | 1,894 | 0.73 | 0.44 | 0.00 | 1.00 |
| Some college + | 1,894 | 0.89 | 0.31 | 0.00 | 1.00 |
| Years of education | 1,894 | 14.67 | 1.73 | 8.00 | 17.00 |
| Total family income (1000s of USD) | 1,894 | 59.63 | 39.29 | 5.00 | 175.00 |
| Hourly earnings on MTurk | 1,894 | 4.43 | 2.85 | 1.50 | 11.50 |

*Notes*: Refer to Online Appendix Figure A1 for a discussion of the experimental sample. Years of education, total income and hourly MTurk earnings are calculated from categorical variables. Non-Parent Intending Children refers to any respondent who either answers that they are pregnant or intend to have children, and currently have no children.

respondents are neither parents, nor intend to become so). In total, 45% of the respondents are married, 73% are employed and 89% have at least some college.

The geographical location of these respondents within the United States (based on their IP address) is provided in the online appendix Figure A1. The geographical coverage is broadly representative of the US population. In the online appendix Table A1 we compare our MTurk respondent coverage with the US population from 2015 (United States Census Bureau, 2015). In general, we see that our MTurk sample lines up well with the national population at the state level, however there are a number of exceptions, such as the lower number of respondents from California, most likely reflecting the earlier time zone on the West Coast.

In the online appendix Table A2 and Figure A2, we compare the observable (average) characteristics of our MTurk sample to those of the US population based on the 2015 American Community Survey (ACS). In each case in Table A2, we present descriptive statistics from each sample, as well as a formal test of equality of means. We observe that the MTurk sample is on average younger and consists of a higher proportion of white and parent respondents. MTurk respondents are more likely to be women, more educated but with lower income. In Figure A2, we compare the distribution of family income in the MTurk versus the ACS: our MTurk sample has fewer individuals in the bins above $80,000 and more individuals in virtually all the bins below $80,000.

## III.  Methodology

In order to estimate the perceived importance of birth weight in terms of willingness to pay, we run discrete choice experiments (DCEs). A DCE is a type of CA: An

experiment in which respondents are asked to choose their preferred option from a set when a number of attributes are varied simultaneously. CA was borne from early work in consumer theory in which tastes for goods owe to the collection of their characteristics (Lancaster, 1966). In the past, CA has been used to measure preferences over medical care in a variety of contexts, including the valuation of waiting times (Propper, 1990, 1995), alternative miscarriage treatment options (Ryan and Hughes, 1997), asthma medications (King *et al.*, 2007), or depression management (Wittink *et al.*, 2010). In other settings, theoretical choices in CA have been documented to agree with actual choice behaviour in the real world, and outperform vignette experiments (Hainmueller, Hangartner and Yamamoto, 2015).

Our birth choice experiments consist of asking respondents to consider a series of paired birth scenarios, while focusing on four attributes of each birth scenario. We use a main-effects, orthogonal (all attribute levels vary independently) and balanced (each level of an attribute occurs the same number of times) experimental design. In the experiment, the attributes are combined to form various (hypothetical) birth scenarios, all about a hospital birth of the first child with no complications. The attributes considered are the baby's weight at birth (5 lbs, 8 oz; 5 lbs, 13 oz; 6 lbs, 3 oz; 6 lbs, 8 oz; 6 lbs, 13 oz; 7 lbs, 3 oz; 7 lbs, 8 oz; 7 lbs, 13 oz; 8 lbs, 3 oz; 8 lbs, 8 oz; or 8 lbs, 13 oz), the out of pocket expenses associated with the birth ($250; $750; $1,000; $2,000; $3,000; $4,000; $5,000; $6,000; $7,500; or $10,000), the sex of the child (boy or girl), and the season in which the baby is born (winter, spring, summer or fall). The latter options are used in order to avoid priming respondents into thinking that we are interested in birth weight. As these attributes are all orthogonally varied, the effect of each characteristic on the likelihood that a particular birth is chosen is separately identified (Marshall *et al.*, 2010).

Each respondent was asked to consider seven pairs of birth scenarios in an iterative fashion. In order to move forward in the experiment, a choice must be made for each pair, and once the choice has been made the respondent may not go back and revise their choice. In each case, the two pairs were displayed side-by-side on a single screen, and respondents were asked to indicate which was their preferred birth scenario. As well as randomizing the level of each attribute on each profile, the order of the attributes was randomised; however, to reduce the cognitive load to respondents the ordering of attributes was only randomised once, and then fixed across the seven pairings that the respondent ranked. The DCE's framing and the explanation of the attributes shown to respondents are displayed in the online appendix Figures A3 and A4. In Figure A5, we display an example of a pair of birth scenarios as presented to respondents.

The levels of attributes were chosen to represent plausible values from the US population (Ryan and Farrar, 2000), and extreme values were avoided to prevent the likelihood of 'grounding effects' (or corner solutions), following Bridges *et al.* (2011). In order to minimise the likelihood that respondents would employ simple heuristics in answers, we limited the number of attributes (four) which need be considered. As discussed in Bridges *et al.* (2011), we observe in experimental responses that such heuristics are not employed given response sensitivity to all dimensions studied. Birth weights were always presented in pounds and ounces, as this experiment was run with

a US sample. As well as indicating that all births were *complication-free*, only birth weights *within* the normal range of 2,500–4,000 g were included (11 evenly spaced weights were defined in this range). This range includes the vast majority of all births in the United States.[5] From our reading of the US literature on out-of-pocket medical expenses, the US insurance system and hospital bills in the United States for the delivery of a baby, the value range from $250 to $10,000 seemed a plausible range for 2016 and 2018. Recent evidence from the United States (Moniz *et al.*, 2020) suggests that for individuals with employer based insurance (around 50% of deliveries in the United States), the average out of pocket costs for a birth were $4,569 in 2015. This is quite close to the mid-point in the values provided in the experiment. Among women not covered by employer based insurance, the large majority of births (90%) are covered by Medicaid, and so likely have lower copayments. In the DCEs, it is stated that the birth scenarios refer to a 'hospital birth of first child with no complications' so that the out-of-pocket expenses are associated to a healthy hospital delivery and represent the monetary valuation of the attribute 'birth weight' in our DCEs. An opt-out option was not included in any of the discrete choices. This has been suggested to have desired properties such as avoiding non-random opt-out of all questions (Bekker-Grob, Ryan and Gerard, 2012; Veldwijk *et al.*, 2014).

We are interested in estimating two quantities. First, we would like to estimate, *ceteris paribus*, the likelihood that a birth scenario is chosen given that a particular birth weight is observed (compared with an omitted base category). Second, we would like to estimate the WTP for birth weight, by combining the information from both variations in birth weight and variations in out-of-pocket costs.

Consider a sample of $i \in \{1,\ldots,N\}$ individuals, each of whom considers $K$ choice tasks in which they must decide between $J$ options (profiles, or in our case, birth scenarios). Each profile contains $L$ attributes, where each particular attribute $l$ consists of discrete levels of the variable. In the case of the DCE described above, we have $N = 1,894$ respondents, $K = 7$ choice tasks per respondent, $J = 2$ profiles per task, and $L = 4$ attributes.[6] We follow Hainmueller, Hopkins and Yamamoto (2013) in defining a treatment vector $T_{ijk}$. This treatment vector has $L$ cells, and summarises for individual $i$, at choice task $k$, for profile $j$, the full set of attributes observed. Each particular attribute $T_{ijkl}$ is randomly assigned from among all the levels of $l$, the assignment of which is orthogonal to all other attributes the respondent sees. Using the potential outcomes framework, we define a binary variable $Y_{ijk}(\bar{\mathbf{t}})$ which takes the value 1 if respondent $i$ would choose profile $j$ on choice set $k$ if faced with the set of attributes $\bar{\mathbf{t}}$, or 0 if the profile would not be chosen.

---

[5]According to full vital statistics of 2013 from the National Vital Statistics System (see Figure A6 in the online appendix), 8.02% of births were LBW (<2,500 g), and 7.89% were large for gestational age at birth (>4,000 g).
[6]These four attributes have 2, 10, 11 and 4 levels, respectively, for sex, out of pocket costs, birth weight and season of birth.

Hainmueller *et al.* (2013) call this first quantity the average marginal component effect (AMCE) and demonstrate that under reasonably weak assumptions,[7] it can be recovered using a non-parametric sub-classification estimator, conditional regression or a simple difference in means. The logic of the AMCE is to capture the change in the likelihood that a given profile would be chosen if the *l*th component were changed from $t_0$ to $t_1$, or in our case, a change in birth weight.[8]

Under the controlled randomisation in CA, Holland (1986)'s fundamental problem of causal inference is resolved by construction, as on average there will be no correlation between observing the particular level of an attribute and individual characteristics. Treatment units are thus those who observe a particular $t_1$, while those who do not act as controls. In practice, to estimate the change in the likelihood that a birth scenario is chosen given a change in birth weight (or any other attribute), we estimate the following two equations:

$$Pr(Y_{ijk}=1) = \Lambda\left(\alpha + \beta\,Costs_{ijk} + \gamma\,BW_{ijk} + \sum_{r=2}^{4} \delta_r\,SOB_{ijk,r} + \kappa\,Girl_{ijk} + \mu_j + \phi_k\right) \quad (1)$$

and

$$Pr(Y_{ijk}=1) = \Lambda\left(\alpha + \beta\,Costs_{ijk} + \sum_{q=2}^{11} \gamma_q\,BW_{ijk,q} + \sum_{r=2}^{4} \delta_r\,SOB_{ijk,r} + \kappa\,Girl_{ijk} + \mu_j + \phi_k\right),$$
$$(2)$$

where $Y_{ijk}=1$ if the birth scenario $j$ is chosen, $\Lambda$ is the *cdf* of the logistic distribution, $Costs_{ijk}$ denotes the out of pocket expenses associated with the birth scenario $j$, $BW_{ijk}$ is the birth weight associated with the birth scenario $j$, $BW_{ijk,q}$ is equal to 1 if the birth weight category of the birth scenario $j$ is $q$, $SOB_{ijk,r}$ is equal to 1 if the season of birth category of the birth scenario $j$ is $r$, $Girl_{ijk}$ is 1 if the gender of the baby of the birth scenario $j$ is girl, and $\mu_j$ and $\phi_k$ are option-profile and choice-task order fixed effects respectively. Standard errors are clustered at the level of the respondent to capture the (likely) positive correlations among choices based on attributes by a particular respondent.

We estimate equations (1) and (2) and report average marginal effects. We omit from equation (2) the lowest birth weight category as the baseline level, implying that all marginal effects of each birth weight should be interpreted as the marginal likelihood of choosing a birth scenario given birth weight $q$ in place of the lowest birth weight (2,500 g).

---

[7]These assumptions relate to randomisation of attributes, and stability of respondent behaviour regardless of the number of profiles that they have seen or the order of the attribute in the profile. This first assumption holds by construction in our experiment. A benefit of the set-up of the DCE is that even if order and round effects are not completely neutral, these can be flexibly captured using fixed effects in a regression.

[8]Formally, the AMCE is defined as (Hainmueller *et al.*, 2013):

$$E[Y_i(t_1, T_{ijk[-l]}, \mathbf{T}_{i[-j]k}) - Y_i(t_0, T_{ijk[-l]}, \mathbf{T}_{i[-j]k}) | (T_{ijk[-l]}, \mathbf{T}_{i[-j]k}) \in \tilde{\mathcal{T}}]$$

which can be quite easily calculated by integrating over all of the other attributes and levels except for $t_1$ (the treatment of interest) and $t_0$ (the baseline level for the attribute). These other attributes and levels are denoted as the set $\tilde{\mathcal{T}}$ here.

We then estimate the average WTP for birth weight in two different ways: using equations (1) and (2) respectively. From equation (1), the marginal effects on the likelihood of choosing a particular birth scenario given an increase in the particular attribute, conditional on all other attributes, are:

$$\frac{\partial Pr(Y_{ijk}=1)}{\partial Costs_{ijk}} = \beta \Lambda'(\cdot) \qquad \frac{\partial Pr(Y_{ijk}=1)}{\partial BW_{ijk}} = \gamma \Lambda'(\cdot),$$

where $\Lambda'$ is the *pdf* of the logistic distribution. Given these marginal effects, the marginal rate of substitution (MRS) between birth weight *BW* and the price of a given birth (the out of pocket costs) – which measures the change in costs that a respondent would be willing to withstand for a marginal increase in birth weight – is given by:

$$MRS_{BW,Costs} = \frac{\frac{\partial Pr(Y_{ijk}=1)}{\partial BW_{ijk}}}{\frac{\partial Pr(Y_{ijk}=1)}{\partial Costs_{ijk}}} = \frac{\gamma}{\beta}.$$

Multiplying this quantity by minus 1 gives precisely the WTP:

$$WTP_{BW}(1) = -\frac{\gamma}{\beta} = -\frac{\partial Cost_{ijk}}{\partial BW_{ijk}}.$$

Note that in the above calculation we take the negative so that costs are interpreted as the positive amount that must be paid rather than the negative change in financial resources. This *WTP* can also be derived quite straightforwardly from a model of the indirect utility function as described in Zweifel, Breyer and Kifmann (2009). In order to calculate the confidence interval associated with the WTP, we use the *delta method*, which is both simple and shown to perform well under simulation (Hole, 2007a).

Finally, we also compute the average WTP based on equation (2) as:

$$WTP_{BW}(2) = -\frac{1}{\beta} \sum_{q=2}^{11} \omega_q \gamma_q,$$

where $\omega_q$ is the fraction of births with weight between $q-1$ and $q$ in the birth data from National Vital Statistics System (NVSS) over the normal birth weight range so that $\sum_{q=2}^{11} \omega_q = 1.$ [9] A 95% confidence interval for this second WTP measure is also constructed using the delta method.

## IV. Experimental results

### Results for the whole sample

Figure 1 shows that the randomisation worked as intended: it balanced observable characteristics across the range of experimental attributes. In examining 12 observable characteristics of respondents, an Omnibus *F*-test suggests no lack of balance at any

---

[9] $q=1$ corresponds to 2,500 g, $q=2$ to 2,637 g, $q=3$ to 2,807 g, $q=4$ to 2,948 g, $q=5$ to 3,090 g, $q=6$ to 3,260 g, $q=7$ to 3,402 g, $q=8$ to 3,544 g, $q=9$ to 3,714 g, $q=10$ to 3,856 g, and $q=11$ to 4,000 g.
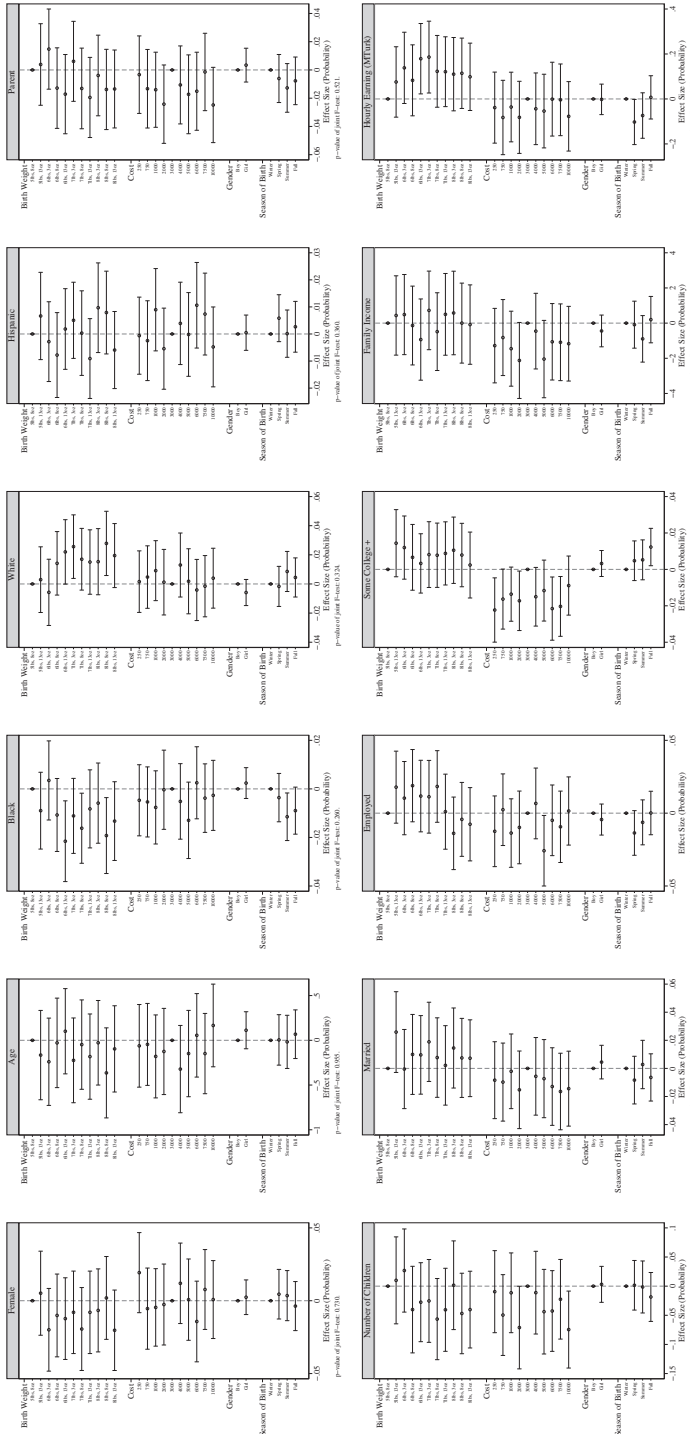
Figure 1. Balance tests

*Notes*: Point estimates and confidence intervals are plotted, documenting whether respondents with a particular observable characteristic were more likely to observe particular attributes in conjoint experiments. Variables described in Table 1 are regressed on whether the respondent observed each possible attribute in the experiment. The observable variable (dependent variable) in each case is indicated in plot titles. In each plot, 95% confidence intervals are reported for each attribute in the experiment, and the omitted category for each characteristic is indicated as a point at 0. Each plot represents a separate regression, and the *P*-value of an Omnibus F-test for each regression is reported at the base of each plot.
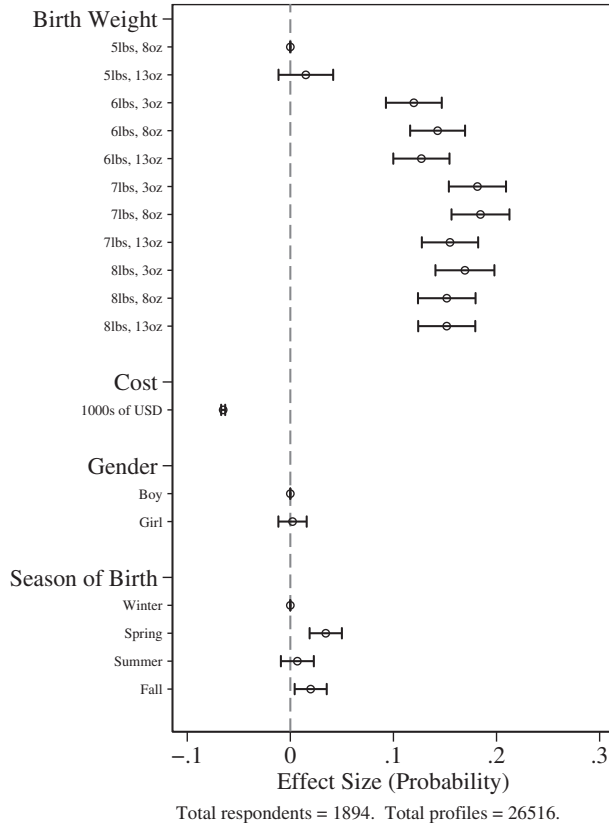
Figure 2. Discrete choice experimental results

*Notes*: Point estimates and confidence intervals are displayed of the change in likelihood of choosing a birth profile given that a particular characteristic was seen. Each characteristic is compared to the omitted base case indicated on the zero line. Each respondent observes 7 paired birth scenarios, resulting in 14 profiles per respondent. 95% confidence intervals are clustered by respondent, and costs are displayed as a linear coefficient. Fully non-parametric costs are displayed in Appendix Figure A7.

conventional significance level in each case. Our main experimental results are presented in Figure 2. This figure displays point estimates of the likelihood of preferring a particular birth scenario given each characteristic, compared with an omitted base category for each characteristic. Along with each point estimate, the 95% confidence interval is plotted, clustering by respondent. While we present cost as a linear variable measured in 1,000s of dollars, in the online appendix Figure A7 the same results are presented with costs displayed as the same categorical measure observed by respondents.[10]

The top panel displays the likelihood of choosing a birth scenario given a particular birth weight, compared to being shown the minimum sample birth weight of 5 lbs, 8 oz (2,500 g). In each case, higher birth weights are associated with a greater likelihood of choosing the corresponding birth scenario. The most preferred birth weight (based on point estimates) is 7 lbs, 8 oz (3,400 g), which results in a birth

---

[10]In the online appendix Figure A8, we document that results are largely unchanged if we work with the full sample of 2,005 respondents rather than the preferred sample of 1,894 respondents meeting inclusion criteria.

TABLE 2

*Birth characteristics and willingness to pay for birth weight*

| | Continuous (1) | Categorical (2) |
|---|---|---|
| Birth weight (in 1000s of g) | 0.092*** | |
| | [0.007] | |
| Cost numerical | −0.063*** | −0.063*** |
| | [0.001] | [0.001] |
| 5 lbs, 13 oz | | 0.016 |
| | | [0.014] |
| 6 lbs, 3 oz | | 0.120*** |
| | | [0.014] |
| 6 lbs, 8 oz | | 0.143*** |
| | | [0.014] |
| 6 lbs, 13 oz | | 0.128*** |
| | | [0.014] |
| 7 lbs, 3 oz | | 0.181*** |
| | | [0.014] |
| 7 lbs, 8 oz | | 0.184*** |
| | | [0.014] |
| 7 lbs, 13 oz | | 0.154*** |
| | | [0.014] |
| 8 lbs, 3 oz | | 0.169*** |
| | | [0.015] |
| 8 lbs, 8 oz | | 0.152*** |
| | | [0.014] |
| 8 lbs, 13 oz | | 0.151*** |
| | | [0.014] |
| Girl | 0.003 | 0.002 |
| | [0.007] | [0.007] |
| Spring | 0.034*** | 0.034*** |
| | [0.008] | [0.008] |
| Summer | 0.006 | 0.006 |
| | [0.008] | [0.008] |
| Fall | 0.019** | 0.019** |
| | [0.008] | [0.008] |
| WTP for birth weight (1000 g) | 1,470.3 | 2,401.0 |
| 95% CI | [1,238.3, 1,702.3] | [2,032.4, 2,769.6] |
| Observations | 26,516 | 26,516 |

*Notes*: Average marginal effects from a logit regression are displayed. All columns include option order fixed effects and round fixed effects. Standard errors are clustered by respondent. Willingness to pay and its 95% confidence interval is estimated based on the ratio of costs to the probability of choosing a particular birth weight. The 95% confidence interval is calculated using the delta method for the ratio.
[*]$P$-value < 0.1, **$P$-value < 0.05, ***$P$-value < 0.01.

scenario being approximately 18 pp more likely to be chosen than the omitted base category. The magnitudes of the estimates are large. With the exception of 5 lbs, 13 oz, all higher birth weights are at least 12 pp more likely to be chosen.

As discussed in section III, we can combine estimates of average marginal component effects to generate estimates of the WTP for each characteristic. In Table 2,
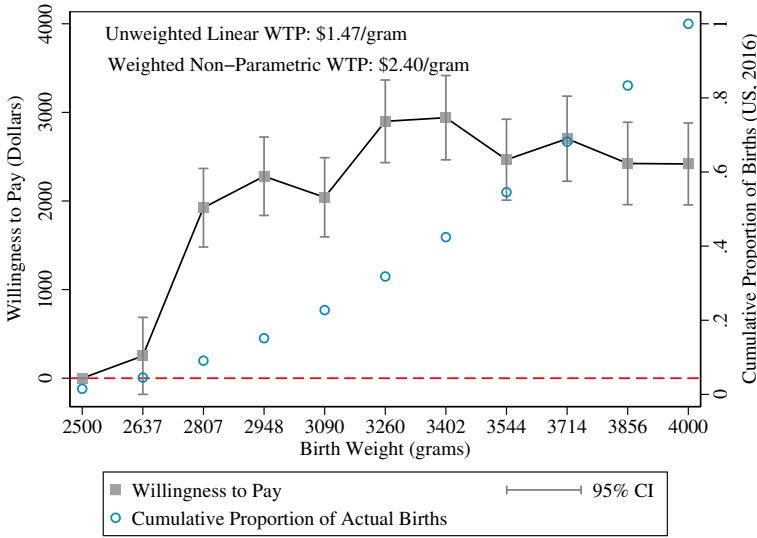
Figure 3. Relative willingness to pay for birth weight
*Notes*: Each point and confidence interval are with respect to the baseline (omitted) category of 2,500 g, the minimum displayed birth weight. Willingness to pay is determined as the ratio between the particular birth weight and out of pocket costs estimated as average marginal effects in a logit regression. 'Cumulative proportion of births' refers to the cumulative proportion in the displayed weight range of 2,500–4,000 g. 95% confidence intervals displayed are calculated using the delta method.

column 1, we assume a linear functional form for birth weight. By comparing the change in the likelihood of choosing a birth scenario based on an increase in birth weight with the change in likelihood due to an increase in costs, we estimate that the average WTP for an additional 1,000 g in the full sample is $1,470.3, or $1.47 per gram (95% CI: [$1.24, $1.70]). As expected, we observe that all else equal, higher costs result in a birth scenario being less likely to be preferred. On average, for each additional $1,000 in out of pocket expenses, the likelihood of choosing a birth scenario falls by 6.3 pp.[11]

When calculating the average WTP of birth weight as a single figure, this is based on a specification in which birth weight (and costs) enter the estimating equation linearly. However, as we observe in column 2 of Table 2, the relationship between birth weight and the likelihood of choosing a birth scenario is non-linear. In Figure 3, we document the WTP of all birth weight options, with respect to the minimum birth weight in the sample. We observe that the largest relative difference occurs at 3,400 g (7 lbs 8 oz, compared with the omitted base of 2,500 g). Using the non-parametric WTP estimates we obtain an average WTP of $2.40 per gram (95% CI: [$2.03, $2.77]).

It is also illustrative to compare WTP for birth weight to estimated WTP for other characteristics. Using point estimates from Table 2, we estimate a WTP for a girl

---

[11]In the online appendix Table A3 we follow Francis-Tan and Mialon (2015) and *re-weight* the sample so that it has a geographical distribution that is representative of the US population. We find quantitatively and qualitatively similar results.
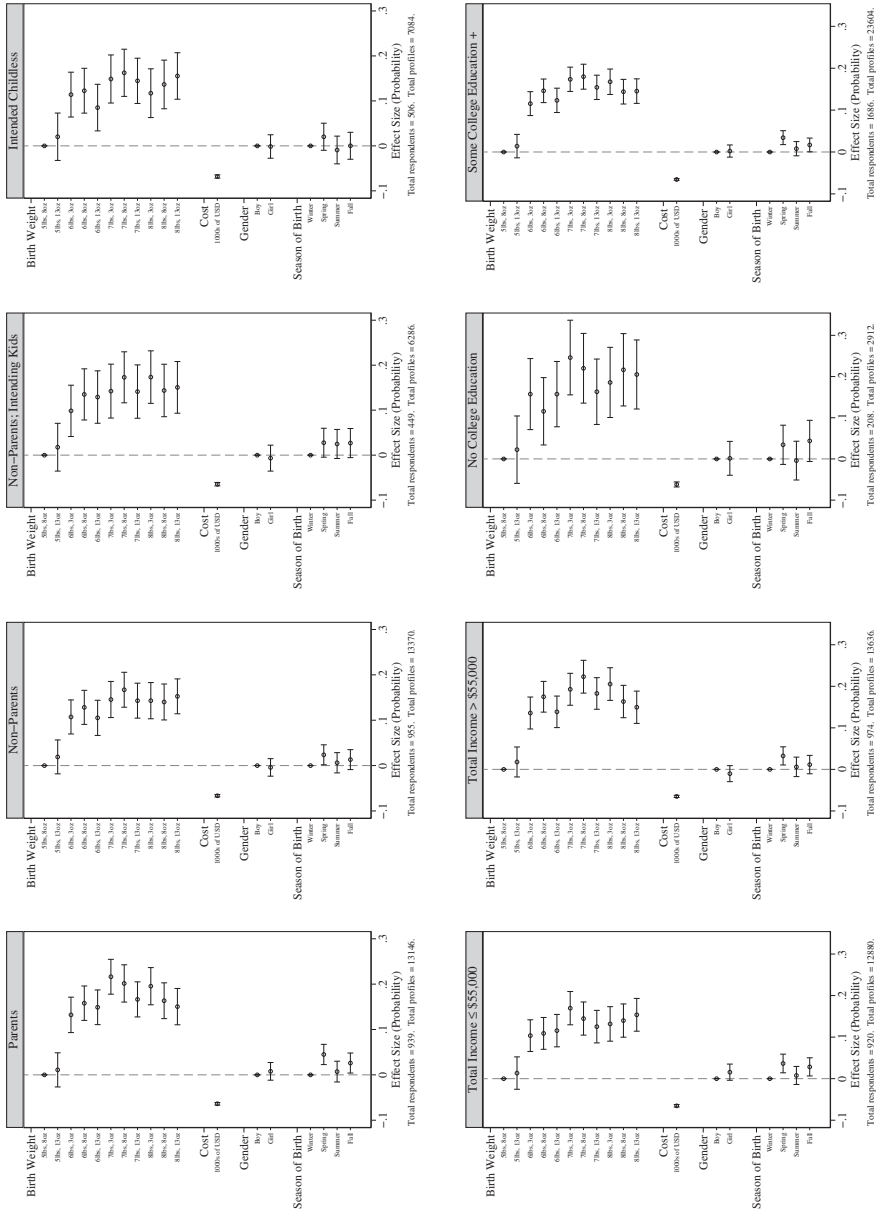
Figure 4. Heterogeneity

*Notes:* Methods are identical to those described in notes to Figure 2. The full sample is split by parents or non-parents (panels A and B), and then non-parents are split into those who report intending to have children (or already being pregnant) versus those who do not intend to have children 'Intended Childless' (panels C and D). Panels E and F compare responses for respondents reporting a total family income of 55,000 USD or lower with those reporting a total family income above 55,000 USD, and panels G and H present estimates by whether or not the individual has attained any college education.

TABLE 3

*Birth characteristics and willingness to pay for birth weight by parental status*

| | All | Parent | | Non-parents | |
| | | Yes | No | Intending | Not intending |
| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Birth weight (in 1000s of g) | 0.092*** [0.007] | 0.098*** [0.010] | 0.086*** [0.010] | 0.092*** [0.014] | 0.082*** [0.013] |
| Cost numerical | −0.063*** [0.001] | −0.061*** [0.001] | −0.064*** [0.001] | −0.062*** [0.002] | −0.065*** [0.002] |
| Girl | 0.003 [0.007] | 0.009 [0.010] | −0.004 [0.010] | −0.007 [0.015] | −0.001 [0.013] |
| Spring | 0.034*** [0.008] | 0.046*** [0.011] | 0.024** [0.011] | 0.026 [0.016] | 0.021 [0.015] |
| Summer | 0.006 [0.008] | 0.007 [0.012] | 0.005 [0.011] | 0.023 [0.016] | −0.009 [0.016] |
| Fall | 0.019** [0.008] | 0.026** [0.011] | 0.012 [0.011] | 0.026 [0.016] | −0.000 [0.015] |
| WTP for birth weight (1000 g) | 1,470.3 | 1,598.2 | 1,353.5 | 1,476.8 | 1,254.1 |
| 95% CI | [1,238.3, 1702.3] | [1,255.4, 1941.1] | [1,038.5, 1,668.5] | [1,010.4, 1,943.2] | [826.4, 1,681.9] |
| Observations | 26,516 | 13,146 | 13,370 | 6,286 | 7,084 |
| P-value for test of equality | | | 0.501 | | 0.676 |
| P-value for equality (2)=(5) | | | | | 0.419 |

*Notes*: Average marginal effects from a logit regression are displayed. All columns include option order fixed effects and round fixed effects. Standard errors are clustered by respondent. Willingness to pay and its 95% confidence interval is estimated based on the ratio of costs to the probability of choosing a particular birth weight. The 95% confidence interval is calculated using the delta method for the ratio. Intending and Not Intending in columns 4 and 5 refer to decisions regarding future children as outlined in Table 1. Tests of equality in the table footer refer to equality of WTP estimates between columns (2) and (3) and between columns (4) and (5) in the first test, and between WTPs estimated between columns (2) and (5), that is between parents and intended childless, in the second test. These *P*-values are directly estimated using a Chow test.
*$P$-value < 0.1, **$P$-value < 0.05, ***$P$-value < 0.01.

(rather than boy) birth of only $47, and estimate a WTP for a spring (rather than winter) birth of $539. Note that this can also be cast in terms of trade-offs related to birth weight. On average in the experimental sample, we estimate a willingness to accept 32 fewer grams of birth weight to achieve a girl birth, and 370 fewer grams of birth weight to achieve a preferred season (spring) birth.

**Results by parental status**

The headline estimated effect for average WTP suggests a value of $1.47 per gram over the range examined (95% CI: [$1.24, $1.70]). This value is calculated using the entire sample of respondents. We briefly consider estimates for particular subgroups of interest, namely parents, non-parents, and non-parents who do and do not intend to have children. All these basic demographic characteristics were asked *after* the completion of the experiments.

Figure 4 displays outcomes of the discrete choice experiments for each group. Panels A and B split by parental status (parents vs. non-parents), and then panels C and D further split non-parents by desired childbearing status (those who intend to have children or are already pregnant versus those who do not intend to have children). These figures reveal that parents are the most sensitive to changes in birth weight. Non-parents display a much flatter profile, and are consistently less likely to choose a birth scenario given a higher birth weight. When further splitting by those who intend to have children and those who do not, we observe that the profile for the former is comparable to that for parents, while those who do not intend are significantly less likely to choose a birth scenario based on an increase in weight. We examine these results, along with precise values for WTP, in Table 3

*Parents versus non-parents*
In columns 2 and 3 of Table 3, we estimate the linear specification for birth weight and costs for parents and all non-parents. We observe, first, that although both groups are similarly impacted by increases in costs (a birth scenario is 6.1 pp or 6.4 pp less likely to be chosen for each $1,000 increase in costs for parents and non-parents respectively), point estimates on birth weight are higher for parents than for non-parents. An increase in 1,000 g of birth weight increases the likelihood that parents choose a profile by 9.8 pp, while only by 8.6 pp for non-parents. This is reflected in different average WTP values. The average WTP for a gram of birth weight among *parents* is $1.60, (95% CI: [$1.26, $1.94]), compared to $1.35 among non-parents, (95% CI: [$1.04, $1.67]). Perhaps unsurprisingly, across the board parents are more likely than non-parents to be swayed by changes in non-pecuniary attributes: For parents, both birth weight and birth season are more important than for non-parents. We estimate a pooled specification where we interact a dummy for being a parent with birth weight in Appendix Table A4. This allows us to estimate the WTP differential between parents and non-parents and its 95% confidence interval. While the average WTP differential is notable – at $186 for an additional 1,000 g – it is not statistically distinguishable from zero at the 5% significance level (95% CI: [−$258, $631]). However, we reject the hypothesis that the average WTP among parents is less than or equal to that among non-parents in favour of the alternative hypothesis that the former is *larger* than the latter at the 5% significance level ($P = 0.042$) in a one-sided test.[12]

If parents are more educated, wealthier and/or older, their estimated WTP could owe to these differences, and potentially be higher than that of non-parents due to greater availability of financial resources and a different information set than non-parents (e.g. they know what a normal birth weight is, and may even appreciate its return). Since parents are both wealthier and older than non-parents (see Table A5 in the online appendix), it is important to check that the WTP for birth weight does not change once we interact characteristics where they differ (namely, age and income).

---

[12]The *P*-value for this one-sided hypothesis test is obtained as 1 minus the proportion of times that the estimated WTP among parents exceeds that of non-parents, when WTP for each group is calculated 500 times in a bootstrap resampling procedure, clustering over respondents.

The comparison of columns (1) and (2) in Table A6 in the online appendix reveals two interesting findings: (a) accounting for interactions between birth weight and individual characteristics leads to very similar point (and interval) estimates and (b) the interactions between birth weight and individual characteristics appear to be unrelated to the probability of choosing a birth scenario.[13]

*Parents versus non-parents who intend to have children*
Columns 4 and 5 of Table 3 display estimates for non-parents, separating by whether they intend to have children or do not intend to have children. If we compare figures for parents with those of non-parents who state that they *do* intend to have children, we see that the point estimate on birth weight is slightly higher among the former. As above, parents are 9.8 pp more likely to choose a birth scenario for each 1,000 g increase in birth weight, while the same figure for non-parents who intend to have children is 9.2 pp. The average WTP of the non-parent planners is \$1.48 per gram (cf 1.60 for parents) with a 95% CI of \$1.01–\$1.94. Once again, if we refer to the online appendix Table A4 we see that the difference in average WTP is not statistically significant (column 3) at the 5% significance level. Moreover, we cannot reject the hypothesis that the average WTP among parents is less than or equal to that among non-parents who intend to have children against the alternative that the former is *larger* than the latter ($P=0.194$).[14]

*Non-parents who intend to have children versus non-parents who do not intend to have children*
Finally, if we compare the two groups of non-parents, those who intend to have children and those who do not, we see a large average difference in the likelihood to choose a birth given an increase in birth weight. As above, non-parents who intend to have children are 9.2 pp more likely to choose a birth scenario for each 1,000 gram increase in birth weight, while non-parent non-planners are only 8.2 pp more likely. The average WTP for each group is \$1.48 per gram for those who intend to have children versus only \$1.25 per gram for those who do not intend to have children, or an 18% increase. We cannot formally reject the null hypothesis that the average WTP among non-parents who intend to have children is less than or equal to that among non-parents who do not intend to have children at the 10% significance level ($P = 0.317$).[15]

While the point estimates suggest the existence of heterogeneity in WTP for birth weight among groups, from \$1.25 per gram for individuals who intend to be childless

---

[13]For completeness, the online appendix Table A7 shows that allowing for heterogeneous valuations of birth weight by age, education and income does not affect our results in the full sample of respondents.

[14]Once again, after controlling for interactions between birth weight and individual characteristics (see Table A6), we obtain very similar point (and interval) estimates. In addition, the interactions between birth weight and individual characteristics appear to be unrelated to the probability of choosing a birth scenario among non-parents who intend to have children (*P*-value: 0.443 in column 3).

[15]When controlling for interactions between birth weight and individual characteristics (Table A6), we obtain similar point (and interval) estimates, and interaction terms are jointly insignificant at typical levels (*P*-value: 0.983 in column 4).

TABLE 4

*Birth characteristics and willingness to pay for birth weight by SES status*

| | All | Family income | | College | |
| --- | --- | --- | --- | --- | --- |
| | | ≤ $ 55,000 | > $ 55,000 | None | Some + |
| | (1) | (2) | (3) | (4) | (5) |
| Birth weight | 0.092*** | 0.085*** | 0.098*** | 0.122*** | 0.089*** |
| (in 1000s of g) | [0.007] | [0.010] | [0.010] | [0.022] | [0.007] |
| Cost numerical | −0.063*** | −0.063*** | −0.062*** | −0.059*** | −0.063*** |
| | [0.001] | [0.001] | [0.001] | [0.003] | [0.001] |
| Girl | 0.003 | 0.016 | −0.010 | −0.000 | 0.003 |
| | [0.007] | [0.010] | [0.010] | [0.021] | [0.007] |
| Spring | 0.034*** | 0.038*** | 0.031*** | 0.034 | 0.034*** |
| | [0.008] | [0.011] | [0.011] | [0.024] | [0.008] |
| Summer | 0.006 | 0.008 | 0.005 | -0.005 | 0.007 |
| | [0.008] | [0.011] | [0.012] | [0.024] | [0.009] |
| Fall | 0.019** | 0.029*** | 0.010 | 0.042 | 0.016* |
| | [0.008] | [0.011] | [0.011] | [0.026] | [0.008] |
| WTP for birth weight (1000 g) | 1,470.3 | 1,357.6 | 1,577.0 | 2,045.2 | 1,407.6 |
| 95% CI | [1,238.3, 1,702.3] | [1,028.3, 1,686.9] | [1,249.7, 1,904.2] | [1,250.0, 2,840.5] | [1,165.4, 1,649.8] |
| Observations | 26,516 | 12,880 | 13,636 | 2,912 | 23,604 |
| P-value for test of equality | | | 0.784 | | 0.373 |

*Notes*: Average marginal effects from a logit regression are displayed. All columns include option order fixed effects and round fixed effects. Standard errors are clustered by respondent. Willingness to pay and its 95% confidence interval is estimated based on the ratio of costs to the probability of choosing a particular birth weight. The 95% confidence interval is calculated using the delta method for the ratio. Estimates are presented separating by family income and by respondent educational level. Tests of equality in the table footer refer to equality of WTP estimates between columns (2) and (3) and between columns (4) and (5), that is testing for equality of WTP by income and by educational level. These *P*-values are directly estimated using a Chow test.
*$P$-value < 0.1, **$P$-value < 0.05, ***$P$-value < 0.01.

to $1.60 per gram for parents, the confidence intervals are quite large, so that homogeneity in WTP across groups is indeed compatible with our findings. This is recognized in the formal tests of equality of coefficients presented in Table 3, both when comparing parents with non-parents, and parents with only that sub-group who intend to remain childless. It is important to note that, while this magnitude in the difference between parents and those who intend to be childless may appear small (at 35 cents per gram according to point estimates), this difference accounts for nearly 30% of the *total* estimated WTP among intended childless, or 22% among parents. As we discuss at more length in the next section (in our comparison with the returns to birth weight in the labour market), there is evidence that birth weight may be considerably undervalued by individuals given its importance as a precursor for labour market and other lifetime outcomes.

**Results by demographic characteristics**

Finally, in panels E-H of Figure 4 and in Table 4 we consider heterogeneity by both respondents' education, and by their family income level. Once again, while we observe considerable heterogeneity in point estimates, these are not sufficiently precisely estimated to allow us to reject tests of equality of WTP between groups. In columns 2 and 3 of Table 4, we observe around a 20 cent difference in WTP per gram of birth weight when comparing families with total incomes below $55,000 USD to those with incomes above this threshold. This difference is significantly larger among individuals with at least some college and those with no college education. In this case, point estimates, while very noisy, suggest that those *without* college education have an estimated WTP of $2.05 USD per gram of birth weight compared with $1.41 for individuals with at least some college education. In both cases, these values are significantly different to zero, however given the small sample of non-college educated individuals, the confidence intervals of the two groups nearly overlap entirely.

In Appendix Table A8, we additionally consider heterogeneity by race and sex of the respondent. This analysis suggests that those reporting 'Other Race' or reporting being Asian have the highest WTP for birth weight followed by individuals reporting being White, and individuals reporting being Black. However the confidence intervals for the average WTP overlap across the various races. We also observe a higher WTP for birth weight among male respondents than among females but with overlapping confidence intervals once again. Clear divergent gender preferences (men preferring boys, women preferring girls) are also documented.

While these should be cast as exploratory analyses, Figure 4 provides some evidence that non-linearity in estimates turning negative at higher birth weights (in line with complications when babies are born at higher weights) are driven more clearly among higher income and college educated groups. In the case of families with a total income of above $55,000 USD, the non-linearity is observed, with WTPs falling from 7 lbs, 8 oz onwards, while no such reduction is observed among families with total incomes below 55,000 USD. Similarly, while based on noisy estimates, non-linearities are not clearly visible among non-college educated groups, potentially explaining higher WTPs in this group given higher point estimates across the entire birth weight distribution.

## V.   Assessing the validity of our experimental estimates

Our experimental estimates are subject to two potential limitations. First, they come from a convenience sample of US residents, namely MTurk respondents. As documented in the online appendix Table A2, MTurk users are different from the US population as a whole. Second, our estimates are based on hypothetical choices, and

one may be worried about hypothetical bias.[16,17] In order to check the validity of our estimates of the private valuation of birth weight, we compare our WTP for birth weight with that coming from observed behavioural changes owing to the impact of changes in birth timing due to financial incentives faced by parents.

After this comparison, we assess our experimental estimates for private WTP in light of a number of results from the economic literature on birth weight. In particular, we ask two questions: First, how does private WTP compare to the WTP inferred from public programmes?[18] Secondly, how does the private WTP compare to the total expected (labour market) benefits accruing to birth weight over the life cycle?[19]

### Comparison of experimental estimates with tax incentives

Schulkind and Shapiro (2014) discuss the impact of tax incentives on birth timing and birth outcomes in the United States. Using their estimates, we can consider what proportion of individuals are incentivised to shift their child's date of birth due to a $1,000 increase in tax incentives as well as what the impact of this incentive is on average birth weight.[20] If we combine Schulkind and Shapiro's estimate that 1 in 134 births are shifted due to a $1,000 tax incentive in the United States, and that this shift has an impact of between 2.41 and 6.37 g in the population, the scaled increase in birth weight on those who are estimated to change date of birth would be between $2.41 \times 134 = 323$ g and $6.37 \times 134 = 854$ g. When expressed in terms of the $1,000

---

[16]While it has been observed that results from hypothetical choices are nearly always replicated on average in *incentivised* choice experiments (Camerer and Hogarth, 1999), and that these results can agree very closely to true behaviour (Hainmueller *et al*., 2013), there are mixed opinions about the appropriateness of using hypothetical choice to value goods in economic research. Alternative perspectives on this are presented by Hausman (2012) and Carson (2012) in a symposium on contingent valuation.

[17]We additionally conduct a test examining whether individuals exhibit preference stability across rounds within the conjoint experiment. As noted in Harris, Gerstenblüth and Triunfo (2018 section 2B), testing preference stability requires holding constant the context of a choice experiment. We thus consider the sub-set of all pairs of profiles that are identical in terms of attributes for sex and season of birth. This lets us isolate changes in birth weight and cost differential between these two comparisons, and observe whether individuals make mutually inconsistent choices. For example, a mutually inconsistent choice would occur if an individual reveals that they are willing to pay *p* for an increase in birth weight of *q* given a particular set of sex and season of birth attributes, but then fails to pay *p'* for an increase in *q'* where $p/q \geq p'/q'$ when later facing the same sex and season of birth attributes. When we conduct this test, we find a relatively small number of cases where we can formally show that individuals have mutually inconsistent preferences, namely in 26 of the 14,035 pairs considered.

[18]We infer the WTP for birth weight from two large social safety net programmes. The first is WIC (the Special Supplemental Nutrition Program for Women Infants and Children), a programme which explicitly targets neonatal health, and the second is the Food Stamp Program (FSP), which, although not designed to target neonatal health outcomes, has been documented to have important impacts on early-life human capital measures.

[19]While the labour market returns to birth weight are a clear lower bound on the value of birth weight, these are all private returns, and so provide a benchmark value with which to compare the private WTP estimates discussed in the previous section.

[20]In terms of birth timing, Schulkind and Shapiro (2014) state "The estimate in Column (1) of 0.0037 indicates that a $1,000 increase in tax benefits is associated with approximately a 0.37 pp increase in the probability of a December birth. This point estimate corresponds to approximately 1 out of every 134 January births being moved to December for a $1,000 increase in tax benefits." And in considering birth weight, they state 'Evaluating the point estimate from Column (1) at its 95% confidence intervals suggests that an additional $1,000 in benefits causes between a 0.07% and 0.19% decrease in average birth weight, or between 2.41 and 6.37 g'.

incentive, this suggests that parents exhibit a WTP for birth weight of 1.17–3.09 per gram.[21]

This range is noteworthy for several reasons. First, it lines up with our experimental estimates from MTurk of $1.47–$2.40. Second, it is obtained from the same country. Finally, it refers to birth weights in a broadly normal range as in our MTurk experiment, given that parents changing the exact date of birth using rescheduled C-sections are those whose babies are in the healthy range, not those in high-risk pregnancies. Thus, it seems that our experimental findings are backed up by choices revealed in real world decisions. Of course, the consideration of tax incentives as a real-world 'natural experiment' is limited given that it is based on selected *compliers*, namely those parents who are willing to move birth timing based on financial incentives. But our findings suggest that in a controlled, albeit hypothetical, experimental setting similar average results are observed in a convenience sample, that is MTurk respondents.

Finally, inherent in the design of the DCE was the decision to focus only on the WTP for birth weight over the normal range of weights of 2,500–4,000 g. While the experimental design precludes the direct estimation of this WTP over the omitted range of LBWs, Almond *et al*. (2005) estimated hospital costs associated with LBW at $4.93 per gram. A back-of-the envelope calculation, where we take the estimate by Almond *et al*. (2005) as the average WTP for gram of birth weight over the range of *low* birth weights, gives an average WTP of $1.77 per gram.[22]

### Comparison with WTP from public programmes

It is of interest to ask how estimates of private WTP from this paper compare with the inferred WTP from public investment. While much of the benefits of increases in birth weight accrue to families, increases in birth weight also have important public returns, including benefits flowing from reductions in public health care spending, and lower usage of means-tested public benefits programmes (Almond *et al*., 2005). In the paragraphs below, we provide back-of-the-envelope calculations of the implied WTP for birth weight based on public investment. However, we note that when estimating returns using public programmes, these should be treated as strict upper bounds given that the benefits of the public programmes considered cover a large number of domains beyond simply early life health (refer to Clarke, Cortés Méndez and Vergara Sepúlveda (2020) for additional discussion). Thus, this exercise should be viewed as at best leading to tentative bounds.

*Comparison with the public WTP estimated from a targeted programme*
We can estimate the WTP for birth weight using estimates from the WIC, which provides food and education to pregnant and postpartum breastfeeding women who earn less than 185% of the US federal poverty guideline. By combining estimates of

---

[21]These values are given as $1,000/854 g = $1.17 per gram and $1,000/323 g = $3.09 per gram.
[22]This value is calculated as follows: $4.93×0.087+$1.47×0.913, where 0.087 and 0.913 are the fractions of LBW babies (500–2,499 g) and normal birth weight babies (2,500–4,000 g) from the population of US births (see Figure A6).

the cost per WIC user with estimates of the benefit in terms of additional birth weight, we can arrive at an estimate of the WTP per gram of birth weight. Ben-Shalom, Moffitt and Scholz (2011) document that WIC participation costs $54 per enrollee per month, and according to WIC administrative data, 56.9%, 34.7% and 7.8% of participants enroll in the first, second or third trimester respectively (Johnson *et al.*, 2013). Using trimester midpoints to calculate months of enrolment, this suggests approximate total costs of covering a single pregnant woman of $321. Among plausibly causal estimates of the impact of the WIC programme, Rossin-Slater (2013) estimates that participation has a mean impact of 27 g of birth weight, and Hoynes, Page and Stevens (2011) estimate impacts of 18–29 g. In the case of the highest estimated impact, the WTP based on the WIC equates to $321/29 g = $11.07 per gram, while for the lowest estimate, the WTP equates to $321/18 g = $17.83 per gram. Both estimates of the WTP based on the WIC exceed our experimental estimates of the private average WTP, for both the whole sample of respondents ($1.47), and the sub-sample of parents ($1.60).

*Comparison with the public WTP estimated from an untargeted programme*
The evidence from WIC discussed above estimates the inferred WTP using a targeted programme which explicitly focuses on maternal and newborn health. Nevertheless, there are a range of other public programmes which, while not explicitly targeting infant health, have been documented to have unintended effects on these outcomes. Perhaps the largest of these is the Food Stamp Program (FSP), now known as the Supplementary Nutrition Assistance Program (SNAP), which provided support for 44.2 million people in 2016 at a total cost of 70.9 billion dollars. Almond, Hoynes and Schanzenbach (2011) provide a particularly well-identified estimate of the effect of the FSP on infant health, and in particular, on birth weight. They estimate individual effects of programme exposure, which amount to 20.27 g for white pregnant women or 31.69 g for black pregnant women. This allows us to estimate the inferred WTP based on the FSP for a gram of birth weight when combined with the costs per pregnant women. In order to determine the costs per pregnant women, we focus on data on current costs and users (in order to be comparable to our estimated WTP in current dollars). Using the final 3 months of pregnancy to estimate the typical costs for a pregnant woman, and average per person monthly costs from 2016 of $134 (i.e. (71000/44)/12), the inferred WTP based on the FSP for an additional gram in birth weight is approximately $17.[23] Once again, the inferred WTP based on the FSP exceeds our estimates for the average private WTP by an order of magnitude.

[23]This is calculated using (134×3)/31.69=$12.7 based on Almond *et al.* (2011)'s estimates for black mothers and (134×3)/20.27=$19.8 for estimates for white mothers. In addition, we know that 40.2% of food stamp users are white and 25.7% are black (Gray, 2014). Hence, we can get a weighted average estimate using 40.2/65.9=0.61 as the weight for white mothers, and 25.7/65.9=0.39 as the weight for black mothers. This leads us to a weighted average of $17=0.61×$19.8+0.39×$12.7.

TABLE 5

*Estimates of the long run returns to birth weight in the US*

| Authors | Weight | Geographical area | Time period | Dependent variable | Estimate | Denominator return | Estimation strategy |
|---|---|---|---|---|---|---|---|
| Panel A: Income | | | | | | | |
| Behrman and Rosenzweig (2004) | $\mu$=90.2 oz ($\mu$=2,557) | Minnesota | 1936–1955 | ln(Wage) | 0.190 (0.077)[a] | oz/week pregnancy | Between MZ twin |
| Cook and Fletcher (2015) | $\mu$=3,367 | Wisconsin | 1957 HS graduates | ln(Wage) | 0.0997 (0.0788) | Birth weight (1 sd) | Between siblings |
| Johnson and Schoeni (2011) | NA | USA (PSID) | 1951–1975 | ln(Earnings) | −0.1667 (0.097) | LBW | Between siblings (males only) |
| Panel B: Completed education | | | | | | | |
| Royer (2009) | $\mu$=2,533 | California | 1960–1982 | Completed education (Years) | 0.16 (0.07) | 1,000 g (3500–2500 g) | Between twins (females only) |
| Currie and Moretti (2007) | $\mu$=3,268 | California | 1970–1974 | Completed education (Years) | −0.079 (0.014) | LBW | Between siblings (females only) |
| Conley and Bennett (2000) | Pr(LBW)=0.07 | USA (PSID) | 1968–1973 | Timely graduation | −2.024 (0.764) | LBW | Between siblings |

*Notes:* [a]Standard error is calculated based on *t*-statistic reported in original paper.

### Comparison with the returns to birth weight in the labour market

It is well accepted that higher birth weight is associated with reductions in morbidity and mortality, and greater educational attainment and achievement throughout childhood.[24] Moreover, these impacts have been well-documented to persist into adulthood and impact labour market outcomes (Bharadwaj, Lundborg and Rooth, 2015). In Table 5, we review the range of papers which have estimated the long-run returns to birth weight in the United States.[25] One way to benchmark (lower bounds) of the parental average WTP for birth weight is to determine how it compares to the flow of expected benefits during the life of their child. Thus, considering these well-estimated cases of the labour market returns to birth weight, we can discount expected returns back to the start of an individual's life, and compare it with our experimentally estimated WTP. This should of course be considered as a lower bound to the true value of birth weight.[26]

For this exercise, we are most interested in those papers which provide estimates of the long-run returns to birth weight in the labour market. Among those papers which have estimated the effect of birth weight on earnings, there are three papers that use twin or sibling fixed effects to leverage within family variation in birth weight to estimate returns conditional on genetic material. These are Behrman and Rosenzweig (2004), Johnson and Schoeni (2011), and Cook and Fletcher (2015). In order to generate a back-of-the-envelope comparison of the WTP for birth weight with the present value of expected labour market returns, we focus on the estimates of Behrman and Rosenzweig (2004). Behrman and Rosenzweig (2004)'s results provide a point estimate of the labour market returns to birth weight in the United States which suggests that 'augmenting a child's birth weight by a 1 lb. increases her adult earnings by over 7%'. According to the United States Census Bureau (2016), the median personal income in the United States in 2015 was $30,240. If we assume a working life which begins at the age of 25 and ends at the age of 60, we can calculate the present value of a 7% increase in wages as a deferred annuity. This calculation suggests that the present value of an additional pound of birth weight is $10,235.[27] Dividing this value by the 454 g in a pound gives the labour market value of a gram of weight of $23. If we assume that only approximately

---

[24]For example, on morbidity, Conley, Strully and Bennet (2003), Almond *et al*. (2005), Oreopoulos *et al*. (2008), and Gupta, Deding and Lausten (2013), and on early-life education, Lin and Liu (2009), Fletcher (2011), Torche and Echevarría (2011), Figlio *et al*. (2014), and Bharadwaj, Eberhard and Neilson (2017), demonstrate a strong and plausibly causal link.

[25]A number of similar estimates exist in a non-US setting (for example Rosenzweig and Zhang (2013) in China, Black *et al*. (2007) in Norway, and Currie and Hyson (1999) in Great Britain), however, in order to benchmark our WTP results in the US population we do not focus on these here.

[26]Labour market returns are a convenient financial metric, but do not include any of the additional pecuniary or non-pecuniary benefits which may flow to parents from a higher birth weight child such as lower expected costs associated with medical care (Almond *et al*., 2005) and the clear intrinsic value of health at birth, regardless of its impacts on economic circumstance during later life.

[27]We calculate the present value as

$$PVBW = (30{,}240 \times 0.07) \times \frac{1 - (1 + 0.05)^{-35}}{0.05} \times \frac{1}{(1 + 0.05)^{25}} = 10{,}235.46.$$

Note that in general, if anything our assumptions are conservative with regard to the estimated present value. For example, if we were not to discount this amount back to the age of 0, or if we were to discount using a lower discount rate to incorporate inflation, this would lead to higher calculated present values.

60% of the working age population will actually be employed in the labour market (United States Bureau of Labor Statistics, 2017), scaling by this value still suggests a labour market return of approximately $14, an order of magnitude higher than our estimated values of average WTP.

This lower bound calculation using Behrman and Rosenzweig (2004)'s estimates relies on a number of assumptions that are unlikely to hold in practice. Chief among these is that the returns to birth weight are stable over the life course, and salary and labour market participation rates are also stable over the life course. Still, we believe this is an informative estimate, if only because the $14 per gram is close to the WTP inferred from WIC and FSP ($11–$18 per gram), but 8–13 times larger than the private average WTP estimated among our respondents.

## VI. Allowing for preference heterogeneity

Our empirical analysis has used a traditional logit model, which assumes that the parameter associated with each birth attribute is *fixed* across individuals. In our case, this is tantamount to assuming *homogeneous* preferences over birth outcomes between individuals. In this subsection, we allow for preference *heterogeneity* in birth outcomes during our estimation process by specifying a *mixed* logit model (Revelt and Train 1998, McFadden and Train 2000 and Train 2003).

This procedure requires the use of a maximum simulated likelihood in place of maximum likelihood, however is now available in many standard software packages.[28] The parameter vector now consists of each individual's specific parameters, which give rise to the mean parameter in the sample as well as measures of its variance.

In the online appendix Table A9, we display the parameters estimated from the mixed logit, as well as the WTP for birth weight using the full sample and each sub-sample of interest. As is common in discrete choice applications with WTP, we model the price (out of pocket expenses) as *fixed* across respondents, while allowing all other coefficients (and preferences) to *vary*. This ensures that the WTP for each attribute is identified, as outlined in Revelt and Train (1998). In panel A, we display the mean estimates for each parameter, and in panel B, the standard deviation of each parameter. As is typical with the mixed logit, the normalisation of the parameters with respect to individual utility means that point estimates are significantly larger than those in the standard logit model. Nevertheless, we are more interested in the WTP of each parameter (as well as the distribution of parameters in the sample) rather than each parameter itself. *On average*, the WTP for birth weight is quite similar to that estimated in the standard logit model. For the full sample, the WTP from the Mixed Logit model is $1.68 per gram (95% CI: [$1.47, $1.90]). Similarly, we observe that this WTP is highest for parents at $1.79 per gram (95% CI: [$1.48, $2.11]), followed by non-parents who plan to have a birth ($1.72 per gram, 95% CI: [$1.29, $2.15]) and the lowest among non-parents who do not intend to have children ($1.38 per gram, 95% CI: [$0.99, $1.77]).

---

[28]See for example Hole (2007b) for a Stata implementation, or a series of packages made available by Kenneth Train in other languages (https://eml.berkeley.edu/ train/software.html).

Using both of these sets of parameters (mean and standard deviation), we are also able to determine the proportion of all respondents who positively value birth weight (and indeed any characteristic) in these linear specifications.[29] These values are displayed at the base of the table, indicating what proportion of respondents positively value birth weight. These values follow a similar pattern as those observed for WTP. Namely, parents and non-parents who intend to have children are the most likely to place a positive value on birth weight (70.7% and 71.0%, respectively), while non-parents who do not plan to have children are the least likely to assign a positive value (68.1%). Using the conditioning of individual taste (COIT) method described in Revelt and Train (2000) we are able to estimate the entire distribution of WTP across respondents, which we display in the online appendix Figure A9. This provides evidence of considerable heterogeneity in tastes for birth weight.

Finally, we extend the mixed Logit to our non-parametric specification where birth weight enters in categories as observed by respondents. The results for the WTP, as well as the percent of respondents who value each birth weight positively, are displayed in the online appendix Figure A10. These are all based on the mean and standard deviations of the parameters estimated from the mixed logit, as displayed in the footer of Table A9. In turning to the proportion of respondents who positively value each birth weight category, we observe that this quickly rises as birth weights diverge from the baseline reference category. Once reaching approximately 2,800 g, over 80% of all respondents prefer this to the baseline value of 2,500 g, and this value rises to close to 100% once exceeding approximately 2,950 g.

All in all, our experimental private WTP estimation is robust to allowing for preference heterogeneity.

## VII.   Conclusion

The use of birth weight as an individual's prominent measure of early-life endowment of human capital is now a well-established practice in the economic literature. Birth weight has increasingly been shown to be a modifiable outcome, being particularly responsive to certain policy measures. Despite considerable public investment in policies to increase birth weight and health at birth, very little is known about the *private* WTP for birth weight.

In this paper, we document that individuals have a positive, economically and statistically significant WTP for birth weight. We find that this WTP is higher among parents than non-parents, and higher among non-parents that intend to have children than among non-parents who do not intend to have children. Among all respondents, the average WTP for a gram of birth weight is estimated at $1.47, while among parents is estimated at $1.60. The average WTP based on non-parametric estimates among all respondents is estimated at $2.40 per gram.

While our experimental findings are based on hypothetical choices made by a convenience sample of respondents (MTurk workers), our range of estimates lines up

---

[29]These can be calculated using the entire vector of parameters, or alternatively as $100 \times \Phi(-\mu_k/\sigma_k)$, where $\Phi$ is the cumulative normal distribution, $\mu_k$ is parameter $k$'s mean, and $\sigma_k$ is its standard deviation.

with the WTP for birth weight inferred from the impact of tax incentives on birth timing and birth outcomes in the United States as reported in Schulkind and Shapiro (2014). Using the estimates from these authors, we compute the WTP for birth weight among parents who were willing and moved birth timing based on financial incentives at $1.17–$3.09 per gram.

Our findings suggest that parents have a WTP of about $1.60 per gram of birth weight, far too little in comparison to the lower-upper bounds of $14–$17 per gram implied from private returns and public investment. Whether this is a behavioural puzzle, is driven by imperfect altruism, or is due to the fact that these returns are simply the means of a very random process and so accounting for dispersion in the returns (risk) would lead any rational parent to have a much lower WTP is something that should be taken up in future research.

*Final Manuscript Received: January 2021*

# References

Almond, D. (2006). 'Is the 1918 influenza pandemic over? Long-term effects of In utero influenza exposure in the post-1940 U.S.Population', *Journal of Political Economy*, Vol. 114, pp. 672–712.

Almond, D. and Currie, J. (2011). 'Killing me softly: the fetal origins hypothesis', *Journal of Economic Perspectives*, Vol. 25, pp. 153–172.

Almond, D., Hoynes, H. W. and Schanzenbach, D. W. (2011). 'Inside the war on poverty: the impact of food stamps on birth outcomes', *The Review of Economics and Statistics*, Vol. 93, pp. 387–403.

Almond, D., Currie, J. and Duque, V. (2018). 'Childhood circumstances and adult outcomes: Act II', *Journal of Economic Literature*, Vol. 56, pp. 1360–1446.

Almond, D., Chay, K. Y. and Lee, D. S. (2005). 'The costs of low birth weight', *The Quarterly Journal of Economics*, Vol. 120, pp. 1031–1083.

Almond, D., Edlund, L. and Palme, M. (2009). 'Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in Sweden', *Quarterly Journal of Economics*, Vol. 124, pp. 1729–1772.

Bardos, J., Friedenthal, J., Spiegelman, J. and Williams, Z. (2016). 'Cloud based surveys to assess patient perceptions of health care: 1000 respondents in 3 days for US $300', *JMIR Research Protocols*, Vol. 5, pp. e166.

Barker, D. J. (1990). 'The fetal and infant origins of adult disease', *British Medical Journal*, Vol. 301, pp. 1111.

Barker, D. J. (1995). 'Fetal origins of coronary heart disease', *British Medical Journal*, Vol. 311, pp. 171–174.

Barker, D. J., Osmond, C., Winter, P., Margetts, B. and Simmonds, S. (1989). 'Weight in infancy and death from ischaemic heart disease', *The Lancet*, Vol. 334, pp. 577–580.

Behrman, J. R. and Rosenzweig, M. R. (2004). 'Returns to birthweight', *The Review of Economics and Statistics*, Vol. 86, pp. 586–601.

Bekker-Grob, E. W., Ryan, M. and Gerard, K. (2012). 'Discrete choice experiments in health economics: a review of the literature', *Health Economics*, Vol. 21, pp. 145–172.

Ben-Shalom, Y, Moffitt, RA., and Scholz, JK (2011). An Assessment of the Effectiveness of Anti-poverty Programs in the United States, National Bureau of Economic Research, Inc., NBER Working Papers 17042.

Berinsky, A. J., Huber, G. A. and Lenz, G. S. (2012). 'Evaluating online labor markets for experimental research: amazon.com's mechanical turk', *Political Analysis*, Vol. 20, pp. 351–368.

Bharadwaj, P., Eberhard, J. and Neilson, C. (2017). 'Health at birth, parental investments and academic outcomes', *Journal of Labor Economics*, Vol. 36, pp. 349–394.

Bharadwaj, P., Løken, K. V. and Neilson, C. (2013). 'Early life health interventions and academic achievement', *American Economic Review*, Vol. 103, pp. 1862–1891.

Bharadwaj, P., Lundborg, P. and Rooth, D.-O. (2015). *Birth Weight in the Long Run*, National Bureau of Economic Research, Inc., NBER Working Papers 21354.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2007). 'From the cradle to the labor market? The effect of birth weight on adult outcomes', *The Quarterly Journal of Economics*, Vol. 122, pp. 409–439.

Bridges, J. F. P., Brett Hauber, A., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., Reed Johnson, F., Mauskopf, J., Ryan, M. and Farrar, S. (2011). 'Conjoint analysis applications in health-a checklist: a report of the ISPOR good research practices for conjoint analysis task force', *Value in Health*, Vol. 14, pp. 403–413.

Camerer, C. F. and Hogarth, R. (1999). 'The effects of financial incentives in experiments: a review and capital-labor-production framework', *Journal of Risk and Uncertainty*, Vol. 19, pp. 7–42.

Carson, R. T. (2012). 'Contingent valuation: a practical alternative when prices aren't available', *Journal of Economic Perspectives*, Vol. 26, pp. 27–42.

Case, A., Fertig, A. and Paxson, C. (2005). 'The lasting impact of childhood health and circumstance', *Journal of Health Economics*, Vol. 24, pp. 365–389.

Chevalier, A., and O'Sullivan, V. (2007). Mother's Education and Birth Weight, Institute for the Study of Labor (IZA) IZA Discussion Papers 2640.

Clarke, D., Cortés Méndez, G. and Vergara Sepúlveda, D, (2020). 'Growing together: assessing equity and efficiency in a prenatal health program', *Journal of Population Economics*, Vol. 33, pp. 883–956.

Clarke, D., Oreffice, S., and Quintana-Domeque, C. (2017). On the Value of Birth Weight, Human Capital and Economic Opportunity Working Group Working Papers 2017-018.

Conley, D. and Bennett, N. G. (2000). 'Is biology destiny? Birth weight and life chances', *American Sociological Review*, Vol. 65, pp. 458–467.

Conley, D., Strully, K., and Bennet, N. G. (2003). A Pound of Flesh or Just Proxy? Using Tin Differences to Estimate the Effect of Birth Weight on Life Chances, National Bureau of Economic Research, Inc NBER Working Papers 9901.

Cook, C. J. and Fletcher, J. M. (2015). 'Understanding heterogeneity in the effects of birth weight on adult cognition and wages', *Journal of Health Economics*, Vol. 41, pp. 107–116.

Currie, J. and Moretti, E. (2007). 'Biology as destiny? Short- and long-run determinants of intergenerational transmission of birth weight', *Journal of Labor Economics*, Vol. 25, pp. 231–264.

Currie, J. and Hyson, R. (1999). 'Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight', *The American Economic Review*, Vol. 89, pp. 245–250.

Dickert-Conlin, S. and Chandra, A. (1999). 'Taxes and the timing of birth', *Journal of Political Economy*, Vol. 107, pp. 161–177.

Figlio, D., Guryan, J., Karbownik, K. and Roth, J. (2014). 'The effects of poor neonatal health on children's cognitive development', *American Economic Review*, Vol. 104, pp. 3921–3955.

Fletcher, J. M. (2011). 'The medium term schooling and health effects of low birth weight: Evidence from siblings', *Economics of Education Review*, Vol. 30, pp. 517–527.

Francis-Tan, A. and Mialon, H. M. (2015). '"A diamond is forever" and other fairy tales: the relationship between wedding expenses and marriage duration', *Economic Inquiry*, Vol. 53, pp. 1919–1930.

Gray, K. F. (2014). Characteristics of Supplemental Nutrition Assistance Program Households. Fiscal Year 2013. United States Department of Agriculture, Food and Nutrition Service Supplemental Nutrition Assistance Program Report SNAP-14-CHAR.

Gupta, N. D., Deding, M. and Lausten, M. (2013). 'The effect of low birth weight on height, weight and behavioral outcomes in the medium-run', *Economics & Human Biology*, Vol. 11, pp. 42–55.

Hainmueller, J., Hopkins, D. J. and Yamamoto, T. (2013). 'Causal inference in conjoint analysis: understanding multidimensional choices via stated preference experiments', *Political Analysis*, Vol. 22, pp. 1–30.

Hainmueller, J., Hangartner, D. and Yamamoto, T. (2015). 'Validating vignette and conjoint survey experiments against real-world behavior', *Proceedings of the National Academy of Sciences*, Vol. 112, pp. 2395–2400.

Harris, J. E., Gerstenblüth, M., and Triunfo, P. (2018). *Smokers' Rational Lexicographic Preferences for Cigarette Package Warnings: A Discrete Choice Experiment with Eye Tracking*, NBER Working Papers 24974. National Bureau of Economic Research, Inc., NBER Working Papers 24974.

Hausman, J. (2012). 'Contingent valuation: from dubious to hopeless', *Journal of Economic Perspectives*, Vol. 26, pp. 43–56.

Hole, A. R. (2007a). 'A comparison of approaches to estimating confidence intervals for willingness to pay measures', *Health Economics*, Vol. 8, pp. 827–840.

Hole, A. R. (2007b). 'Fitting mixed logit models by using maximum simulated likelihood', *The Stata Journal*, Vol. 3, pp. 388–401.

Holland, P. W. (1986). 'Statistics and causal inference', *Journal of the American Statistical Association*, Vol. 81, pp. 945–960.

Hoynes, H., Page, M. and Stevens, A. H. (2011). 'Can targeted transfers improve birth outcomes? Evidence from the introduction of the WIC program', *Journal of Public Economics*, Vol. 95, pp. 813–827.

Jamison, J. (2016). 'Perceptions regarding the value of life before and after birth? Reproductive system & sexual disorders', *Current Research*, Vol. 5, pp. 1000195.

Johnson, B., Thorn, B., McGill, B., Suchman, A., Mendelson, M., Patlan, K. L., Freeman, B., Gotlieb, R. and Connor, P. (2013). WIC Participant and Program Characteristics 2012 Final Report. U.S. Department of Agriculture, Food and Nutrition Service Prepared by Insight Policy Research under Contract No. AG-3198-C-11-0010 Alexandria, VA.

Johnson, R. C. and Schoeni, R. F. (2011). 'The influence of early-life events on human capital, health status, and labor market outcomes over the life course?', *The B.E. Journal of Economic Analysis & Policy*, Vol. 11, pp. 1–57.

Jordan, J. J., Hoffman, M., Bloom, P. and Rand, D. G. (2016). 'Third-party punishment as a costly signal of trustworthiness', *Nature*, Vol. 530, pp. 473–476.

King, M. T., Hall, J., Lancsar, E., Fiebig, D., Hossain, I., Louviere, J., Reddel, H. K. and Jenkins, C. R. (2007). 'Patient preferences for managing asthma: results from a discrete choice experiment', *Health Economics*, Vol. 16, pp. 703–717.

Kuziemko, I., Norton, M. I., Saez, E. and Stantcheva, S. (2015). 'How elastic are preferences for redistribution?',. Evidence from Randomized Survey Experiments', *American Economic Review*, Vol. 105, pp. 1478–1508.

LaLumia, S., Sallee, J. M. and Turner, N. (2015). 'New evidence on taxes and the timing of birth', *American Economic Journal: Economic Policy*, Vol. 7, pp. 258–293.

Lancaster, K. J. (1966). 'A new approach to consumer theory', *Journal of Political Economy*, Vol. 74, pp. 132–157.

Lin, M.-J. and Liu, J.-T. (2009). 'Do lower birth weight babies have lower grades? Twin fixed effect and instrumental variable method evidence from Taiwan', *Social Science & Medicine*, Vol. 68, pp. 1780–1787.

Marshall, D., Bridges, J. F. P., Hauber, B., Cameron, R., Donnalley, L., Fyie, K. and Reed Johnson, F. (2010). 'Conjoint analysis applications in health - how are studies being designed and reported?', *The Patient: Patient-Centered Outcomes Research*, Vol. 3, pp. 249–256.

Maruyama, S. and Heinesen, E. (2020). 'Another look at returns to birthweight', *Journal of Health Economics*, Vol. 70, pp. 1–20.

McFadden, D. and Train, K. (2000). 'Mixed MNL models for discrete response', *Journal of Applied Econometrics*, Vol. 15, pp. 447–470.

Moniz, M. H., Fendrick, A. M., Kolenic, G. E., Tilea, A., Admon, L. K. and Dalton, V. K. (2020). 'Out-of-pocket spending for maternity care among women with employer-based insurance', *Health Affairs*, Vol. 39, pp. 18–23.

Mortensen, K. and Hughes, T. L. (2018). 'Comparing amazon's mechanical turk platform to conventional data collection methods in the health and medical research literature', *Journal of General Internal Medicine*, Vol. 33, pp. 533–538.

Oreopoulos, P., Stabile, M., Walld, R. and Roos, L. L. (2008). 'Short-, medium-, and long-term consequences of poor infant health: an analysis using siblings and twins', *Journal of Human Resources*, Vol. 43, pp. 88–138.

Propper, C. (1990). 'Contingent valuation of time spent on NHS waiting list', *Economic Journal*, Vol. 100, pp. 193–199.

Propper, C. (1995). 'The disutility of time spent on the United Kingdom's national health service waiting lists', *Journal of Human Resources*, Vol. 30, pp. 677–700.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: households choices of appliance efficiency level', *The Review of Economics and Statistics*, Vol. 80, pp. 647–657.

Revelt, D. and Train, K. (2000). Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier. University of California at Berkeley Economics Working Papers E00-274.

Rosenzweig, M. R. and Zhang, J. (2013). 'Economic growth, comparative advantage, and gender differences in schooling outcomes: Evidence from the birthweight differences of Chinese twins', *Journal of Development Economics*, Vol. 104, pp. 245–260.

Rosenzweig, M. R. and Schultz, T. P. (1983). 'Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight', *Journal of Political Economy*, Vol. 91, pp. 723–746.

Rossin-Slater, M. (2013). 'WIC in your neighborhood: New evidence on the impacts of geographic access to clinics', *Journal of Public Economics*, Vol. 102, pp. 51–69.

Royer, H. (2009). 'Separated at girth: US twin estimates of the effects of birth weight', *American Economic Journal: Applied Economics*, Vol. 1, pp. 49–85.

Ryan, M. and Hughes, J. (1997). 'Using conjoint analysis to assess women's preferences for miscarriage management', *Health Economics*, Vol. 6, pp. 261–273.

Ryan, M. and Farrar, S. (2000). 'Using conjoint analysis to elicit preferences for health care', *British Medical Journal*, Vol. 320, pp. 1530–1533.

Schulkind, L. and Shapiro, T. M. (2014). 'What a difference a day makes: Quantifying the effects of birth timing manipulation on infant health', *Journal of Health Economics*, Vol. 33, pp. 139–158.

Sexton, M. and Hebel, J. R. (1984). 'A clinical trial of change in maternal smoking and its effects on birth weights', *Journal of the American Medical Association*, Vol. 251, pp. 911–915.

Torche, F. and Echevarría, G. (2011). 'The effect of birthweight on childhood cognitive development in a middle-income country', *International Journal of Epidemiology*, Vol. 40, pp. 1008–1018.

Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

United States Bureau of Labor Statistics (2017). Civilian Employment-Population Ratio [EMRATIO], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/EMRATIO, Accessed: 2017-01-26.

United States Census Bureau (2015). State Population Totals Datasets: 2010-2015?, https://www2.census.gov/programs-surveys/popest/datasets/2010-2015/state/asrh/scprc-est2015-18-pop-res.csv [Accessed: 26 January, 2017.]

United States Census Bureau. (2016). Selected Characteristics of People 15 Years and Over, by Total Money Income, Work Experience, Race, Hispanic Origin, and Sex, https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-01.html [Accessed: 25 January, 2017.]

Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E. W., Smit H. A. and de Wit, G. A. (2014). 'The effect of including an opt-out option in discrete choice experiments', *PLoS One*, Vol. 9, pp. 1–9.

Wittink, M. N., Cary, M., Tenhave, T., Baron, J. and Gallo, J. J. (2010). 'Towards patient-centered care for depression', *The Patient: Patient-Centered Outcomes Research*, Vol. 3, pp. 145–157.

Wu, C., Scott, , Hultman, C., Diegidio, P., Hermiz, S., Garimella, R., Crutchfield, T. M. and Lee, C. N. (2017). 'What do our patients truly want? Conjoint analysis of an aesthetic plastic surgery practice using internet crowdsourcing', *Aesthetic Surgery Journal*, Vol. 37, pp. 105–118.

Zizzo, D. J. (2010). 'Experimenter demand effects in economic experiments', *Experimental Economics*, Vol. 13, pp. 75–98.

Zweifel, P., Breyer, F., and Kifmann, M. (2009). *Health Economics*, Springer, Berlin.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Appendix**: Online Appendix for On the Value of Birth Weight.