## RESEARCH

# Do small effects matter more in vulnerable populations? an investigation using Environmental influences on Child Health Outcomes (ECHO) cohorts

Janet L. Peacock[1*], Susana Diaz Coto[1], Judy R. Rees[1], Odile Sauzet[2], Elizabeth T. Jensen[3], Raina Fichorova[4], Anne L. Dunlop[5], Nigel Paneth[6], Amy Padula[7], Tracey Woodruff[7], Rachel Morello-Frosch[8], Jessica Trowbridge[7], Dana Goin[9], Luis E. Maldonado[10], Zhongzheng Niu[10], Akhgar Ghassabian[11], Leonardo Transande[12], Assiamira Ferrara[13], Lisa A. Croen[13], Stacey Alexeeff[13], Carrie Breton[10], Augusto Litonjua[14], Thomas G. O'Connor[14], Kristen Lyall[15], Heather Volk[16], Akram Alshawabkeh[17], Justin Manjourides[17], Carlos A. Camargo Jr[18], Dana Dabelea[19], Christine W. Hockett[20,21], Casper G. Bendixsen[22], Irva Hertz-Picciotto[23], Rebecca J. Schmidt[23], Alison E. Hipwell[24], Kate Keenan[25], Catherine Karr[26], Kaja Z. LeWinn[27], Barry Lester[28], Marie Camerota[28], Jody Ganiban[29], Cynthia McEvoy[30], Michael R. Elliott[31], Sheela Sathyanarayana[32,33,34], Nan Ji[10], Joseph M. Braun[35], Margaret R. Karagas[1] and on behalf of Program Collaborators for Environmental influences on Child Health Outcomes

## Abstract

**Background** A major challenge in epidemiology is knowing when an exposure effect is large enough to be clinically important, in particular how to interpret a difference in mean outcome in unexposed/exposed groups. Where it can be calculated, the proportion/percentage beyond a suitable cut-point is useful in defining individuals at high risk to give a more meaningful outcome. In this simulation study we compute differences in outcome means and proportions that arise from hypothetical small effects in vulnerable sub-populations.

**Methods** Data from over 28,000 mother/child pairs belonging to the Environmental influences on Child Health Outcomes Program were used to examine the impact of hypothetical environmental exposures on mean birthweight, and low birthweight (LBW) (birthweight < 2500g). We computed mean birthweight in unexposed/exposed groups by sociodemographic categories (maternal education, health insurance, race, ethnicity) using a range of hypothetical exposure effect sizes. We compared the difference in mean birthweight and the percentage LBW, calculated using a distributional approach.

**Results** When the hypothetical mean exposure effect was fixed (at 50, 125, 167 or 250g), the absolute difference in % LBW (risk difference) was not constant but varied by socioeconomic categories. The risk differences were greater

*Correspondence:
Janet L. Peacock
janet.l.peacock@dartmouth.edu
Full list of author information is available at the end of the article

Peacock *et al. BMC Public Health*     (2024) 24:2655

Page 2 of 13

in sub-populations with the highest baseline percentages LBW: ranging from 3.1–5.3 percentage points for exposure effect of 125g. Similar patterns were seen for other mean exposure sizes simulated.

**Conclusions** Vulnerable sub-populations with greater baseline percentages at high risk fare worse when exposed to a small insult compared to the general population. This illustrates another facet of health disparity in vulnerable individuals.

**Keywords** Pregnancy outcomes, Child health outcome, Health disparities, Environmental exposure, Social determinants of health

## Introduction

A major challenge in epidemiological research is to determine whether or not an observed negative or positive effect of an exposure is large enough to be of clinical or public health importance. There are many ways that an exposure might impact the distribution of a continuous outcome and one summary measure is to compare the mean health outcome in exposed and unexposed groups. However, the clinical relevance of observed differences between such group means is often hard to interpret. To address this, researchers may compare an observed difference with a recognized minimal clinically meaningful difference (MCID), or if the MCID is unknown, Cohen's standardized effect size 'd' (estimate/standard deviation) where an effect size of 0.8 is considered 'large', 0.5 is 'medium' and 0.2 is 'small', with these terms interpreted as indicating importance of the effect size [1]. We note that the erroneous interpretation of the p-value as a measure of effect size and hence of clinical importance is sometimes still seen in papers [2]; underpowered non-significant studies are commonly interpreted as showing 'no difference' or 'no association' and conversely large studies showing statistical significance may not imply a clinically relevant effect. This circles back to the importance of reporting and understanding the actual effect sizes alongside the challenges in interpreting mean differences that are described above.

One approach to assist the interpretation of a mean difference is to additionally consider whether there is a known cut-point for the continuous health outcome that can be used to define individuals at 'high risk' of poor later outcome and to calculate the difference in the proportion at high risk in the exposed and unexposed groups – the absolute risk difference – to provide more information. For example, 'low birthweight' is commonly defined using the World Health Organization cut-point of 2500g to describe a dichotomous outcome. At the population level, this division can be used to determine the proportion of individuals with low versus normal birthweight who are more likely to suffer a poor perinatal outcome [3].

The calculation of the complementary dichotomous outcome, the absolute risk difference, based on the proportion at high risk, can be reported alongside the difference in means to give a 'dual outcome' that provides more information and so helps with the interpretation of study results. Further, if the dichotomized outcome is calculated using the whole distribution using a distributional approach [4–7], then the usual loss in precision associated with dichotomization does not occur and a study can provide a fully powered continuous and dichotomous outcome with the same sample size. This dual outcome approach was used to determine the potential impact of a statistically significant but very small difference in mean lung function z-score by randomized group in a RCT follow up [8].

A further challenge arises in that these two measures of effect size, the difference in means and the absolute risk difference are not constant across different populations. This is because the proportion at high risk and hence the difference between those exposed and not exposed is greater in vulnerable populations whose baseline mean outcome is usually more extreme than that of the general population [9]. To illustrate this consider the impact of an exposure causing a small reduction in mean birthweight in births that are full-term and preterm. As the risk of low birthweight is already higher among preterm than term births, even a small shift in mean birthweight in a preterm population has a relatively greater impact in the proportion with low birthweight. In general, small differences in mean health outcomes in vulnerable populations may be missed due to insufficient statistical power – they are 'non-significant', or they may be dismissed as too small to be clinically important despite the potential for a larger proportion of the population to move into a high risk category.

In this paper we use data from the pan-US NIH-funded "Environmental influences on Child Health Outcomes" (ECHO) program to simulate the impact of hypothetical small differences on a child health outcome in

Peacock *et al. BMC Public Health* (2024) 24:2655

Page 3 of 13

sub-populations defined by socioeconomic variables. The aim and focus of this paper is not to report new findings on effect sizes for specific outcomes in themselves but to illustrate the impact of environmental exposures across a range of plausible effect sizes. In this way we highlight the importance of the choice of outcome to assist ECHO and other researchers in child health in interpreting small effects.

## Methods
### Study participants
The Environmental influences on Child Health Outcomes (ECHO) program http://www.echochildren.org/ funded by NIH, was established in 2016 to investigate the impacts of a wide range of environmental exposures on children's health to "enhance the health of children for generations to come" [10, 11]. In cycle 1, 2016–2023, the ECHO program included 69 cohorts in 31 consortia representing the diversity in sociodemographic features and exposures among children born in the United States. The cycle 1 program includes participant data for over 43,000 pregnancies and children who have consented to the ECHO-wide protocol. Further details of the study sample are given in the Additional file (eMethods1).

### Characterization of health outcomes by social factors
These ECHO-wide cohort data were used to characterize the distributions of the birthweight outcome according to a range of populations stratified by known social determinants of health. This outcome was chosen to illustrate a key perinatal child health outcome. The following social factors were included: maternal education (categorized as less than high school, high school, college-no degree, bachelor's degree, master's degree or higher), health insurance ('HI': No HI/public HI, employer/market/private HI, more than one HI), race (White, Black, Native Hawaiian/Pacific Islander, American Indian/Alaskan Native, More than one race, Other races) and ethnicity (non-Hispanic, Hispanic). Children's race and ethnicity were based on caregiver reports and categories were collapsed (e.g. into More than one race, Other races) where the numbers were small. All social factors were included as proxy indicators of socio-environmental health risks [12].

We chose to examine the effect of a set of fixed absolute mean differences since absolute mean difference is commonly used to characterize the effect of an exposure on a continuous outcome. The baseline mean outcome and standard deviation (SD) values in each subgroup were used as the mean and SD in the respective unexposed subgroups. The SDs were assumed to be the same in each pair of unexposed and exposed sub-groups. We then considered four plausible scenarios to illustrate the effect of a hypothetical exposure on a population: i) a very small

effect on mean birthweight where the mean is reduced by 50g, ii) a small effect where mean birthweight is reduced by 125g, iii) a medium effect where mean birthweight is reduced by 167g, and iv) a larger effect where the mean is reduced by 250g. These were used to assess the effect on the difference in population at high risk between the groups. We assume that a reduction in mean indicates a poorer birthweight [13]. For each of the four scenarios we computed the equivalent mean outcome in the hypothetically unexposed and exposed groups and the proportions at high risk using below 2500g, 'low birthweight' to indicate high risk. These analyses were repeated by population subgroup according to the socioeconomic categories.

Figure 1 demonstrate the principles using hypothetical data to illustrate the distribution of the unexposed and exposed groups and the proportion at high risk (z < -1.645 or below the 5th percentile) and shows that the proportions at high risk are much greater in the vulnerable populations and so the impact of the same small decrease in the mean is greater than it is in general populations.

The proportions at high risk and their differences were calculated using a distributional approach, a statistical method that uses the whole distribution to calculate the proportion beyond a given cut-point, to avoid the loss of power and precision usually associated with dichotomization. This method works in a similar way to the calculation of reference ranges [4–6]. The distributional method used permits analysis with a range of distributional forms, symmetric and skewed, and is described more fully in the Additional file (eMethods2). Analyses were conducted using the R 4.2.2 package DistdichoR [14, 15].

## Results
Data were drawn from the first phase of the ECHO program (ECHO Cycle 1, 2016–2023) as part of 50 participant cohorts listed in Additional file eTable1. Analyses included a maximum of 28,496 mother/child pairs with a birthweight recorded (Table 1).

### Birthweight (g) and Low Birthweight (LBW)
Mean (SD) birthweight was 3308g (569) overall and the distribution was slightly negatively skewed (Additional file eFigure1). The birthweight means varied by the social factors analyzed: ranging from 3190 to 3345g in maternal education categories, 3266g to 3363g by health insurance status, 3131g to 3373g by race and 3243g to 3319g by ethnic group (Table 1). In parallel the percentage of infants with low birthweight (LBW) was 7.8% overall and varied markedly by maternal education, health insurance, race and ethnicity. The participant categories with the highest percentage of LBW were mothers with less than high school education (12.2%) and Black race (13.5%).
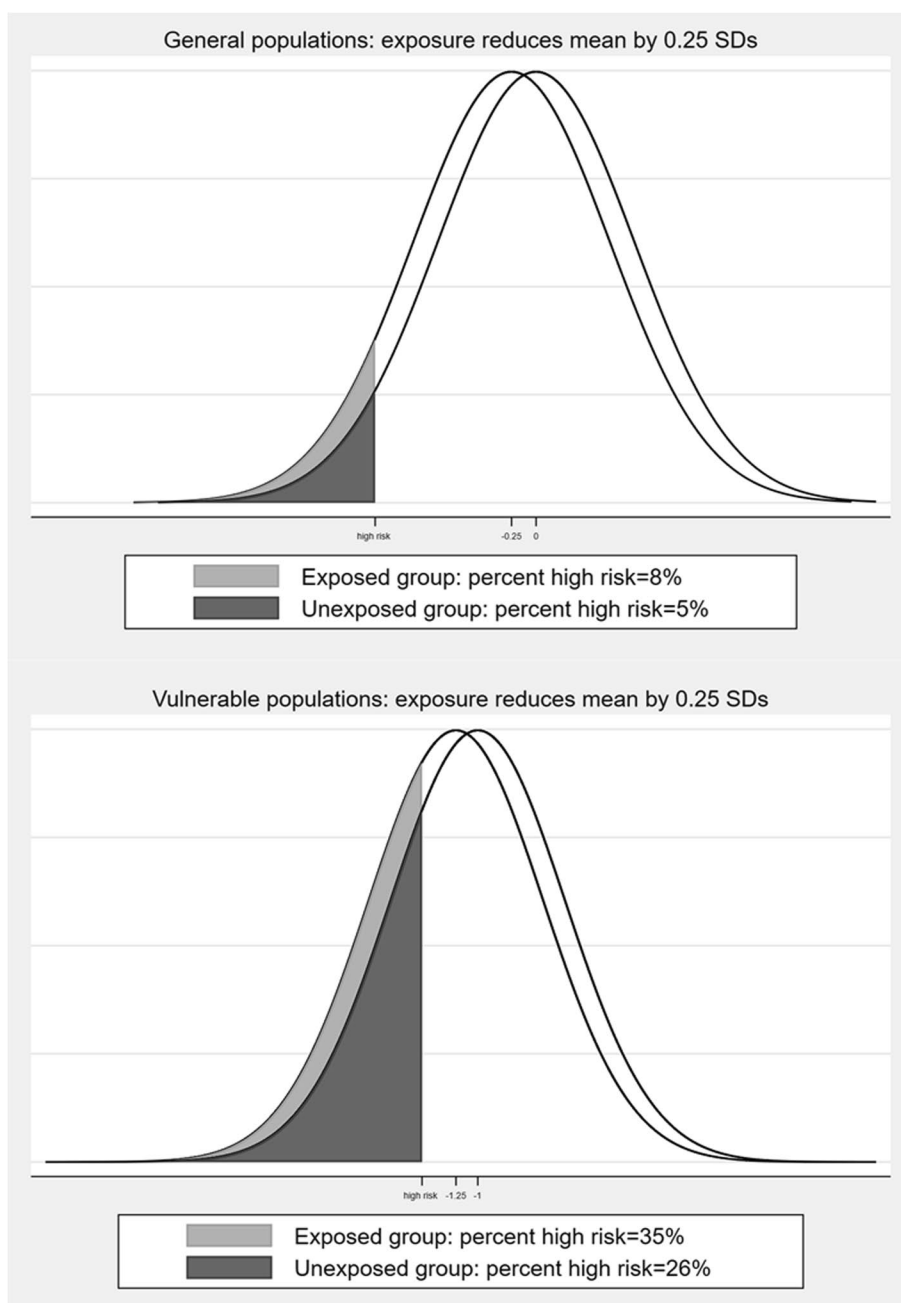
Peacock *et al. BMC Public Health*      (2024) 24:2655

Page 4 of 13



**Fig. 1** Illustration of the impact of a small effect, 0.25 standard deviations, in general populations compared with vulnerable populations showing that the same decrease in mean value has a greater impact in vulnerable populations. The hypothetical distributions are standard Normal with means in the unexposed distribution of 0 (general population) and -1 (vulnerable population), and standard deviation 1. High risk is defined as below 5th centile (< -1.645)

Table 2 and Fig. 2 depict the impact of a hypothetical and constant small change in mean birthweight, 50g, on the percentage of infants with low birthweight (% LBW) according to the categories of the four social factors analyzed. This shows the expected trend in LBW by maternal education whereby % LBW is inversely associated with educational attainment. Further, the absolute difference in % LBW was not constant, ranging from 1.2 percentage points to 1.8 percentage points, and was greatest for those with the highest baseline % LBW. A similar pattern was seen for LBW by race where Black mothers had more than twice the % LBW compared to White mothers

Peacock *et al. BMC Public Health* (2024) 24:2655

Page 5 of 13

**Table 1** Characteristics of the ECHO participants included in the study by outcome: Birthweight (*N* = 28,496)

| | N | Mean | SD | % Low birthweight |
|---|---|---|---|---|
| **All participants** | 28,496 | 3308 | 569 | 7.8% |
| **Maternal education** | | | | |
| Less than high school | 1064 | 3190 | 591 | 12.2% |
| High school | 2283 | 3249 | 579 | 9.8% |
| College, no degree | 4985 | 3273 | 601 | 9.9% |
| Bachelor's degree | 5557 | 3340 | 584 | 7.5% |
| Master's degree or higher | 5177 | 3345 | 558 | 6.5% |
| Missing/unknown | 9430 | 3315 | 539 | 6.5% |
| **Health Insurance (HI)** | | | | |
| No HI/public HI | 6781 | 3266 | 577 | 9.2% |
| Employer/market/private | 9743 | 3363 | 552 | 5.9% |
| More than one | 2916 | 3273 | 591 | 9.5% |
| Missing/unknown | 9056 | 3292 | 569 | 8.2% |
| **Race** | | | | |
| White | 16,393 | 3373 | 557 | 5.9% |
| Black | 4303 | 3131 | 573 | 13.5% |
| Asian | 993 | 3152 | 522 | 10.6% |
| Native Hawaiian/Pacific Islands | 102 | 3285 | 556 | 7.9% |
| American Indian/Alaskan Native | 570 | 3356 | 600 | 7.7% |
| Multiple Races | 3002 | 3271 | 585 | 9.4% |
| Other races | 1295 | 3249 | 568 | 9.4% |
| Missing/unknown | 1838 | 3309 | 538 | 6.6% |
| **Ethnicity** | | | | |
| Hispanic | 7174 | 3283 | 563 | 8.2% |
| Non-Hispanic | 20,450 | 3319 | 572 | 7.6% |
| Missing/unknown | 872 | 3243 | 540 | 8.5% |

**Table 2** Modeled proportions with low birthweight (< 2500g) associated with a shift in mean of 50g, overall and by social determinants in the ECHO consortium (*N* = 28,496)

| | % LBW unexposed | % LBW exposed | Difference in percentage points (exposed-unexposed) | 95% Confidence Interval |
|---|---|---|---|---|
| **All participants** | 7.8 | 9.1 | 1.4 | 1.1, 1.6 |
| **Maternal education** | | | | |
| Less than high school | 12.2 | 14.0 | 1.8 | 0.0, 3.6 |
| High school | 9.8 | 11.4 | 1.6 | 0.5, 2.6 |
| College, no degree | 9.9 | 11.5 | 1.5 | 0.8, 2.3 |
| Bachelor's degree | 7.5 | 8.8 | 1.3 | 0.7, 1.9 |
| Master's degree or higher | 6.5 | 7.7 | 1.2 | 0.7, 1.7 |
| Missing/unknown | 6.5 | 7.8 | 1.3 | 0.9, 1.7 |
| **Health Insurance (HI)** | | | | |
| No HI/public HI | 9.2 | 10.8 | 1.5 | 0.9, 2.1 |
| Employer/market/private | 5.9 | 7.1 | 1.1 | 0.8, 1.5 |
| More than one | 9.5 | 11.1 | 1.5 | 0.6, 2.4 |
| Missing/unknown | 8.2 | 9.6 | 1.4 | 0.9, 1.9 |
| **Race** | | | | |
| White | 5.9 | 7.0 | 1.1 | 0.9, 1.4 |
| Black | 13.5 | 15.5 | 2.0 | 1.0, 3.0 |
| Asian | 10.6 | 12.4 | 1.9 | 0.2, 3.6 |
| Native Hawaiian/Pacific Islands | 7.9 | 9.3 | 1.4 | -2.9, 5.7 |
| American Indian/Alaskan Native | 7.7 | 9.0 | 1.3 | -0.5, 3.1 |
| Multiple Races | 9.4 | 10.9 | 1.5 | 0.6, 2.4 |
| Other races | 9.4 | 10.9 | 1.6 | 0.2, 2.9 |
| Missing/unknown | 6.6 | 7.9 | 1.3 | 0.4, 2.2 |
| **Ethnicity** | | | | |
| Hispanic | 8.2 | 9.7 | 1.4 | 0.9, 2.0 |
| Non-Hispanic | 7.6 | 8.9 | 1.3 | 1.0, 1.6 |
| Missing/unknown | 8.5 | 10.0 | 1.5 | 0.0, 3.1 |

(13.5% vs 5.9%) and the same small decrease in mean birthweight was associated with a greater increase in % LBW among Black than White mothers (2.0% vs 1.1%) (Table 2, Fig. 2). In general, the offspring of women identifying as non-White race, i.e. Asian, Native Hawaiian/Pacific Island, American Indian/Alaskan Natives, Multiple races and Other races fared worse than the offspring of White mothers in terms of % LBW and the impact of a constant small decrease in mean birthweight. The patterns by health insurance and ethnic group were visible but less marked (Table 2, Fig. 2). Overall, the greatest risk difference was seen in Black children, 2.0 percentage points, and least risk difference was seen in White children, 1.1 percentage points when the exposure shifted mean birthweight by 50g.

Almost identical overall patterns were seen when mean birthweight was reduced by 125g, 167g and 250g with the effect sizes in differences of proportions increasing as expected. The trend in risk was particularly marked for maternal education and race (Table 3, eTable 2-eTable 3, Figs. 3– 5): for example the risk difference in Black children was 5.3% points compared to 3.1% points in White children for an exposure effect of 125g Table 3, Fig. 3), and was 7.4% (11.8%) vs 4.4% (7.3%) points for the exposure effect of 167g (250g) (eTables 2–3, Figs. 4– 5). Figure 6 (eTable 4) shows the overall relationship between mean birthweight in each of the 21 sub-populations that comprised the four social factors, by the size of the exposure, 50-250g. This figure displays the clear and strong inverse relationship between mean birthweight and the difference in percentage LBW across all categories.

In general, 95% confidence intervals around the estimated difference in proportions were narrow although they were naturally wider in small subgroups.
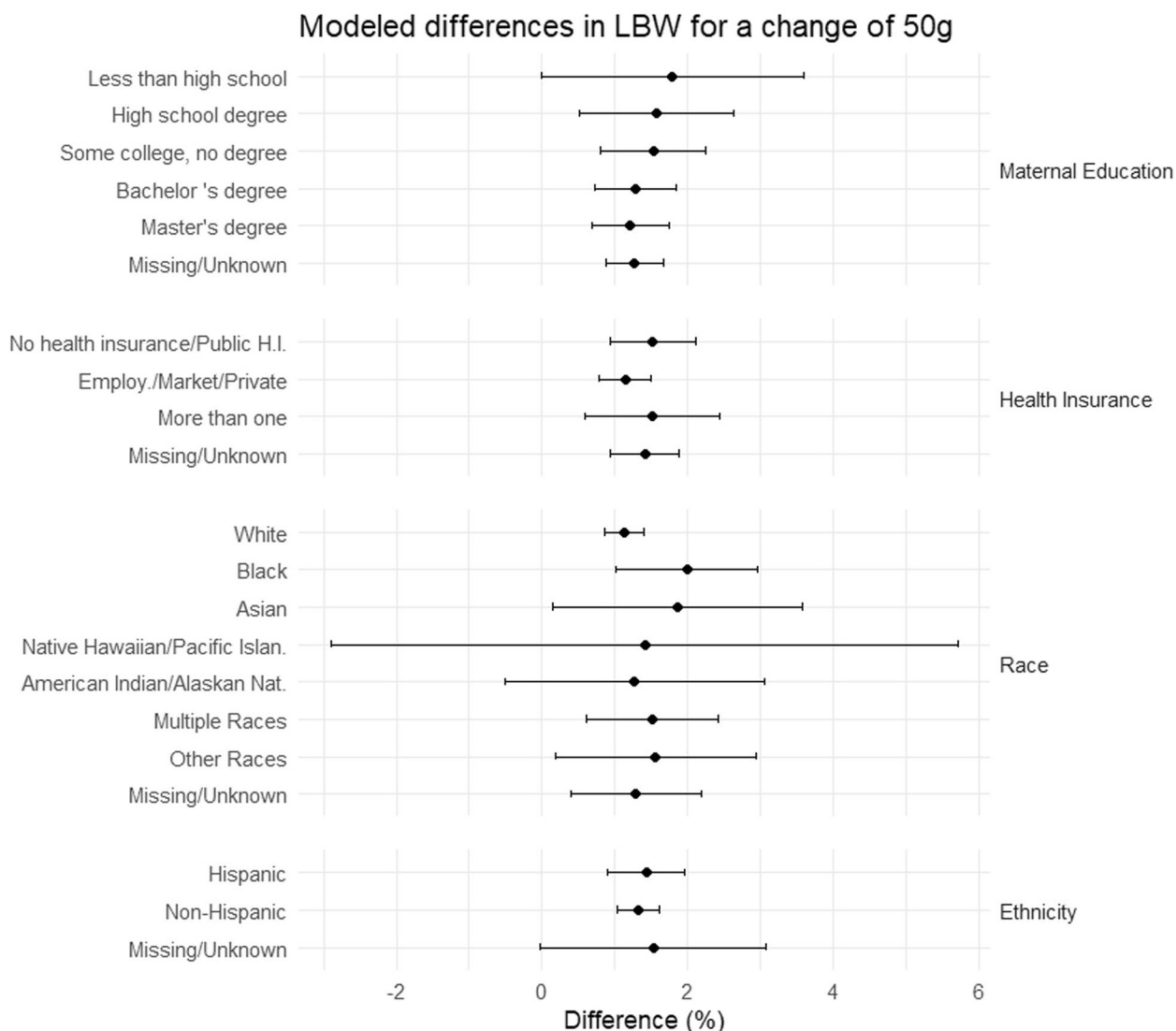
Peacock *et al. BMC Public Health*      (2024) 24:2655

Page 6 of 13



**Fig. 2** Modeled percentage of low birthweight (LBW) in unexposed and exposed populations by social factors associated with a change in mean birthweight of 50g (*N*=28,496)

## Discussion

This simulation study has shown the variability in distribution of a key child health outcome according to four social factors and illustrates the considerable disparities in children's health across the United States. The analyses have shown clear differences in both mean birthweight and the percentage of children with low birthweight (LBW) across categories of maternal education, health insurance status, race and ethnicity. Particularly marked effects were seen for maternal education ranging from 6.5% LBW to 12.2% and even more extreme, for race with 13.5% LBW among Black mothers compared to 5.9% in White mothers.

The current paper sought to explore the potential for adverse exposures to have a disproportionately harmful impact on the most vulnerable. Using a statistical approach, we have looked at the proportion of participants whose health outcomes put them in a high risk category to see how that risk changes when a hypothetical environmental exposure leads to a small shift in their outcome distribution's mean value [13]. Our findings have shown that the highest-risk individuals (i.e., at the extreme of an outcome distribution) are impacted more than those who are less vulnerable when both are exposed to the same small insult. The most striking contrast was in analyses by race where the same a

Peacock *et al. BMC Public Health* (2024) 24:2655

Page 7 of 13

**Table 3** Modeled proportions with low birthweight (< 2500g) associated with, a shift in means of 125g overall and by social determinants in the ECHO consortium (*N* = 28,496)

| | % LBW unexposed | % LBW exposed | Difference in percentage points (exposed-unexposed) | 95% Confidence Interval |
|---|---|---|---|---|
| **All participants** | 7.8 | 11.5 | 3.7 | 3.4, 4.0 |
| **Maternal education** | | | | |
| Less than high school | 12.2 | 17.0 | 4.8 | 2.9, 6.8 |
| High school | 9.8 | 14.1 | 4.3 | 3.1, 5.4 |
| College, no degree | 9.9 | 14.1 | 4.1 | 3.3, 4.9 |
| Bachelor's degree | 7.5 | 11.0 | 3.5 | 2.9, 4.1 |
| Master's degree or higher | 6.5 | 9.8 | 3.4 | 2.8, 3.9 |
| Missing/unknown | 6.5 | 10.0 | 3.5 | 3.1, 3.9 |
| **Health Insurance (HI)** | | | | |
| No HI/public HI | 9.2 | 13.4 | 4.1 | 3.5, 4.8 |
| Employer/market/ private | 5.9 | 9.1 | 3.2 | 2.8, 3.6 |
| More than one | 9.5 | 13.6 | 4.1 | 3.1, 5.1 |
| Missing/unknown | 8.2 | 12.1 | 3.9 | 3.3, 4.4 |
| **Race** | | | | |
| White | 5.9 | 9.0 | 3.1 | 2.8, 3.4 |
| Black | 13.5 | 18.8 | 5.3 | 4.3, 6.4 |
| Asian | 10.6 | 15.6 | 5.1 | 3.2, 6.9 |
| Native Hawaiian/ Pacific Islands | 7.9 | 11.8 | 3.9 | -0.9, 8.6 |
| American Indian/ Alaskan Native | 7.7 | 11.2 | 3.5 | 1.5, 5.4 |
| Multiple Races | 9.4 | 13.5 | 4.1 | 3.1, 5.1 |
| Other races | 9.4 | 13.6 | 4.2 | 2.7, 5.7 |
| Missing/unknown | 6.6 | 10.2 | 3.6 | 2.5, 4.6 |
| **Ethnicity** | | | | |
| Hispanic | 8.2 | 12.1 | 3.9 | 3.3, 4.5 |
| Non-Hispanic | 7.6 | 11.2 | 3.6 | 3.3, 4.0 |
| Missing/unknown | 8.5 | 12.6 | 4.2 | 2.5, 5.9 |

hypothetical exposure would have a markedly greater impact in Black mothers who are already more vulnerable than White mothers due to their higher baseline risk of low birthweight (LBW): the respective LBW risk differences are 2.0 versus 1.1 percentage points – an almost twofold difference for a 50g mean birthweight exposure effect, and increasing to 11.8 versus 7.3 percentage points for a 250g mean birthweight exposure effect. Hence the disparities were observed for any exposure effects. This illustrates the compounded health disparities in already vulnerable women and their children when they are subject to adverse environmental exposures.

These findings are particularly relevant for ECHO-wide studies where the broad aim is to identify adverse effects of environmental exposures. In particular, we need to understand how environmental exposures impact not only the population as a whole but also sub-populations, particularly those that are more vulnerable because their baseline risk is already high. An understanding of impacts of environmental exposures in vulnerable populations is critical in building the evidence base that will guide public and environmental health policy and preventive programs.

This study reinforces the importance of the statistical phenomenon that a constant exposure effect expressed as a shift in mean of an outcome does not equate to a constant shift in the proportions at high risk. This can be expressed in terms of Cohen's indices for effect sizes of differences in means ('Cohen's d') and proportions ('Cohen's h') [1], and shows that for sub-populations within the same overall population, the relationship between d and h is not constant. Hence small effects might be dismissed as not clinically important or are non-significant overall and might be of clinical relevance in vulnerable sub-populations [9]. In order to avoid missing important effects, analyses should be stratified by sociodemographic factors with both mean values and proportions at high risk reported for continuous outcomes. Further the potential for a small shift in mean to have important impact in vulnerable groups requires consideration in study planning and analysis and in designing preventive programs where seemingly small changes might be very important.

Limitations of this work include the choice of the child health outcome and the four social factors. These were chosen to give representation of outcomes measured at birth and of key social indicators, but we acknowledge this represents a small subset of the vast number of outcomes and social factors assessed in ECHO. We could have added more social factors and/or more outcomes, but we believe that the results would be the same and that adding more data risks obscuring the message. We also recognize that the ECHO cycle 1 sample is skewed towards higher educated, higher income women and so there may be less power to detect effects at the lower end of the distribution, as shown by the wider confidence intervals in the small subgroups.

The use of the distributional method to calculate the proportions at high risk avoided the usual loss of precision associated with dichotomization [4]. However these estimates have been computed to illustrate principles rather than to be used to generate hypotheses and should be treated as such. We could have chosen to use quantile regression to estimate the distribution tail areas; previous work has indicated that estimates using a distributional approach and quantile regression would be similar but that estimates calculated using quantile regression are less precise [7]. However,
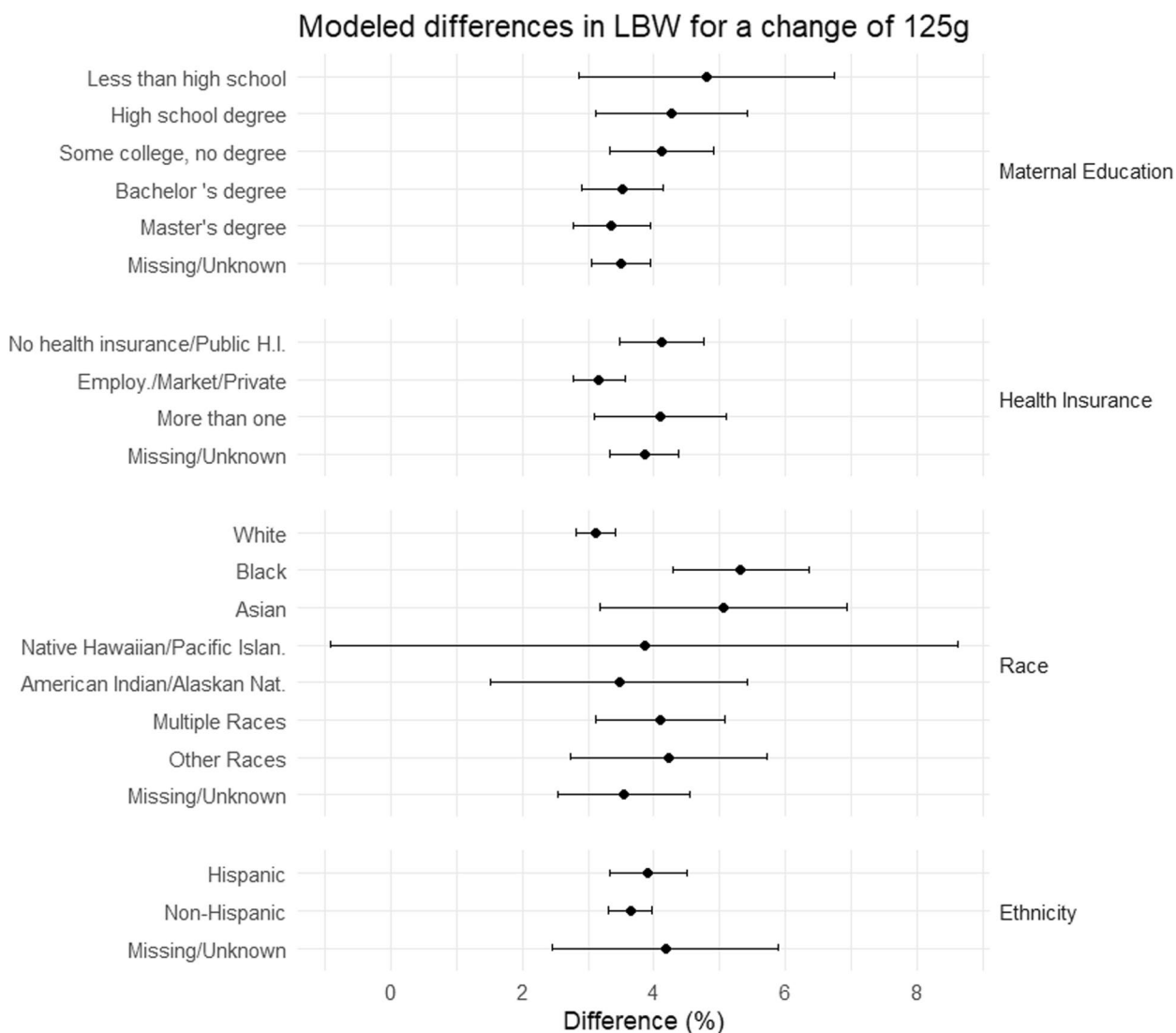
Peacock *et al. BMC Public Health*    (2024) 24:2655

Page 8 of 13

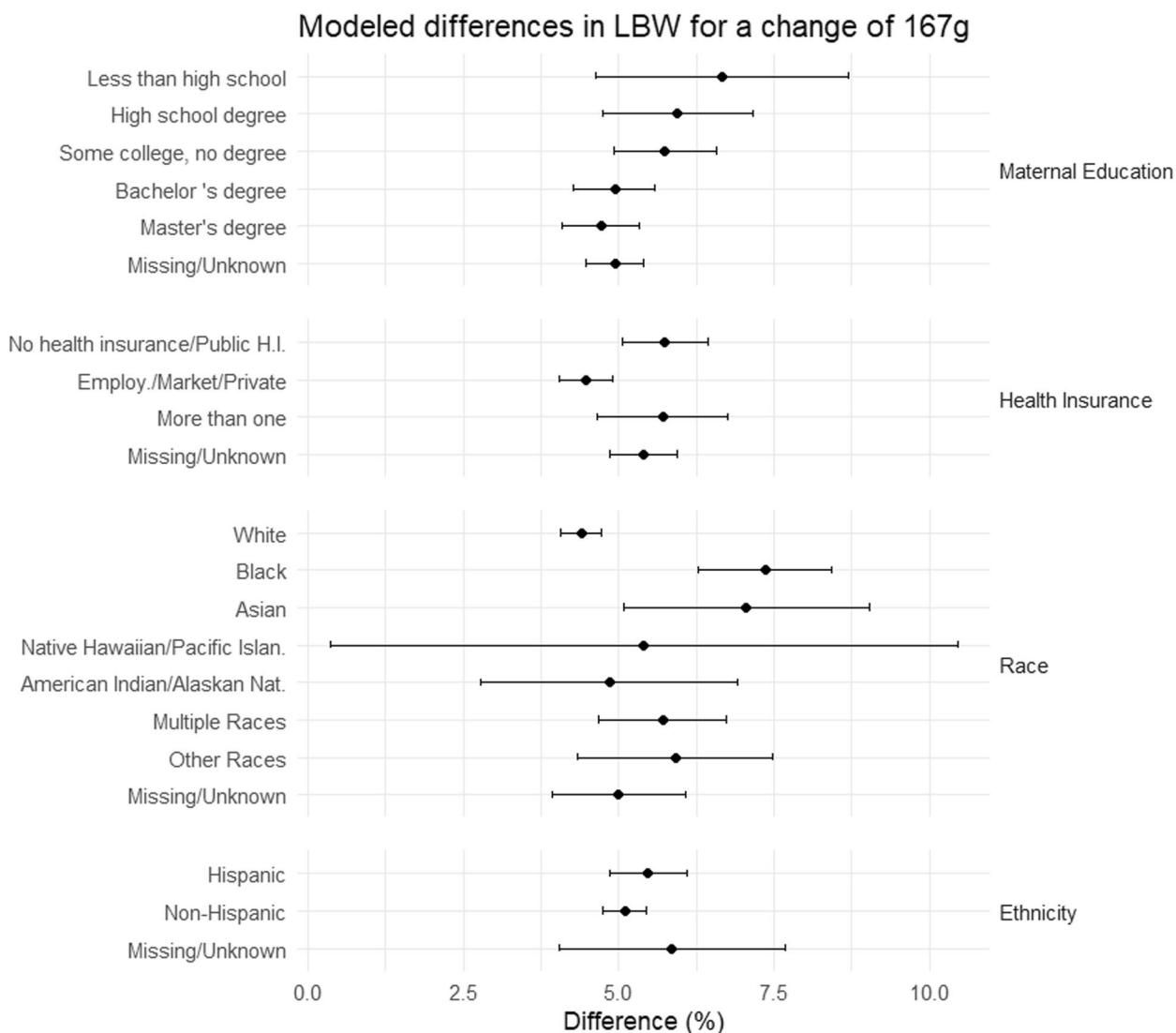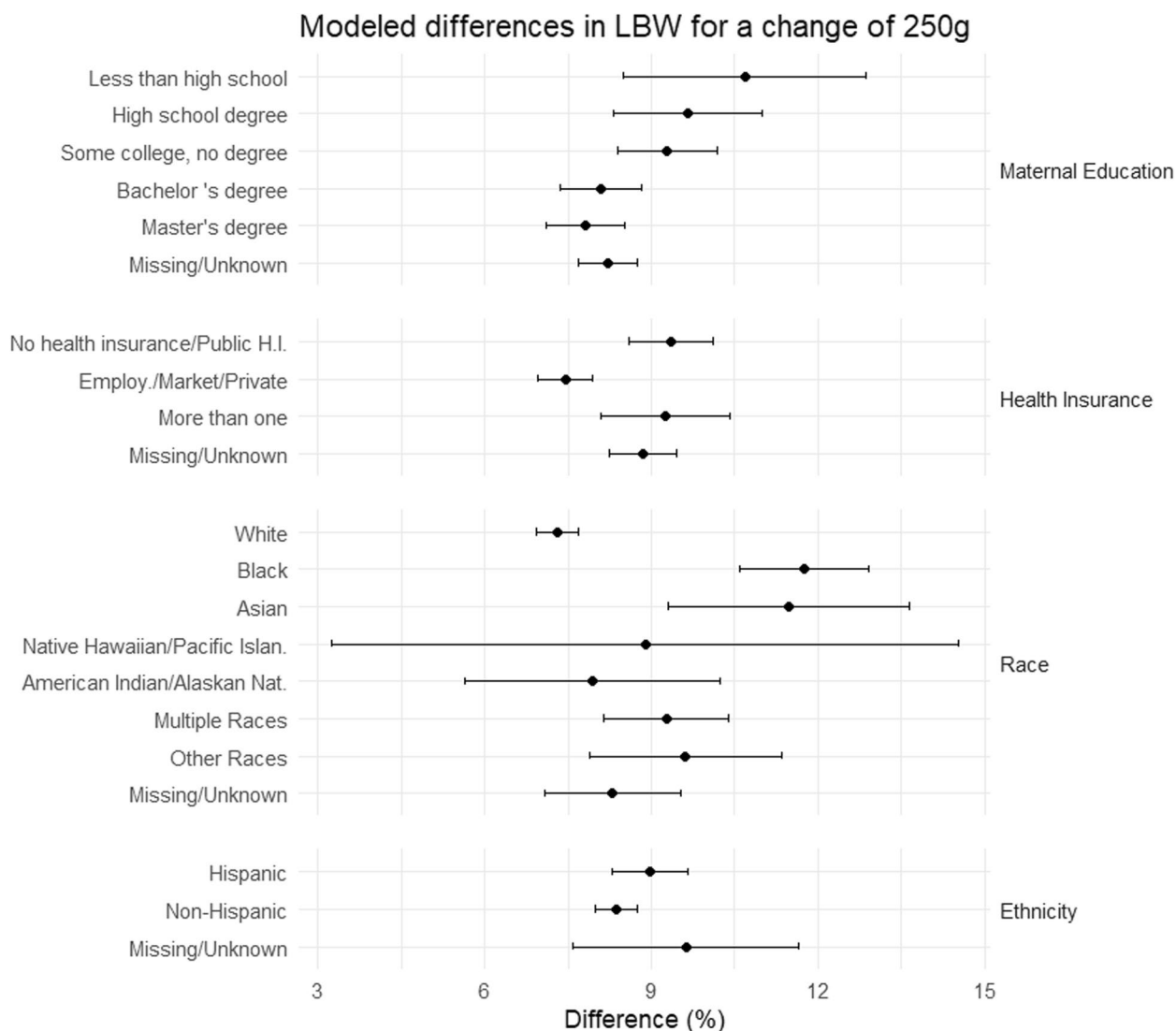## Modeled differences in LBW for a change of 125g



**Fig. 3** Modeled percentage of low birthweight (LBW) in unexposed and exposed populations by social factors associated with a change in mean birthweight of 125g ($N = 28,496$)

in general, quantile regression with its extensions using a Bayesian approach, offers a very flexible approach to modeling multiple quantiles [16].

In raising awareness of the importance of considering multiple effect measures when comparing effects between subgroups, it can be argued that dichotomization is not essential as there are other summary measures such as relative mean difference which are more natural choices. However,

it is our observation that dichotomization is widely used by clinicians because it is natural and meaningful to them. They use dichotomization to aid diagnosis and treatment decisions as well as in the interpretation of

population-level data as it provides a clinically meaningful outcome, as we have described in our previous publications [4, 7].

Analyzing combinations of factors was beyond the scope of this work but it seems very plausible that at least some vulnerable groups of mothers and their children will be subject to a combination of neighborhood stressors for example as assessed in ECHO using the instrument 'Combined social and environmental stressor exposure' measure. Hence the by-category estimates given for the four single social factors in our paper will be underestimates of the impact of the totality of adverse factors vulnerable families are exposed to [17].

**Fig. 4** Modeled percentage of low birthweight (LBW) in unexposed and exposed populations by social factors associated with a change in mean birthweight of 167g (*N* = 28,496)

We have focused on estimating proportions at high risk and on hypothetical exposures that increase the risk of poor outcome. We could have chosen to look at factors that lead to positive health [18] i.e. benefits where the same principles will apply, namely that a small change in the mean outcome in a positive direction will increase the proportion at normal/low risk by a greater amount in a vulnerable population with a lower baseline value. Therefore it follows that an intervention or preventive strategy that makes just a small difference will have a greater effect in the vulnerable than in the general population, and therefore may be worthwhile to consider.

This paper has directed attention to the choice of outcome in estimating the impact of an exposure but this is a simplification of the real world. We have not considered the prevalence or distribution of the hypothetical exposure which are likely to vary by socioeconomic factors. It seems very plausible that the prevalence of exposure may be greater in more deprived areas. In like manner, we have ignored the severity of the overall child outcome by considering birthweight alone. These simplifications seem to make it likely that the illustrative impacts may underestimate their true values in real life. Further, by using our

Peacock *et al. BMC Public Health*     (2024) 24:2655

Page 10 of 13



**Fig. 5** Modeled percentage of low birthweight (LBW) in unexposed and exposed populations by social factors associated with a change in mean birthweight of 250g (*N* = 28,496)

observed subgroup standard deviations for both the unexposed and exposed subgroups, we have assumed that the exposure affects only the mean and not the shape of each subgroup distribution. We chose to use a set of fixed absolute differences for birthweight in our subgroup calculations – we might have fixed on the number of SDs but considered that an absolute difference for birthweight was more intuitive.

The major strengths of this work are that we have analyzed pooled participant data across the ECHO Cycle 1 Program including up to 50 pan-US cohorts with up to 28,000 mother/child pairs and so the sample is not only very large but covers the diversity of

populations of mothers and children in the United States [10, 11]. Therefore the principles that have arisen here are relevant in research into factors affecting child health across the US and worldwide.

**Conclusions**

We have used data from the pan-US ECHO program to illustrate the importance of carefully considering the impact of an environmental exposure in vulnerable sub-populations of mothers and children using the dual outcome between exposed and unexposed individuals and families: difference in means and absolute risk difference based on the proportion at high risk. This
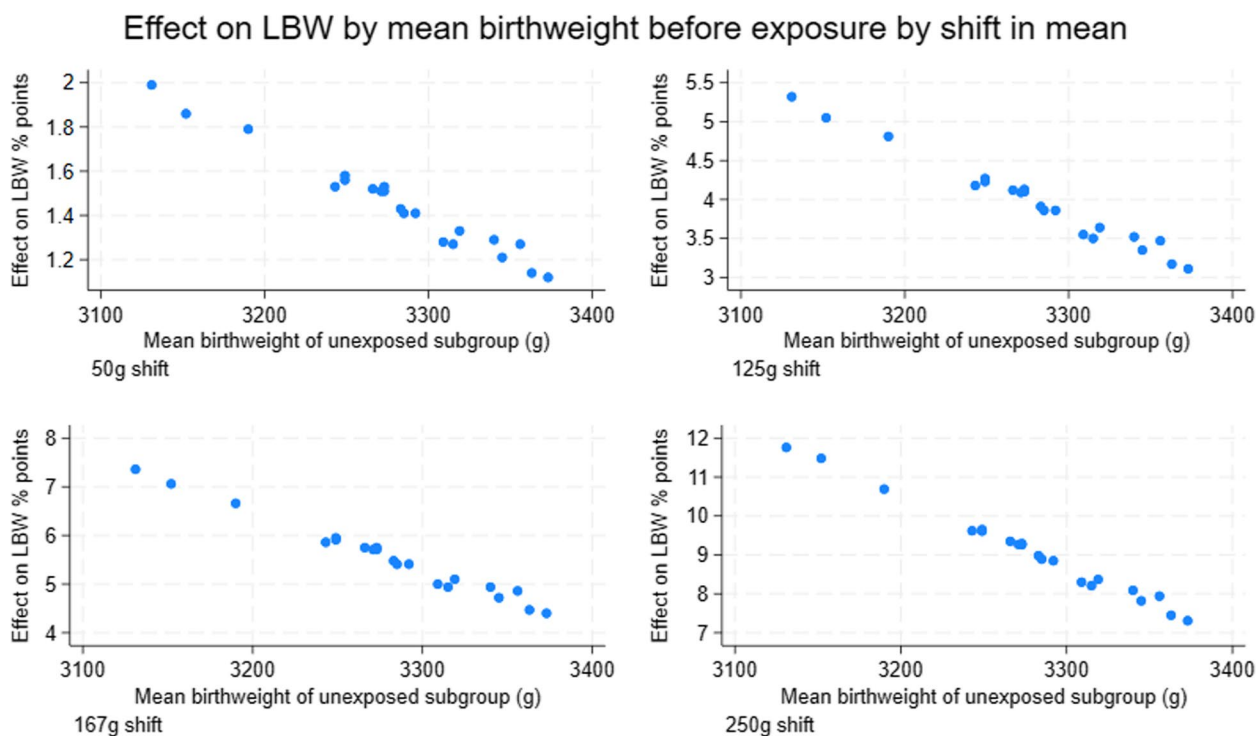
**Fig. 6** Summarizing the effects on low birthweight by the subgroup mean birthweight by the shift in the mean. 21 sub-populations

matters greatly because a small perturbation in mean outcome translates to a range of effect sizes for the proportions at high risk. Since vulnerable populations start with a higher proportion at high risk than the general population, a small perturbation will lead to a larger effect on the proportion at high risk than happens in the general population.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-024-20075-x.

Supplementary Material 1.

of Philadelphia - Center for Autism Research, Philadelphia, PA: Schultz R; University of North Carolina at Chapel Hill, Chapel Hill, NC: Piven J; University of Pittsburgh Medical Center, Magee Women's Hospital, Pittsburgh, PA: Simhan H; Magee Women's Hospital of University of Pittsburgh Medical Center, Pittsburgh, PA: Buss, C, Wadhwa, P; University of North Carolina, Chapel Hill, Huff, K; University of Rochester Medical Center, Rochester, NY: Miller RK; Harvard Pilgrim Health Care Institute, Boston, MA: Oken E; Michigan State University, East Lansing, MI: Kerver JM; o     Henry Ford Health, Detroit, MI: Barone C; Michigan Department of Health and Human Services, Lansing, MI: Fussman C; University of Michigan, Ann Arbor, MI: Elliott M; Wayne State University, Detroit, MI: Ruden D; Columbia University Medical Center, New York, NY: Herbstman J; University of Illinois, Beckman Institute, Urbana, IL: Schantz S; University of Utah, Salt Lake City, UT: Stanford J, Porucznik C, Giardino A; Icahn School of Medicine at Mount Sinai, New York, NY: Wright RJ; Boston Children's Hospital, Boston, MA: Bosquet-Enlow M; George Mason University, Fairfax, VA: Huddleston K; University of Minnesota, Minneapolis, MN: Nguyen R; Rutgers School of Public Health Environmental and Occupational Health Sciences Institute, Piscataway, NJ: Barrett E; Icahn School of Medicine at Mount Sinai, New York, NY: Swan S; Columbia University Medical Center, New York, NY: Perera F.

### Role of funder

### Authors' contributions

Conception/Design: JLP, SDC, OS, JRR. Data Acquisition: ETJ,RF,AD,NP,AP,TW,RMF,JT,DG,LEM,ZN,AG,LT,AF,LAC,SA,CB,AL,TGO,KL,HV,AA,JM,CAC,DD,CWH,CGB, IHP,RJS,AEH,KK,CK,KZL,BL,MC,JG,CM,MRE,SS,NJ,JMB,MRK. Data analysis: SDC. Data Interpretation: JLP, SDC, OS, AEH, ZN, AP, KZL, JMB, LAC, SS, MRE, CGB, MRK. Drafted the work: JLP, SDC. All authors approved the submitted version.

### Funding

### Availability of data and materials

R-script used for the present study are found in Open Science Framework: https://osf.io/sqzrx/?view_only = 1bd510766cad4489bc5a46471258868d. Select de-identified data from the ECHO Program are available through NICHD's Data and Specimen Hub (DASH). Information on study data not available on DASH, such as some Indigenous datasets, can be found on the ECHO study DASH webpage.
Select de-identified data from the ECHO Program are available through NICHD's Data and Specimen Hub (DASH). Information on study data not available on DASH, such as some Indigenous datasets, can be found on the ECHO study DASH webpage.

## Declarations

### Ethics approval and consent to participate

A properly constituted IRB of record, formally designated to review and monitor research involving human subjects, was accountable for compliance with regulatory requirements for the ECHO Cohort Data and Biospecimen Collection Protocol at participating Cohort Study Sites. The work of the ECHO Data Analysis Center is approved through the Johns Hopkins Bloomberg School of Public Health Institutional Review Board. The sIRB reviewed the protocol and all informed consent/assent forms, HIPAA Authorization forms, recruitment materials, and other relevant information prior to the initiation of any ECHO Cohort Data and Biospecimen Collection Protocol-related procedures or activities. The sIRB also reviewed any amendments to the protocol prior to their implementation. The WCG IRB tracking number for the ECHO-wide Protocol is 20181210.

### Consent for publication

This work does not contain any identifiable information images.

### Competing interests

The authors declare no competing interests.

### Author details

[1]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Dartmouth College, Hanover, NH 03756, USA. [2]Bielefeld School of Public Health and Department of Economy and Business Administration, Bielefeld University, Bielefeld, Germany. [3]Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC, USA. [4]Department of Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA, USA. [5]Emory University School of Medicine, Atlanta, GA, USA. [6]Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA. [7]Department of Reproductive Sciences, School of Medicine, University of California San Francisco, San Francisco, CA, USA. [8]Department of Environmental Health Sciences, University of California, Berkeley, CA, USA. [9]Children's Environmental Health Center, University of Illinois, Urbana, IL, USA. [10]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [11]Harvard TH Chan School of Public Health, Harvard University, Boston, MA, USA. [12]Department of Pediatrics, NYU Grossman School of Medicine, New York, NY, USA. [13]Kaiser Permanente Northern California Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [14]University of Rochester Medical Center, Rochester, NY, USA. [15]AJ Drexel Autism Institute, Drexel University, Philadelphia, PA, USA. [16]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [17]Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA. [18]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, USA. [19]Lifecourse Epidemiology of Adiposity and Diabetes (LEAD) Center, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [20]Avera Research Institute, Sioux Falls, SD, USA. [21]Department of Pediatrics, University of South Dakota School of Medicine, Sioux Falls, SD, USA. [22]National Farm Medicine Center, Marshfield Clinic Research Institute, Marshfield, WI, USA. [23]Department of Public Health Sciences, School of Medicine, The MIND Institute, University of California Davis, Davis, CA, USA. [24]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA. [25]University of Chicago, Chicago, IL, USA. [26]Department of Pediatrics, University of Washington, Seattle, WA, USA. [27]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, CA, USA. [28]Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Brown University, Providence, RI, USA. [29]George Washington University, Washington, DC, USA. [30]Department of Pediatrics, School of Medicine, Oregan Health and Science University, Portland, OR, USA. [31]University of Michigan School of Public Health and Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [32]Department of Pediatrics, Center for Child Health, Behavior, and Development, Seattle Children's Research Institute, Seattle, WA, USA. [33]Department of Environmental and Occupational Health Sciences, School of Public Health, University of Washington, Seattle, WA, USA. [34]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA. [35]Warren Alpert Medical School of Brown University, Brown University, Providence, RI, USA.

Peacock *et al. BMC Public Health*     (2024) 24:2655

Page 13 of 13

## References

1. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates; 1988.
2. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31(4):337–50. https://doi.org/10.1007/s10654-016-0149-3.
3. Organisation WH. Low birth weight. https://www.who.int/data/nutrition/nlis/info/low-birth-weight. Accessed 21 Sept 2024.
4. Peacock JL, Sauzet O, Ewings SM, Kerry SM. Dichotomising continuous data while retaining statistical power using a distributional approach. Stat Med. 2012;31(26):3089–103. https://doi.org/10.1002/sim.5354.
5. Sauzet O, Ofuya M, Peacock JL. Dichotomisation using a distributional approach when the outcome is skewed. BMC Med Res Methodol. 2015;15:40. https://doi.org/10.1186/s12874-015-0028-8.
6. Sauzet O, Peacock JL. Estimating dichotomised outcomes in two groups with unequal variances: a distributional approach. Stat Med. 2014;33(26):4547–59. https://doi.org/10.1002/sim.6255.
7. Sauzet O, Breckenkamp J, Borde T, et al. A distributional approach to obtain adjusted comparisons of proportions of a population at risk. Emerg Themes Epidemiol. 2016;13:8. https://doi.org/10.1186/s12982-016-0050-2.
8. Zivanovic S, Peacock J, Alcazar-Paris M, et al. Late outcomes of a randomized trial of high-frequency oscillation in neonates. N Engl J Med. 2014;370(12):1121–30. https://doi.org/10.1056/NEJMoa1309220.
9. Peacock JL, Lo J, Rees JR, Sauzet O. Minimal clinically important difference in means in vulnerable populations: challenges and solutions. BMJ Open. 2021;11(11): e052338. https://doi.org/10.1136/bmjopen-2021-052338.
10. Knapp EA, Kress AM, Parker CB, et al. The Environmental Influences on Child Health Outcomes (ECHO)-Wide Cohort. Am J Epidemiol. 2023;192(8):1249–63. https://doi.org/10.1093/aje/kwad071.
11. LeWinn KZ, Caretta E, Davis A, Anderson AL, Oken E, program collaborators for Environmental influences on Child Health O. SPR perspectives: Environmental influences on Child Health Outcomes (ECHO) Program: overcoming challenges to generate engaged, multidisciplinary science. Pediatr Res. 2022;92(5):1262–9. https://doi.org/10.1038/s41390-021-01598-0.
12. Duncan AF, Montoya-Williams D. Recommendations for reporting research about racial disparities in medical and scientific journals. JAMA Pediatr. 2024. https://doi.org/10.1001/jamapediatrics.2023.5718.
13. Rose G. Sick individuals and sick populations. Int J Epidemiol. 1985;14(1):32–8. https://doi.org/10.1093/ije/14.1.32.
14. Sauzet ORJ, Breiding JH. DistdichoR a R Package for the distributional dichotomisation of continuous outcomes. 2018. https://arxiv.org/abs/1809.03279. Accessed 11 Sept 2018.
15. Team RC. R: A Language and environment for statistical computing_. https://www.R-project.org/. Accessed 21 Sept 2024.
16. Yu KM, Moyeed RA. Bayesian quantile regression. Stat Probabil Lett. 2001;54(4):437–47. https://doi.org/10.1016/S0167-7152(01)00124-9. (In English).
17. Martenies SE, Zhang M, Corrigan AE, et al. Developing a National-Scale Exposure Index for Combined Environmental Hazards and Social Stressors and Applications to the Environmental Influences on Child Health Outcomes (ECHO) Cohort. Int J Environ Res Public Health 2023;20(14). https://doi.org/10.3390/ijerph20146339.
18. Forrest CB, Blackwell CK, Camargo CA Jr. Advancing the science of children's positive health in the national institutes of health Environmental Influences on Child Health Outcomes (ECHO) research program. J Pediatr. 2018;196:298–300. https://doi.org/10.1016/j.jpeds.2018.02.004.

## Publisher's Note