



Published in final edited form as:

*Matern Child Health J.* 2015 September ; 19(9): 1925–1935. doi:10.1007/s10995-015-1723-0.

## Maternal recall versus medical records of metabolic conditions from the prenatal period: A validation study

Paula Krakowiak<sup>a,b</sup>, Cheryl K. Walker<sup>b,c</sup>, Daniel J. Tancredi<sup>d</sup>, and Irva Hertz-Picciotto<sup>a,b</sup>

<sup>a</sup>Division of Epidemiology, Department of Public Health Sciences, School of Medicine, University of California, Davis

<sup>b</sup>MIND (Medical Investigations of Neurodevelopmental Disorders) Institute, University of California, Davis

<sup>c</sup>Department of Obstetrics and Gynecology, School of Medicine, University of California, Davis

<sup>d</sup>Department of Pediatrics, School of Medicine, University of California, Davis

### Abstract

**Objectives**—To assess validity of maternally-reported diabetes and hypertensive disorders, and reliability of BMI measurements during periconception and pregnancy compared with medical records when mothers are interviewed 2-5 years after delivery. To investigate whether reporting accuracy differed by child's case status (autism, delays, typical development).

**Methods**—Participants were mothers of 2-5 year old children with and without neurodevelopmental disorders from the CHARGE (*CH*ildhood Autism Risks from Genetics and the Environment) Study who had both prenatal/delivery records and telephone interviews. Sensitivity and specificity of self-report in telephone interview was assessed by comparison with medical records; agreement was evaluated by kappa statistics. Deviations in reported BMI were evaluated with Bland-Altman plots and concordance correlation coefficient (CCC).

**Results**—Mothers of children with neurodevelopmental disorders (autism or developmental delay) reported metabolic conditions slightly more accurately than control mothers. For diabetes, sensitivity ranged from 73% to 87% and specificity was 98% across groups. For hypertensive disorders, sensitivity ranged from 57% to 77% and specificity from 93% to 98%. Reliability of BMI was high (CCC=0.930); when grouped into BMI categories, a higher proportion of mothers of delayed children were correctly classified ( $\kappa_{wt}=0.93$ ) compared with the autism group and controls ( $\kappa_{wt}=0.85$  and  $\kappa_{wt}=0.84$ , respectively;  $P=0.05$ ). Multiparity was associated with higher discrepancies in BMI and misreporting of hypertensive disorders.

**Conclusions**—For purposes of etiologic studies, self-reported diabetes and hypertensive disorders during periconception and pregnancy show high validity among mothers irrespective of child's case status. Recall of pre-pregnancy BMI is reliable compared with self-reported values in medical records.

## Keywords

validation study; diabetes; preeclampsia; hypertension; body mass index; pregnancy; neurodevelopmental disorders

---

## Introduction

Metabolic conditions (MCs), including type 2 and gestational diabetes, hypertension, preeclampsia, and obesity, have been associated with labor and delivery complications (e.g., preterm labor, unscheduled cesarean delivery)<sup>1-3</sup> and adverse developmental outcomes in children.<sup>4,5</sup> Nearly 9% of women in California had type 2 or gestational diabetes during pregnancy;<sup>6</sup> 5% of U.S. pregnancies were complicated by chronic or gestational hypertension;<sup>7</sup> 34% of adult women under age 40 in the U.S. are obese and 16% have metabolic syndrome.<sup>8,9</sup> Approximately half of all pregnancies in the U.S. are unintended, leaving few opportunities for early intervention to address MCs.<sup>10</sup> Undoubtedly, MCs are a major health concern for pregnant women or those planning to become pregnant as well as for the developing fetus.

Epidemiologic studies of prenatal risk factors in relation to outcomes such as congenital malformations and neurodevelopmental disorders often rely on maternal recall of medical conditions during pregnancy. Medical records are not always complete or available on all participants. Moreover, in case-control study designs there is concern that medical conditions may be reported with differing degrees of accuracy by cases compared to controls, or with regard to demographic factors such as education level. Understanding the quality and consistency of self-reported risk factor information is critical to interpreting results of exposure-to-outcome relationships. Although previous validation studies from population-based cohorts of women have individually examined the quality of self-reported diabetes,<sup>11-15</sup> hypertension,<sup>11-17</sup> and body mass index (BMI)<sup>18-24</sup> in relation to more objective data sources (e.g., physician diagnoses, biological markers/measures, anthropometric measurements using calibrated instruments), few population-based studies have evaluated MCs during periconception and pregnancy.

We were interested in assessing the validity of maternally-reported diabetes, hypertensive disorders, and the reliability of BMI measurements during periconception and pregnancy compared with medical records when mothers were interviewed several years after delivery. This project was based on a population-based sample of mothers of 2 to 5 year-old children with and without a neurodevelopmental disorder. We also sought to investigate whether reporting accuracy differed in relation to the child's case status.

## Methods

This study followed the STARD (Standards for Reporting of Diagnostic Accuracy)<sup>25</sup> and GRRAS (Guidelines for Reporting Reliability and Agreement Studies)<sup>26</sup> guidelines for reporting validity and reliability studies.

## Participants

All participant mothers from the CHARGE (*CH*ildhood Autism Risks from Genetics and the Environment) Study<sup>27</sup> for whom we had both telephone interviews and prenatal/delivery records and whose child had a confirmed diagnosis of autism spectrum disorder (ASD), developmental delay (DD), or typical development (TD) were included in the validation study (Figure 1). This study was also limited to one child per mother. Participants were enrolled in the CHARGE Study between January 2003 and June 2010.

Briefly, CHARGE is an ongoing population-based case-control study comprised of children with ASD, DD, and controls from the general population. Eligible children are between the ages of 24 and 60 months, born in California, living with at least one biologic parent who speaks English or Spanish, and residing in the catchment areas of a specified list of Regional Centers in California. Child's case status was confirmed at the study clinic visit using gold standard instruments. Additional details regarding recruitment and child evaluation are published elsewhere.<sup>5</sup>

The CHARGE Study protocol was approved by institutional review boards of the University of California in Davis and Los Angeles and the State of California Committee for the Protection of Human Subjects. Written informed consent was obtained prior to participation.

## Maternal metabolic conditions

Self-reported diabetes (type 2 or gestational), hypertension, preeclampsia, and measurements to calculate body mass index (BMI; kg/m<sup>2</sup>) were obtained from the CHARGE Environmental Exposure Questionnaire (EEQ), a structured telephone-administered interview with the biological mother (completed by 97% of participants at the time of this study).

In the EEQ, mothers were asked, "During this [index] pregnancy were you ever told by a physician or nurse that you had [gestational diabetes; preeclampsia or toxemia]?" or "At any time before you became pregnant with [index child], were you ever told by a doctor that you had [diabetes; high blood pressure]?" (The question regarding diabetes was followed up with, "What type of diabetes did you have?") Questions about height and weight were asked as follows: "What is your height without shoes?" and "How much did you weigh before your pregnancy with [index child]?"

Pre-pregnancy BMI was calculated using height and pre-pregnancy weight reported in the EEQ and in medical records. Continuous, categorical, and dichotomous BMI variables were constructed. The categorical variable was defined as follows: underweight (BMI<18.5), normal weight (BMI 18.5-24.9), overweight (BMI 25.0-29.9), and obese (BMI 30.0). The dichotomous variable indicated whether or not the mother was obese (BMI 30 vs. BMI<30).

## Validity study

Data from medical records were regarded as the gold standard for analyses of diabetes and hypertensive disorders. Type 2 and gestational diabetes were combined in our analyses

because only 3 women had type 2 diabetes. Hypertensive disorders were examined in two ways: (1) chronic hypertension (onset before pregnancy or <20 weeks gestation), gestational hypertension, and/or preeclampsia (mild, severe, or HELLP syndrome), and (2) preeclampsia including gestational hypertension (hereafter referred to as preeclampsia). Chronic hypertension was not examined separately because only 17 mothers had this condition. After receiving authorization from mothers to obtain their prenatal and delivery records, we contacted healthcare providers to request these records. Providers were re-contacted if the medical charts were incomplete. At the time of this study, 69% of prenatal and 81% of delivery records had been obtained and abstracted; the major reasons for missing data in our study included receipt of partial records and unavailability of requested records (had been purged or in storage and practice non-response). Trained staff extracted information from medical records, under the supervision of an obstetrician, using standardized abstraction forms. All abstractors received at least two weeks of intensive training from a senior abstractor and/or supervising physician. All records included in this study were abstracted at least twice by different abstractors..

### Reliability study

Height and pre-pregnancy weight recorded in prenatal records are typically self-reported; therefore, we only measured the reliability of pre-pregnancy BMI derived from height and pre-pregnancy weight reported at the first prenatal visit and in the telephone interview conducted 3 to 6 years later. Heights from the EEQ and medical records were compared for consistency in reporting, and discrepancies  $\leq 2.54$  cm (1 inch) were considered to be consistent. For discrepancies  $>2.54$  cm, medical abstractors checked the values recorded in medical charts again to minimize transcription errors. Weights from both data sources were also compared for discrepancies. Differences  $>2.27$  kg (5 pounds) were verified.

### Covariates

Demographic and medical information was obtained using the EEQ, birth certificates, and medical records. Covariates selected *a priori* included the following: mother's age at delivery (<25, 25-29, 30-34, 35), race/ethnicity (Hispanic [any race] or other race [including American Indian, Black, Asian, Pacific Islander/Hawaii native, and multiple race], non-Hispanic White), education (no Bachelor degree, Bachelor degree or higher), language preference (English, Spanish or other), delivery payer (government program, private insurance), parity (primipara [no previous livebirths], multipara), pregnancy intendedness (planned, not planned), trimester prenatal care began (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, no care), and years elapsed between conception and the telephone interview.

### Statistical analyses

The validity of maternally-reported diabetes and hypertensive disorders was evaluated using sensitivity (Se) and specificity (Sp), and positive and negative predictive values (PPV, NPV). Non-parametric receiver operating characteristic (ROC) curves were generated to obtain the area under the curve (AUC) c-statistic to define validity globally as follows:  $<0.70$ =poor,  $0.70-0.79$ =fair,  $0.80-0.89$ =good, and  $0.90-1$ =excellent. Logistic regression models were fitted to evaluate differential misclassification (Se, Sp) and reclassification

(PPV, NPV) by case status.<sup>28</sup> Separate models were fitted to assess (1) underreporting of a MC among women diagnosed with that MC (false negative vs. true positive); (2) overreporting of a MC among women not diagnosed with that MC (false positive vs. true negative); (3) overreported MC among women reporting that MC (false positive vs. true positive); and (4) underreported MC among women reporting absence of that MC (false negative vs. true negative).

Reliability was assessed using Cohen's kappa ( $\kappa$ ) and weighted kappa ( $\kappa_{wt}$ ) statistics. Fleiss-Cohen weights<sup>29</sup> were used to compute the weighted kappa coefficient; BMI categories were scored in 1 unit increments, with score=1 for BMI <18.5 and score=4 for BMI  $\geq$  30.0. Strength of agreement was defined as follows: 0=poor, 0.01-0.20=slight, 0.21-0.40=fair, 0.41-0.60=moderate, 0.61-0.80=substantial, and 0.81-1=almost perfect.<sup>30</sup> To evaluate whether agreement differed by case status, kappa statistics were calculated within case groups and compared using a Chi-square test for equal kappa coefficients.<sup>31</sup> Continuous values of pre-pregnancy BMI from the EEQ and medical records were compared graphically using Bland-Altman plots.<sup>32,33</sup> Specifically, the differences between the EEQ and medical record measurements were plotted against the average values of these measurements. The two sets of measurements were considered to be in good agreement if 95% of the differences were within 2 standard deviations (SD) of the mean difference (i.e., the limits of agreement). The concordance correlation coefficient (CCC), derived from variance components, was also calculated for BMI.<sup>34,35</sup> To evaluate whether discrepancies in BMI values were associated with case status, linear regression models were fitted with outcome of (natural) log-transformed squared deviation in pre-pregnancy BMI values between the EEQ and medical records ( $\ln [\text{deviation}^2 + 0.0001]$ ).

All analyses were performed using SAS version 9.2 (SAS Institute Inc., Cary, NC). Bland-Altman plots were constructed using GraphPad Prism version 5.0 (GraphPad Software, San Diego, CA).

## Results

Demographic and pregnancy characteristics of the main study and subsets with complete data on MCs from both medical records and the EEQ are presented in Table 1. Briefly, slightly more than half of mothers were  $\geq$  30 years, White, without a Bachelor degree, and multiparous; two-thirds had planned their pregnancy; and most started prenatal care in the first trimester, had private medical insurance, and were English-speaking. The time elapsed between conception and the EEQ was <5 years for the vast majority of women, with an average of 4.4 years ( $\pm$  0.8). Mothers included in the validation substudies, compared to those who were excluded, were more likely to be White (64% vs. 48%), college educated (47% vs. 43%), and primiparous (46% vs. 42%) (Table 1, eTable 1); we observed similar patterns of differences in characteristics within case groups (data not shown).

### Validity study

Table 2 shows sensitivity, specificity and predictive values calculated for diabetes and hypertensive disorders. Overall, case mothers reported MCs more accurately than controls. The validity of maternally-reported diabetes ranged from good among controls (AUC=0.86)

to excellent among DD (AUC=0.93) and ASD (AUC=0.93) case mothers. Specifically, among those with diabetes, 73% of controls and 87% of case mothers (DD and ASD) reported having diabetes; specificity was high (98% to 100%) for all groups. The validity of hypertensive disorders was good among DD (AUC=0.83) and ASD (AUC=0.86) case mothers but fair for controls (AUC=0.78). However, these conditions were uncommon in the DD and control groups, reflected by the low precision of sensitivity and positive predictive values. Among those with hypertensive disorders, sensitivity ranged from 57% among controls to 77% among ASD case mothers; specificity was high (93% to 98%) for all groups. Validity indices (and their precision) were similar when we examined preeclampsia separately.

We used logistic regression models to determine whether child's case status or maternal characteristics were associated with correctly reporting the presence or absence of MCs. Child's case status and maternal characteristics (age, race/ethnicity, education, delivery payer, parity, pregnancy intendedness, time elapsed between conception and the EEQ) were individually included as predictors in the models; language preference and the trimester prenatal care began were excluded due to certain cells being small (cells <2). Neither the case status nor any maternal characteristics were associated with false positive or false negative reporting of diabetes in bivariate models (eTable 2). In contrast, multiparity was associated with misreporting of hypertensive disorders in bivariate analyses. Among mothers reporting hypertensive disorders, multiparae were 3.7 times more likely to overreport than primiparae (OR 3.71, 95% CI 1.15, 12.00) whereas underreporting was associated with primiparity (OR 3.03, 95% CI 1.14, 7.69) among mothers who did not report hypertensive disorders. Among women with no hypertensive disorders, the proportions of ASD and DD case mothers reporting hypertensive disorders were greater than for controls (4.7% [n=13] and 3.9% [n=3], respectively, vs. 1.0% [n=2]); however, these data were too sparse for conclusive interpretation (i.e., 95% confidence limit ratios >10). Neither the case status nor any maternal characteristics were associated with misreporting of preeclampsia. Among women with no preeclampsia, the proportion of those with no Bachelor degree was slightly higher than for those with a Bachelor degree (2.4% [n=9] vs. 0.3% [n=1]), but again these data were too sparse for meaningful conclusions.

### Reliability study

The agreement between medical records and the EEQ was in the almost perfect range for case mothers and substantial for controls (ASD:  $\kappa=0.82$ ; DD:  $\kappa=0.81$ ; TD:  $\kappa=0.70$ ; eTable 3); these kappa coefficients did not statistically differ according to child's case status. Using the categorical BMI variable (BMI 18.5 [underweight], 18.5-24.9 [normal weight], 25-29.9 [overweight], 30 [obese]), a higher proportion of mothers of DD children were correctly classified ( $\kappa_{wt}=0.93$ ) compared with the ASD case group ( $\kappa_{wt}=0.85$ ) and TD controls ( $\kappa_{wt}=0.84$ ;  $P=0.05$ ) (Table 3); nonetheless, agreement between telephone interviews and medical records was in the almost perfect range for all groups. Nearly all mothers with discordant BMI classifications were within one category of the classification in medical records (e.g., obese women were categorized as overweight based on EEQ data).

Logistic regression models analogous to those described earlier were carried out to examine whether case status or maternal characteristics were associated with BMI (derived from reported height and weight) in the EEQ relative to medical records. All maternal characteristics were examined except for language preference. Hispanic or other race and no Bachelor degree were both associated with underreporting of obesity (BMI  $\geq 30$ ) in the EEQ among women with BMI  $<30$  in medical records (other race/ethnicity vs. White: OR 2.83, 95% CI 1.28, 6.26; no Bachelor vs. Bachelor: OR 2.49, 95% CI 1.08, 5.70; eTable 4). The proportion of women who delayed prenatal care was also higher among women underreporting obesity in the EEQ (11.1% [n=3] vs. 2.5% [n=8]), but these data were sparse.

Discrepancies between BMI values derived from the EEQ and medical records were graphically assessed using a Bland-Altman plot (Figure 2). This plot revealed a mean difference of -0.87 and 95% limits of agreement ranging from -4.81 (BMI underestimation in medical records) to 3.07 (BMI overestimation), indicating good agreement, in clinical judgment, between BMI values from the EEQ and medical records. While, only 3.9% (16 of 415) women had discrepancies outside of these limits of agreement, the discrepancies appeared to become larger as the average of the EEQ and medical records BMI increased indicating that overweight and obese women misreported measurements to a greater extent in the EEQ than women in the normal BMI range; there was a trend toward overestimation of BMI in the EEQ among underweight women. The same patterns of reporting were observed in all groups (eFigures 1-3). The CCC for pre-pregnancy BMI was 0.930 (95% CI: 0.916, 0.942) indicating high reliability of BMI derived from the EEQ.

Using the log of squared difference in BMI between the EEQ and medical records as the outcome, we fitted bivariate and multivariable linear regression models to identify whether case status or maternal characteristics were associated with discrepancies in pre-pregnancy BMI (eTable 5, eTable 6). Case status was not associated with BMI discrepancies in a multivariable model adjusted for mother's race/ethnicity, education, parity and pregnancy intendedness; however, multiparity was associated with greater BMI discrepancies ( $\beta=0.716$ ,  $SE=0.330$ ,  $P=0.03$ ).

## Discussion

Collecting detailed, and therefore more accurate, data in large epidemiologic studies is expensive and, in the instance of case-control studies, sources of prospectively collected data are limited. Furthermore, there is concern that the accuracy of self-report might be impacted by participant case status or sociodemographic characteristics resulting in differential misclassification of an exposure or outcome. Our study of over 600 mothers of preschool-aged children with and without neurodevelopmental disorders showed that the validity of maternally-reported MCs was fair to excellent, with case mothers reporting MCs slightly more accurately than controls.

The majority of women correctly reported whether or not they had diabetes regardless of child's case status, and our validity indices fell within the ranges of 70% to 85% for sensitivity and 98% for specificity, previously reported in validation studies using subsets of participants from population-based cohorts.<sup>11-15</sup> All of these studies included middle-

aged or elderly women while our study was limited to women of reproductive age; also, many of these studies used blood glucose or glycosylated hemoglobin (HbA1c) to determine the presence of diabetes,<sup>11,13,14</sup> yet despite these demographic and methodological differences, our findings remained consistent with these studies. One case-control study involving mothers of children with and without leukemia found greater discrepancies in self-report when compared with medical records for cases (Se=70%, Sp=99%) and controls (Se=60%, Sp=100%) than our study.<sup>36</sup>

The validity of reporting hypertensive disorders was fair to good, with mothers of children with neurodevelopmental disorders reporting more accurately than controls. This pattern of reporting was also apparent when we examined preeclampsia individually; however, hypertensive disorders were especially uncommon among controls, and consequently, the validity indices lacked precision. In our study, the percentage of women who correctly reported having hypertensive disorders was on the upper end of a wide spectrum of sensitivities for hypertension reported in population-based studies, ranging from 23% to 95% (specificities ranged from 86% to 99%).<sup>11-17</sup> In contrast, a lower percentage of women correctly reported having preeclampsia (including gestational hypertension) compared with previous population-based cohorts (85% to 87%).<sup>37,38</sup> Higher sensitivity in those studies may have been partly attributable to shorter lag times between delivery and interview (2-6 months) compared with our study (2-5 years). With respect to case-control discrepancies in maternally-reported hypertensive disorders, similar patterns were previously observed in a case-control study of childhood leukemia (cases: Se=75%, Sp=94%; controls: Se=59%, Sp=94%).<sup>36</sup>

Not surprisingly, women recalled diabetes better than hypertensive disorders. Both type 2 and gestational diabetes are diagnosed by the end of the second trimester, and management during the remainder of pregnancy is intensive, requiring changes in diet and lifestyle, close monitoring of glucose levels, and possibly oral medication or insulin for glycemic control. While chronic hypertension is diagnosed prior to or early in pregnancy and is typically considered a lifelong condition treated with medication, preeclampsia is transient with onset after 20 weeks and most commonly immediately prior to delivery. Women who underreported diabetes or hypertensive disorders could have had milder forms of these conditions and may have been unaware of them. Furthermore, healthcare providers may have failed to adequately communicate these diagnoses to their patients, a scenario that could plausibly occur for preeclampsia with a late onset for example. Although we did not have adequate details on the severity of MCs, we explored whether child's case status and a variety of sociodemographic and pregnancy factors could have influenced our study participants' ability to correctly report these conditions. Neither case status nor maternal characteristics were associated with misreporting of diabetes. In contrast, parity was associated with misreporting of hypertensive disorders, indicating that adjustment for parity needs to be considered when using self-reported hypertensive disorders in analyses of similar populations. Approximately 6% of case mothers and 2% of controls overreported hypertensive disorders whereas fewer than 2% of women overreported preeclampsia. After further investigation, we discovered that some multiparae with histories of preeclampsia incorrectly reported having chronic hypertension possibly due to ambiguity in the wording of the interview question about high blood pressure.



Agreement of BMI, derived from maternally-reported height and pre-pregnancy weight, between medical records and the EEQ was assessed using categorical and continuous values, which allowed us to examine patterns of reporting and degrees of misclassification introduced through categorization. Women's reports of height and weight gave reliable estimates of BMI irrespective of case status. Overweight and obese women tended to underestimate their weight resulting in a lower estimate of BMI while underweight women tended to overestimate BMI. Our findings are consistent with population-based studies of adult women that compared self-reported BMI with measurements taken in a physical exam.<sup>18-24</sup> Multiparity was associated with greater absolute discrepancies in BMI between the telephone interview and medical records in our study. Parity had not been previously investigated; however, previous studies focused on middle-aged and elderly populations where parity may not have been a relevant factor in relation to biased reporting of BMI. In our BMI substudy, 70% of mothers planned their pregnancies and 97% started prenatal care in the first trimester. Neither of these factors influenced the reporting of BMI in the telephone interview. Likewise, years elapsed between conception and telephone interview was not associated with discrepancies in BMI reporting.

We relied on medical records as our gold standard. Although MCs are critical to recognize during pregnancy and should, therefore, be noted in medical records, we are aware that medical records may not always be complete. We took particular care and verified medical records multiple times when there were discrepancies with the EEQ. We examined glucose screening results, blood pressure measurements, and medications whenever those were available to confirm diagnoses of maternal conditions. Therefore, we have confidence that the errors in chart abstraction were minimal.

The validity of BMI reported in the EEQ could not be evaluated in our study because prenatal records do not contain measurements taken pre-pregnancy; instead, we assessed the reliability of self-reported height and pre-pregnancy weight at the first prenatal visit recorded in charts in comparison with height and weight reported in the EEQ 3 to 6 years after conception. In evaluating consistency of reporting over time, we assumed that weight and height data from the medical records would be close to the true values for several reasons. First, women's weight is recorded at every prenatal visit. Second, on average, most women only gain between 0.5 to 2 kg (1 to 10 lbs) during the entire first trimester.<sup>39</sup> Third, the classification of BMI derived from self-reported pre-pregnancy weight compared with BMI derived from weight measured at the first prenatal visit was in nearly perfect agreement in a cohort of women who initiated care in the first trimester<sup>40</sup>, which should be applicable to our cohort given their early initiation of prenatal care. Therefore, maternally-reported pre-pregnancy weight values from medical records are likely to be close to the true values. Yet we recognize that data from charts on pre-pregnancy weight and height are not free of information bias.

Medical records were not available for all participants. Although there were slight differences in maternal characteristics between the mothers included in validation substudies and those who were excluded, the primary reasons for missing charts included receipt of partial records and unavailability of requested records (e.g., in storage, practice non-response), which were not related to participant characteristics. Interviews were completed

by nearly all mothers, with only 3% missing EEQs at the time of our validation analyses. For these reasons, it appears to be unlikely that missingness of medical records or interview data would have substantially influenced our evaluation of the quality of maternally-reported MCs.

Concern about MCs has grown as their prevalence has increased in recent decades, including among women of childbearing age. Furthermore, MCs are associated with adverse outcomes for the mother and the fetus, and nearly half of pregnancies in the U.S. are unintended. Yet few validation studies using population-based samples have evaluated the accuracy of self-reported maternal MCs, including diabetes, hypertensive disorders, and obesity, in periconception and pregnancy. Factors such as cost and time will continue to be obstacles for collecting accurate data, and some studies by design will continue to rely largely on self-report when medical records are unavailable. Concerns over the accuracy of self-reported medical conditions have undermined the confidence and interest in the use of surveys; however, our study provides reassuring findings about the validity of maternally-reported MCs, particularly diabetes, showing fair to excellent validity for these conditions; we demonstrated consistency in reporting of pre-pregnancy BMI at two time points (first prenatal visit and telephone interview 3 to 6 years later); nevertheless, it should be noted that multiparity influenced the accuracy of reporting of some MCs.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

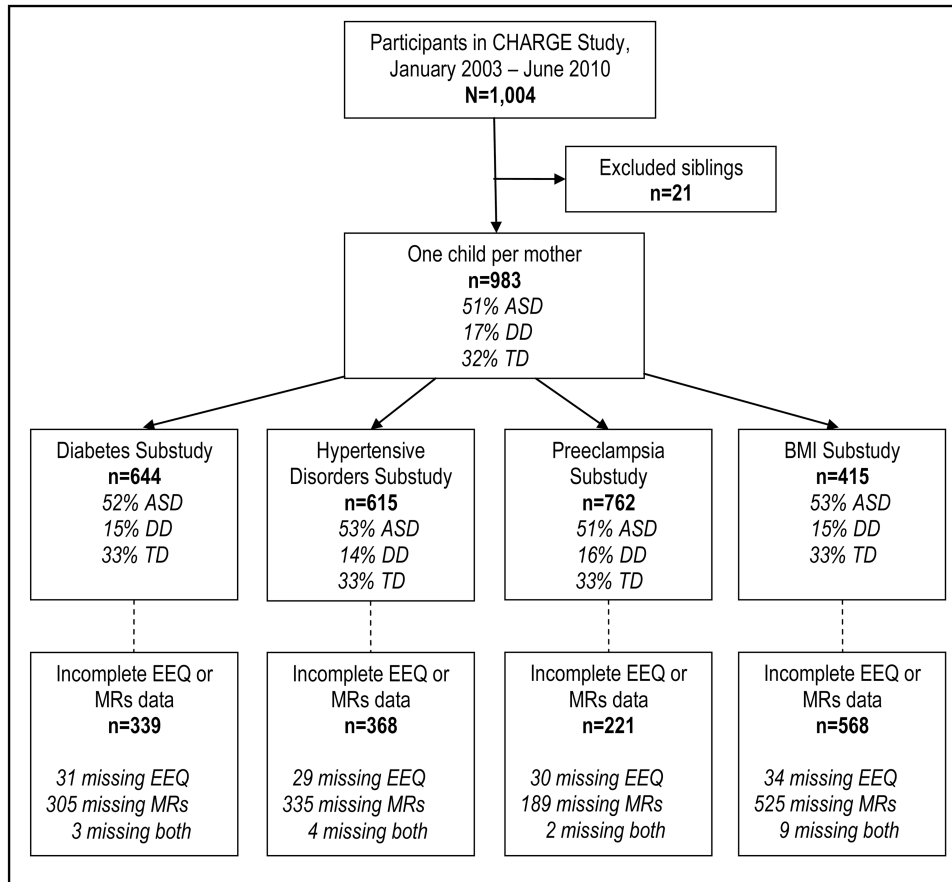
This research was supported by the National Institutes of Health (P01 ES11269 and R01 ES015359), the U.S. Environmental Protection Agency through the Science to Achieve Results (STAR) program (R829388 and R833292) and by the MIND Institute, University of California -Davis. None of the authors have any financial relationships or conflict of interest relevant to this article. The authors would like to thank the CHARGE Study participants and staff for their dedication and effort.

## References

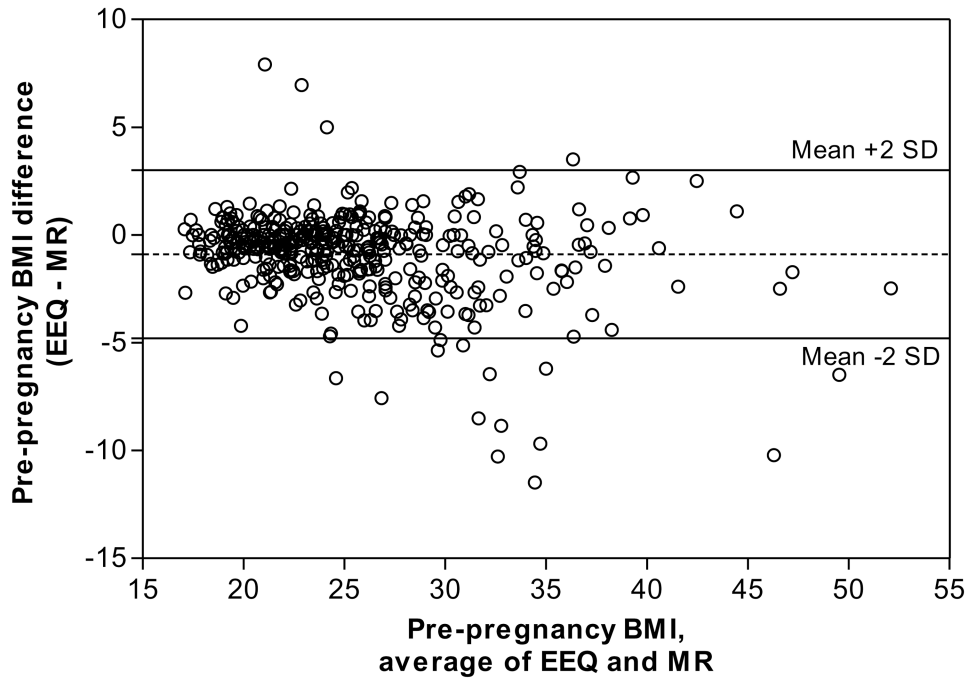
1. Rosenberg TJ, Garbers S, Lipkind H, Chiasson MA. Maternal obesity and diabetes as risk factors for adverse pregnancy outcomes: differences among 4 racial/ethnic groups. *American Journal of Public Health*. 2005; 95(9):1545–1551. [PubMed: 16118366]
2. Orbach H, Matok I, Gorodischer R, Sheiner E, Daniel S, Wiznitzer A, Koren G, Levy A. Hypertension and antihypertensive drugs in pregnancy and perinatal outcomes. *American Journal of Obstetrics and Gynecology*. 2013 Apr; 208(4):301.e1–6. [PubMed: 23159698]
3. Gaillard R, Durmu B, Hofman A, Mackenbach JP, Steegers EA, Jaddoe VW. Risk factors and outcomes of maternal obesity and excessive weight gain during pregnancy. *Obesity (Silver Spring)*. 2013 May; 21(5):1046–1055. [PubMed: 23784909]
4. Walker CK, Krakowiak P, Baker A, Hansen RL, Ozonoff S, Hertz-Picciotto I. Preeclampsia, placental insufficiency, and autism spectrum disorder or developmental delay. *JAMA Pediatrics*. 2014 Dec 8. Epub ahead of print.
5. Krakowiak P, Walker CK, Bremer AA, Baker AS, Ozonoff S, Hansen RL, Hertz-Picciotto I. Maternal metabolic conditions and risk for autism and other neurodevelopmental disorders. *Pediatrics*. 2012; 129(5):e1121–1128. [PubMed: 22492772]

6. Lawrence JM, Contreras R, Chen W, Sacks DA. Trends in the prevalence of preexisting diabetes and gestational diabetes mellitus among a racially/ethnically diverse population of pregnant women, 1999-2005. *Diabetes Care*. 2008; 31(5):899-904. [PubMed: 18223030]
7. Martin JA, Hamilton BE, Ventura SJ, Osterman MJ, Kirmeyer S, Mathews TJ, Wilson EC. Births: final data for 2009. *National Vital Statistics Reports*. 2011 Nov 3; 60(1):1-70. [PubMed: 22670489]
8. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999-2008. *Journal of the American Medical Association*. 2010; 303(3):235-241. [PubMed: 20071471]
9. Ervin RB. (2009). Prevalence of metabolic syndrome among adults 20 years of age and over, by sex, age, race and ethnicity, and body mass index: United States, 2003-2006. *National Health Statistics Reports*. 2009; (13):1-7. [PubMed: 19634296]
10. Finer LB, Zolna MR. Unintended pregnancy in the United States: incidence and disparities, 2006. *Contraception*. 2011 Nov; 84(5):478-485. [PubMed: 22018121]
11. Huerta JM, Tormo MJ, Egea-Caparrós JM, Ortolá-Devesa JB, Navarro C. Accuracy of self-reported diabetes, hypertension and hyperlipidemia in the adult Spanish population. DINO study findings. *Revista Española de Cardiología*. 2009; 62(2):143-152. [PubMed: 19232187]
12. Wada K, Yatsuya H, Ouyang P, Otsuka R, Mitsuhashi H, Takefuji S, Matsushita K, Sugiura K, Hotta Y, Toyoshima H, Tamakoshi K. Self-reported medical history was generally accurate among Japanese workplace population. *Journal of Clinical Epidemiology*. 2009; 62(3):306-313. [PubMed: 18774692]
13. Molenaar EA, Van Ameijden EJ, Grobbee DE, Numans ME. Comparison of routine care self-reported and biometrical data on hypertension and diabetes: results of the Utrecht Health Project. *European Journal of Public Health*. 2007; 17(2):199-205. [PubMed: 16887835]
14. Goldman N, Lin IF, Weinstein M, Lin YH. Evaluating the quality of self-reports of hypertension and diabetes. *Journal of Clinical Epidemiology*. 2003; 56(2):148-154. [PubMed: 12654409]
15. Martin LM, Leff M, Calonge N, Garrett C, Nelson DE. Validation of self-reported chronic conditions and health services in a managed care population. *American Journal of Preventive Medicine*. 2000; 18(3):215-218. [PubMed: 10722987]
16. Alonso A, Beunza JJ, Delgado-Rodríguez M, Martínez-González MA. (2005). Validation of self reported diagnosis of hypertension in a cohort of university graduates in Spain. *BioMed Central Public Health*. 2005; 5:94-100. [PubMed: 16156889]
17. Tormo MJ, Navarro C, Chirlaque MD, Barber X. Validation of self diagnosis of high blood pressure in a sample of the Spanish EPIC cohort: overall agreement and predictive values. EPIC Group of Spain. *Journal of Epidemiology and Community Health*. 2000; 54(3):221-226. [PubMed: 10746117]
18. Lin CJ, DeRoo LA, Jacobs SR, Sandler DP. Accuracy and reliability of self-reported weight and height in the Sister Study. *Public Health Nutrition*. 2012; 15(6):989-999. [PubMed: 22152926]
19. Lee DH, Shin A, Kim J, Yoo KY, Sung J. Validity of self-reported height and weight in a Korean population. *Journal of Epidemiology*. 2011; 21(1):30-36. [PubMed: 20953091]
20. Lucca A, Moura EC. (2010). Validity and reliability of self-reported weight, height and body mass index from telephone interviews. *Cadernos de Saúde Pública*. 2010; 26(1):110-122. [PubMed: 20209215]
21. Burton NW, Brown W, Dobson A. Accuracy of body mass index estimated from self-reported height and weight in mid-aged Australian women. *Australian and New Zealand Journal of Public Health*. 2010; 34(6):620-623. [PubMed: 21134066]
22. Craig BM, Adams AK. Accuracy of body mass index categories based on self-reported height and weight among women in the United States. *Maternal and Child Health Journal*. 2009; 13(4):489-496. [PubMed: 18607705]
23. Wada K, Tamakoshi K, Tsunekawa T, Otsuka R, Zhang H, Murata C, Nagasawa N, Matsushita K, Sugiura K, Yatsuya H, Toyoshima H. Validity of self-reported height and weight in a Japanese workplace population. *International Journal of Obesity*. 2005; 29(9):1093-1099. [PubMed: 15925952]
24. Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutrition*. 2002; 5(4):561-565. [PubMed: 12186665]

25. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *British Medical Journal*. 2003; 326(7379):41–44. [PubMed: 12511463]
26. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*. 2011; 64(1):96–106. [PubMed: 21130355]
27. Hertz-Picciotto I, Croen LA, Hansen R, Jones CR, van de Water J, Pessah IN. The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Environmental Health Perspectives*. 2006; 114(7):1119–1125. [PubMed: 16835068]
28. Pepe, MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, UK: Oxford University Press; 2003.
29. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*. 1973; 33(3):613–619.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [PubMed: 843571]
31. Fleiss, JL.; Levin, B.; Paik, MC. *Statistical methods for rates and proportions*. 3rd. Hoboken, NJ: John Wiley & Sons; 2003.
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1(8476):307–310. [PubMed: 2868172]
33. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound in Obstetrics and Gynecology*. 2003; 22(1):85–93. [PubMed: 12858311]
34. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45(1): 255–268. [PubMed: 2720055]
35. Carrasco JL, Phillips BR, Puig-Martinez J, King TS, Chinchilli VM. Estimation of the concordance correlation coefficient for repeated measures using SAS and R. *Computer Methods and Programs in Biomedicine*. 2013; 109(3):293–304. [PubMed: 23031487]
36. Olson JE, Shu XO, Ross JA, Pendergrass T, Robison LL. Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Cancer Group. *American Journal of Epidemiology*. 1997; 145(1):58–67. [PubMed: 8982023]
37. Klemmensen AK, Olsen SF, Osterdal ML, Tabor A. Validity of preeclampsia-related diagnoses recorded in a national hospital registry and in a postpartum interview of the women. *American Journal of Epidemiology*. 2007; 166(2):117–124. [PubMed: 17556761]
38. Coolman M, de Groot CJ, Jaddoe VW, Hofman A, Raat H, Steegers EA. Medical record validation of maternally reported history of preeclampsia. *Journal of Clinical Epidemiology*. 2010; 63(8): 932–937. [PubMed: 20189760]
39. Fattah C, Farah N, Barry SC, O'Connor N, Stuart B, Turner MJ. Maternal weight and body composition in the first trimester of pregnancy. *Acta Obstetrica et Gynecologica Scandinavica*. 2010; 89(7):952–955. [PubMed: 20380598]
40. Holland E, Moore Simas TA, Doyle Curiale DK, Liao X, Waring ME. Self-reported pre-pregnancy weight versus weight measured at first prenatal visit: effects on categorization of pre-pregnancy body mass index. *Maternal and Child Health Journal*. 2013; 17(10):1872–1878. [PubMed: 23247668]



**Figure 1.** Flowchart illustrating participants included in the validation substudies. EEQ refers to the Environmental Exposures Questionnaire, a structured telephone interview with the mother; MRs refers to prenatal and/or delivery medical records. Participant groups are ASD (autism spectrum disorder), DD (developmental delay), and TD (typical development). Diabetes Substudy includes type 2 and gestational forms of diabetes. Hypertensive Disorders Substudy includes chronic and gestational (e.g., preeclampsia) hypertension.



**Figure 2.** Bland-Altman plot comparing pre-pregnancy BMI from the EEQ and medical records (MRs). Pre-pregnancy BMI was calculated using height and pre-pregnancy weight reported in the EEQ or recorded in the medical records. The difference between pre-pregnancy BMI in the EEQ and BMI in the MR is plotted against the average for these two data sources with 95% limits of agreement. The dashed horizontal line denotes mean bias. The solid horizontal lines above and below the dashed line indicate the upper and lower limits of agreement. The mean bias detected was -0.87, and only 3.9% of the BMI values were outside of the 95% limits of agreement (-4.81 to 3.07).

**Table 1**  
**Demographic and reproductive characteristics of CHARGE Study participants and participants included in the validation study subsets**

	CHARGE Study		Diabetes <sup>a</sup> Substudy		Hypertensive Disorders <sup>b</sup> Substudy		Preeclampsia <sup>c</sup> Substudy		BMI Substudy	
	n	%	n	%	n	%	n	%	n	%
All	983	--	644	65.5	615	61.3	762	77.5	415	42.2
Age at delivery										
<25	161	16.4	95	14.8	89	14.5	122	16.0	63	15.2
25-29	247	25.1	160	24.8	153	24.9	190	24.9	103	24.8
30-34	325	33.1	223	34.6	210	34.2	252	33.1	144	34.7
35	250	25.4	166	25.8	163	26.5	198	26.0	105	25.3
Race/Ethnicity										
White, non-Hispanic	570	58.0	409	63.5	395	64.2	465	61.0	267	64.3
Other	413	42.0	235	36.5	220	35.8	297	39.0	148	35.7
Education										
No Bachelor	537	54.6	343	53.3	330	53.7	412	54.1	203	48.9
Bachelor or higher	446	45.4	301	46.7	285	46.3	350	45.9	212	51.1
Language preference										
English	922	93.8	619	96.1	593	96.4	726	95.3	403	97.1
Spanish or other	61	6.2	25	3.9	22	3.6	36	4.7	12	2.9
Delivery payer										
Government program	170	17.3	90	14.0	88	14.3	122	16.0	54	13.0
Private insurance	813	82.7	554	86.0	527	85.7	640	84.0	361	87.0
Parity										
Primipara	434	44.1	297	46.1	278	45.2	338	44.4	204	49.2
Multipara	549	55.9	347	53.9	337	54.8	424	55.6	211	50.8
Planned pregnancy										
Yes	627	65.2	431	67.1	407	66.4	501	65.9	291	70.1
No	334	34.8	211	32.9	206	33.6	259	34.1	124	29.9
Prenatal care started										
1 <sup>st</sup> trimester	920	93.6	618	96.0	592	96.3	726	95.3	401	96.6
2 <sup>nd</sup> trimester	63	6.4	26	4.0	23	3.7	36	4.7	14	3.4

	CHARGE Study		Diabetes <sup>a</sup> Substudy		Hypertensive Disorders <sup>b</sup> Substudy		Preeclampsia <sup>c</sup> Substudy		BMI Substudy	
	n	%	n	%	n	%	n	%	n	%
Years elapsed										
<3.5	152	15.5	113	17.6	113	18.4	130	17.1	65	15.7
3.5-4.4	342	34.8	234	36.3	231	37.6	273	35.8	155	37.4
4.5-5.4	384	39.0	234	36.3	266	35.1	283	37.1	154	37.1
5.5	105	10.7	63	9.8	55	8.9	76	10.0	41	9.9

<sup>a</sup>Diabetes, type 2 or gestational; three mothers had type 2 diabetes (1 ASD case mother, 1 DD case mother, and 1 TD control mother); the remaining 57 had gestational diabetes.

<sup>b</sup>Hypertensive disorders, chronic or pregnancy-induced and preeclampsia; seventeen mothers had chronic hypertension (12 ASD case mothers, 2 DD case mothers, and 3 TD control mothers); the remaining 54 had pregnancy-induced hypertension/preeclampsia.

<sup>c</sup>Preeclampsia, including pregnancy-induced hypertension.



**Table 2**  
**Comparison of metabolic conditions in the EEG and medical records<sup>a</sup>**

		Medical Record:										
EEQ:	Yes	No	Se	[95% CI]	Sp	[95% CI]	PPV	[95% CI]	NPV	[95% CI]	AUC	
<b>Diabetes<sup>b</sup>:</b>												
All	Yes	50	7	83.3	[71.5, 91.7]	98.8	[97.6, 99.5]	87.7	[76.3, 94.9]	98.3	[96.9, 99.2]	0.91
	No	10	577									
ASD	Yes	26	5	86.7	[69.3, 96.2]	98.4	[96.2, 99.5]	83.9	[66.3, 94.6]	98.7	[96.7, 99.6]	0.93
	No	4	300									
DD	Yes	13	0	86.7	[59.5, 98.3]	100	[95.4, 100]	100	[75.3, 100]	97.5	[91.4, 99.7]	0.93
	No	2	79									
TD	Yes	11	2	73.3	[44.9, 92.2]	99.0	[96.4, 99.9]	84.6	[54.6, 98.1]	98.0	[95.0, 99.5]	0.86
	No	4	198									
<b>Hypertensive Disorders<sup>c</sup>:</b>												
All	Yes	51	27	71.8	[59.9, 81.9]	95.0	[92.9, 96.7]	65.4	[53.8, 75.8]	96.3	[94.3, 97.7]	0.84
	No	20	517									
ASD	Yes	36	18	76.6	[62.0, 87.7]	93.5	[89.9, 96.1]	66.7	[52.5, 78.9]	95.9	[92.8, 97.9]	0.86
	No	11	257									
DD	Yes	7	5	70.0	[34.8, 93.3]	93.4	[85.3, 97.8]	58.3	[27.7, 84.8]	96.0	[88.6, 99.2]	0.83
	No	3	71									
TD	Yes	8	4	57.1	[28.9, 82.3]	97.9	[94.8, 99.4]	66.7	[34.9, 90.1]	96.9	[93.4, 98.9]	0.78
	No	6	189									
<b>Preeclampsia<sup>d</sup>:</b>												
All	Yes	38	10	64.4	[50.9, 76.5]	98.6	[97.4, 99.3]	79.2	[65.0, 89.5]	97.1	[95.5, 98.2]	0.82
	No	21	693									

Medical Record:											
EEQ:	Yes	No	Se	[95% CI]	Sp	[95% CI]	PPV	[95% CI]	NPV	[95% CI]	AUC
ASD	Yes	27	69.2	[52.4, 83.0]	98.0	[95.9, 99.2]	79.4	[62.1, 91.3]	96.6	[94.2, 98.2]	0.84
	No	12									
DD	Yes	6	66.7	[29.9, 92.5]	99.1	[95.2, 100]	85.7	[42.1, 99.6]	97.4	[92.6, 99.5]	0.83
	No	3									
TD	Yes	5	45.5	[16.8, 76.6]	99.2	[97.0, 99.9]	71.4	[29.0, 96.3]	97.6	[94.8, 99.1]	0.72
	No	6									

<sup>a</sup> Se=Specificity, Sp=Specificity, PPV=Positive Predictive Value, NPV=Negative Predictive Value, CI=Confidence Interval

<sup>b</sup> Diabetes, type 2 or gestational (N=644); three mothers had type 2 diabetes (1 ASD case mother, 1 DD case mother, and 1 TD control mother); the remaining 57 had gestational diabetes.

<sup>c</sup> Hypertensive disorders, chronic or pregnancy-induced and preeclampsia (N=615); seventeen mothers had chronic hypertension (12 ASD case mothers, 2 DD case mothers, and 3 TD control mothers); the remaining 54 had pregnancy-induced hypertension/preeclampsia.

<sup>d</sup> Preeclampsia (N=762); includes pregnancy-induced hypertension.

**Table 3**  
**Comparison of BMI categories in the EEQ and medical records, N=415<sup>a</sup>**

EEQ:	Medical Record:					Kappa <sup>b</sup>	[95% CI]
	<18.5	18.5-24.9	25-29.9	30			
All	<18.5	10	10	0	0	0.86	[0.83, 0.90]
	18.5-24.9	1	188	25	1		
	25-29.9	1	9	79	26		
	30	0	0	1	64		
ASD	<18.5	6	8	0	0	0.85	[0.80, 0.90]
	18.5-24.9	0	101	15	0		
	25-29.9	1	7	39	11		
	30	0	0	0	30		
DD	<18.5	2	1	0	0	0.93	[0.87, 0.98]
	18.5-24.9	0	19	1	0		
	25-29.9	0	0	17	4		
	30	0	0	1	16		
TD	<18.5	2	1	0	0	0.84	[0.77, 0.91]
	18.5-24.9	1	68	9	1		
	25-29.9	0	2	23	11		
	30	0	0	0	18		

<sup>a</sup> CI=Confidence Interval

<sup>b</sup> Weighted kappa statistic