Published in final edited form as:

Intellect Dev Disabil. 2009 October ; 47(5): 348–357. doi:10.1352/1934-9556-47.5.348.

Propensity Scores: A method for matching on multiple variables in Down syndrome research

Jennifer Blackford

Vanderbilt University Medical Center

Families of children with Down syndrome are different from both families of children with other or no disabilities (Most, D. E., Fidler, D. J., Laforce-Booth, C., & Kelly, J., 2006). Many of these differences exist before the child's birth, and these pre-existing differences may affect child outcomes. For example, mothers of children with Down syndrome are more likely to be older (Hook, E. B., 1981; Urbano, R. C. & Hodapp, R. M., 2007), with rates of Down syndrome significantly increased in mothers older than 35 years. Older mothers are at higher risk for having health related complications, such as hypertension, and adverse birth outcomes including preterm delivery and low birthweight (Cleary-Goldman, J. et al., 2005).

Pre-existing group differences can confound, or confuse, the interpretation of study findings. When groups differ on a variable also related to the outcome of interest, it is difficult to interpret a finding of significant group differences. For example, maternal age confounds the relationship between Down syndrome and birthweight (Figure 1). Children with Down syndrome have lower average birthweight compared to children without Down syndrome (Cronk, C. E., 1978), and newborns of older mothers typically weigh less than newborns of younger mothers (Cleary-Goldman, J. et al., 2005). In addition, older mothers are more likely to have a child with Down syndrome (Hook, E. B., 1981; Urbano, R. C. & Hodapp, R. M., 2007). Because both maternal age and Down syndrome may jointly contribute to lower birthweight, maternal age is a confounder in studies of the effect of Down syndrome on birthweight. If the confounder, maternal age, is not measured and accounted for, then the effect of Down syndrome on birthweight will be overestimated.

When uncontrolled, confounder variables can bias study results by underestimating or overestimating the true relationship between two variables. Biased estimates, in either direction can negatively affect research, practice, and policy. When effect sizes are underestimated, important factors related to developmental disabilities are not identified, and when effect sizes are overestimated, time, energy and monetary resources might be devoted to factors that are not actually related to developmental disabilities.

Confounders are a known problem in Down syndrome research (Hodapp, R. M., Glidden, L. M., & Kaiser, A. P., 2005; Seltzer, M. M., Abbeduto, L., Krauss, M. W., Greenberg, J., & Swe, A., 2004; Stoneman, Z., 2005) and methods, such as matching and covariate analysis, exist to control for confounders. Matching is a common strategy for selecting a group of comparison participants who are similar to the participants of interest on the potential confounders, especially when the group of interest is limited. Covariate analyses can adjust for the confounders statistically, but matching has several advantages (Rosenbaum, P. R. & Rubin, D. B., 1983): analysis of matched data is relatively easy; matching produces effect size estimates with smaller variance than covariate adjustment; analyses on matched data are more robust; and matching can control for more confounders than covariate adjustment, for a given sample size.

Covariate Matching

Covariate matching, the systematic selection of participants to balance groups (e.g., with or without Down syndrome) on select variables, is one method for controlling confounders. Covariate matching methods have two major practical limitations: the number of confounders is restricted by the sample size and participants must be available for both groups.

For covariate matching, the number of confounding variables must be relatively small. As the number of confounding variables increases, the sample size requirement increases exponentially. Each additional confounder adds a group to the study design. For example, in a study of Down syndrome with two groups (with and without Down syndrome) of 20 children each, adding advanced maternal age (yes/no) as a confounder adds an additional factor (2×2). With 20 children per group, the sample size is now *doubled* to 80 children ($2 \times 2 \times 20$). This is problematic because the number of potential confounders in Down syndrome research is likely to be large. In studies of children with Down syndrome and their families, potential confounders include both factors related to risk for Down syndrome (e.g., maternal age, maternal race) and factors that may be related to child or family outcome (e.g., marital status, birth order, child temperament, and sibling temperament [see Stoneman (2005)]). Adding a second confounder would increase the research design to three factors ($2 \times 2 \times 2$) and the sample size to 160. Thus, matching on more than two confounding variables quickly becomes prohibitive.

A related issue is potential participants must be available for both groups. As described above, as sample size requirements increase, it can be difficult to find sufficient numbers of study participants. This is especially true when the participant population, such as newborns with Down syndrome, is relatively small. Availability of potential participants can be further compounded by the distribution of the confounder across the two groups. For example, advanced maternal age is a good candidate for matching because maternal age is associated with both increased risk for Down syndrome and multiple outcomes, such as low birthweight. However, the proportion of mothers with advanced maternal age is different for newborns with and without Down syndrome. In Tennessee births from 2001, 23% of the children born with Down syndrome had mothers 40 years or older (Tennessee Birth Records, 2001), but for children without Down syndrome only 1.5% had mothers 40 years or older. Thus, the population of possible newborns without Down syndrome who also have older mothers is substantially reduced, making it difficult to find matches.

Propensity Matching

Propensity scores provide a method for matching on multiple confounding variables, without the limitations of covariate matching methods. Propensity scores methods were first introduced 25 years ago (Rosenbaum, P. R. & Rubin, D. B., 1983) and have been used in several different fields, such as cardiology (Normand, S. L. T. et al., 2001), psychology (Hill, J. L., Brooks-Gunn, J., & Waldfogel, J., 2003), and psychiatry (Sernyak, M. J., Desai, R., Stolar, M., & Rosenheck, R., 2001). However, to our knowledge, propensity scores have only been used once in developmental disabilities research (Witt, W. P., Kasper, J. D., & Riley, A. W., 2003) and have yet to be used in research on Down syndrome.

A propensity score represents a group of variables (confounders) with a single number (Rosenbaum, P. R. & Rubin, D. B., 1983). Specifically, the propensity score is a number between zero and one that represents the predicted probability that a person is in a particular group, given the confounders. For example, using a group of potential confounders such as maternal age, education, and weight gain, a propensity score represents the predicted likelihood that a specific newborn has Down syndrome. A propensity score of .90 means that the newborn has a .90 predicted probability of having Down syndrome, given her scores on the specific

group of confounding variables. Propensity scores are computed using logistic regression, a standard statistical method. In the logistic regression, the outcome variable is a binary variable representing group (e.g., with or without Down syndrome) and the predictor variables are the potential confounders. The logistic regression analysis produces an empirically derived formula (i.e., weighting of predictor variables) that best discriminates between the two groups. Applying this formula to each individual's values on all of the predictor values produces a predicted score. This predicted score is the individual's propensity score.

Propensity matching uses propensity scores to find matched participants from a comparison group. There are many methods for creating matched pairs, such as nearest available, Mahalanobis metric, and caliper matching (D'Agostino Jr., R. B., 1998; Rosenbaum, P.R. & Rubin, D.B., 1985). Successful matching requires both variability in the propensity scores within each group and overlap of propensity scores across groups (i.e., common support region). Once the matched groups are created, standard statistical analyses are used to test for group differences on the outcome variables. See Blackford (2007) for a detailed description of propensity score methods.

Propensity matching has multiple strengths. First, because propensity matching requires only the single propensity score to select a comparison group, the sample size limitations associated with covariate matching are avoided. Second, the independence of the propensity score from the outcome variable of interest makes it a good proxy for random assignment. Third, research shows that estimated propensity scores are better than the true propensity score at removing bias (Joffe, M. & Rosenbaum, P. R., 1999). Finally, propensity methods can be implemented using standard software packages.

Propensity matching also has several caveats and requirements. A major caveat is that propensity scores can only control for observed confounders (Joffe, M. & Rosenbaum, P. R., 1999). Unobserved confounders likely still exist and may influence study findings. Computation of the propensity score using logistic regression generally requires 10 observations per confounder. Thus, the number of confounders that can be controlled is proportional to sample size, limiting the use of this approach for small samples. Also, propensity scores can only be used for matching two groups at a time. Studies with three or more groups require separate propensity scores for each two group comparison.

To determine the utility of propensity scores in Down syndrome research, we compared propensity score matching to no matching and covariate matching in a statewide sample of newborn births. We tested the hypothesis that newborns with Down syndrome have lower birthweight than a comparison group of newborns without Down syndrome. Matching success rates, statistical significance, and effect size and were compared.

Method

Participants

Participants included newborns born in Tennessee during the year 2001. Administrative databases of births and hospital discharges were provided by the State of Tennessee. Birth records from year 2001 (Tennessee Birth Records, 2001) were linked to hospital discharge records from years 2001–2003 (Tennessee Hospital Discharge Data System, 2001–2003). The birth record included demographic variables, birth variables, and identification of Down syndrome. Although data collected for administrative purposes, such as birth records, may be contain errors (Iezzoni, L. I., 1997), descriptive demographic and birthweight data from Tennessee birth certificates have been reported to be reliable for both healthy births and those with adverse outcomes (Piper, J. M. et al., 1993). However, birth records in that study showed low sensitivity to identify congenital anomalies. Birth records may underreport diagnoses, such

as Down syndrome, for multiple reasons including delays in genetic testing. To increase the accuracy of Down syndrome diagnosis in this study, the birth records were linked to hospital discharge records to identify additional children with Down syndrome who were not identified at birth. See Urbano and Hodapp (2007) for details on data linkage and Down syndrome identification.

In 2001, 84,398 infants were born. Of these newborns, 85 (0.10%) had Down syndrome; 1,150 (1.4%) had a congenital anomaly; and 83,163 did not have Down syndrome or other congenital anomalies. For this study, we compared newborns with Down syndrome to newborns without Down syndrome or other congenital anomalies. For these two groups, complete data for all analytic variables was available for 90% of the sample. For the purpose of this simulation study, only complete cases were included in the sample. The final analytic sample included 77 newborns with Down syndrome and 75,155 typically-developing newborns without congenital anomalies.

Confounders

We selected possible confounders based on putative contributions to Down syndrome or birthweight. Potentially confounding variables included maternal age (Hook, E. B., 1981), maternal education (Chapman, D. A. et al., 2008; Urbano, R. C. & Hodapp, R. M., 2007), birth order (Seidman, D. S. et al., 1988), plural birth (Luke, B., 1994), maternal weight gain (Abrams, B. & Selvin, S., 1995), maternal race (Canfield, M. A. et al., 2006), marital status (Luo, Z.-C., Wilkins, R., & Kramer, M. S., 2004), and child gender (Seidman, D. S. et al., 1988). Maternal age, maternal education and maternal weight gain were continuous variables. Birth order (first/second/third+), plural birth (yes/no), maternal race (white/non-white), marital status (married/not married), and child gender (male/female) were categorical.

Matching

Covariate matches were created by selecting newborns from the Down syndrome group and a comparison group with the same values on each of the confounding variables. For propensity matching, propensity scores were first computed for each newborn using logistic regression. Matches were made by finding the closest propensity score match for each newborn with Down syndrome from the comparison group of newborns (Parsons, L. S., 2001). This process was performed without replacement so that once a match was made, the comparison newborn was no longer available for other matches. This method is known as nearest neighbor, nearest available, or greedy matching (Cochran, W. G. & Rubin, D. B., 1973; Rosenbaum, P. R. & Rubin, D. B., 1985).

Analyses

Following propensity score matching, *t*-tests were performed to test for balance between groups for each of the potentially confounding variables. Variables that remained unbalanced, based on a significant difference between the two groups, were included as covariates in subsequent analyses (Rosenbaum, P. R. & Rubin, D. B., 1985). Tests of group differences were unnecessary in the covariate matching sample because each match was exact.

The effect of matching method on newborn birthweight was tested for three subsamples: 1) a non-matched group (75,155 typically-developing newborns without congenital anomalies), 2) to a covariate-matched group, and 3) to a propensity matched group. For each of the matching methods, an analysis of variance (ANOVA) was performed to test for group (Down syndrome/comparison) differences in birthweight. Repeated measures ANOVAs were performed to test for group differences in birthweight for the covariate matched and propensity matched groups because matched pairs represent dependent samples. Down syndrome and comparison group differences are reported with both effect size (*d*) and statistical significance.

Effect sizes provide a standardized estimate of the magnitude of the relationship between Down syndrome and birthweight, regardless of sample size. For the two matched groups, we used a standard Cohen's d (1992) to measure standard differences in group means. However, for very large sample sizes, standard methods for computing effect size may be unsuitable because effect sizes become small and unrepresentative when the sample size gets very large. For the very large, non-matched group, we used a bootstrap procedure (Efron, B., 1979) to compute effect size for repeated samples ($N=77$, the same as the propensity matched group) from the full group. The bootstrap procedure provides a precise estimate of effect size comparable to the other matching methods.

To determine the utility of propensity scores in Down syndrome research, we used matching success rate, statistical significance, and effect size to compare each of the three matching methods. All statistical analyses were performed using SAS (version 9.1; Cary, NC) or R (version 2.5.1; The R Foundation for Statistical Computing). An alpha of .05 was used for all tests.

Results

Group Differences

Newborns with Down syndrome and their mothers differed from the comparison newborns on the eight confounders (Table 1). There were significant group differences in maternal age, maternal weight gain, birth order, plurality, marital status and newborn gender. Newborns with Down syndrome were more likely to be male, plural birth, and later born, and on average their mothers were older, gained less weight during pregnancy, and were more likely to be married.

Matching Success

Using covariate matching methods, only 48% of the newborns with Down syndrome had at least one match from the comparison group. Out of the 75,155 newborns in the comparison group, only 37 were exact matches to at least one infant in the Down syndrome group on all eight confounding variables. At that rate, at least 156,572 newborns would be needed in the comparison group to achieve 100% matching. Moreover, of the 37 newborns with matches, 41% only matched to one child in the comparison group. The high number of single matches indicates that the probability of finding an exact match on a moderate number of confounders is very low.

In contrast, with propensity score matching, 100% of the newborns with Down syndrome had a match from the comparison group. The matched groups were balanced on all confounding variables (all p s $> .20$) except maternal race ($p = .03$), where mothers of newborns with Down syndrome were more likely to be white.

It is also noteworthy that for covariate matching, differences emerged when comparing the "matched" versus "unmatched" newborns with Down syndrome. "Matched" infants—or those newborns for whom it was possible to find a match with the comparison group—differed from "unmatched" infants—those 40 for whom an appropriate match could not be found—on maternal age, $t(75) = -4.02$, $p = .0001$ and maternal weight gain, $t(75) = 3.19$, $p = .002$, with trends for plurality, $\chi^2(1) = 3.52$, $p = .06$ and birth order, $\chi^2(2) = 4.98$, $p = .08$ (see Table 2). Matching each newborn with Down syndrome, and deleting from the study infants for whom matches could not be found, produced a group that differed from the overall sample of newborns with Down syndrome on several confounding variables.

Statistical Significance and Effect Size

The presence of Down syndrome was significantly related to birthweight when the non-matched and covariate match methods were used. However, this effect was no longer significant ($p = .16$) when the propensity match method was used to control for confounders. This statistical significance difference was accompanied by differences in effect size. The effect size (d) of Down syndrome on birthweight based on the three matching methods differed (Table 3). The effect sizes for the non-matched ($d = .60$) and covariate matched ($d = 1.02$) groups were larger than the effect size for the propensity matched group ($d = .33$).

Discussion

This study introduced propensity scores as a method for controlling for multiple confounding variables in Down syndrome research. Most studies in developmental disabilities research have been able to control for a few confounders at most. When uncontrolled, confounders bias study results by underestimating or overestimating true effect sizes, which can negatively affect research, practice, and policy. When effect sizes are underestimated, we fail to identify important factors related to developmental disabilities, and when effect sizes are overestimated, we may devote time, energy and monetary resources to the wrong factors.

This study compared three different methods—non-matched, covariate matched, and propensity matched—to test the relationship between Down syndrome and birthweight. There were four major study findings: (1) pre-existing group differences, (2) low matching rates for covariate matching, (3) bias introduced by covariate matching and (4) overestimation of effect size and statistical significance by non-matching and covariate matching.

First, there were many pre-existing differences between the newborns with Down syndrome. Newborns with Down syndrome were more likely to be male, plural birth and later born. The mothers of newborns with Down syndrome were older and gained less weight. The pre-existing maternal and newborn differences reported here are similar to differences found by others (Hook, E. B., 1981; Urbano, R. C. & Hodapp, R. M., 2007). Prior studies have also reported significant differences in maternal race (Canfield, M. A. et al., 2006) and education (Urbano, R. C. & Hodapp, R. M., 2007). While there were mean differences in this sample, none reached statistical significance.

Second, when attempting to match on eight confounding variables, there were substantial differences in the number of successful matches for covariate matching compared to propensity matching methods. With covariate matching, only half of the newborns with Down syndrome could be matched to a newborn from the comparison group. In contrast, all of the Down syndrome group could be matched using propensity score matching. Covariate matching methods resulted in throwing out over half of the newborns in the Down syndrome group, reducing the overall study sample size and negatively affecting statistical power.

Third, covariate matching methods introduced a systematic bias into the study. The newborns with Down syndrome who were successfully matched were not representative of the entire group of newborns with Down syndrome. That is, the matching process resulted in a biased sample of newborns with Down syndrome. Compared to the newborns who were not matched, those with matches were born to younger mothers who gained more weight during pregnancy. Maternal age and maternal weight gain showed pre-existing group differences and were included as potential confounders. The bias resulted in an overestimation of the effect of Down syndrome on birthweight. A study comparing propensity to categorical matching (Rosenbaum, P.R. & Rubin, D.B., 1985) in children with prenatal barbiturate exposure reported similar low matching success and bias. Only 57% of the children with prenatal barbiturate exposure were successfully matched, and the unmatched children were significantly different from the

matched children. Thus, methods that leave some participants unmatched introduce serious bias due to systematic differences in the matched and unmatched groups.

Finally, non-matched and covariate matched methods overestimated the effect size of Down syndrome and birthweight, and produced erroneously significant findings. That is, the two methods produced estimates of effect size that were larger than the true effect. The effect size estimated from the propensity score method was .33. Compared to the propensity score method, the effect sizes from the non-matched and covariate match methods were 80% and 276% larger, respectively. For the non-matched group, the overestimation resulted from uncontrolled confounders. Covariate matching introduced systematic bias that inflated effect size. The overestimation of effect size was greatest for covariate matching, suggesting that, at least in this case, covariate matching was worse than no matching. The effect size differences were accompanied by differences in statistical significance. The group differences in birthweight found with the non-matched and covariate matching methods lost statistical significance once the confounders were controlled with propensity matching.

This study has several limitations. First, the sample includes only a single year of births in one Southern state. Infants born in Tennessee, and their mothers, may not be representative of the United States or other countries. These likely demographic differences might affect matching results. Second, administrative databases may contain errors. While we were able to improve accuracy in the identification of newborns with Down syndrome by record linkage with hospital data, we were not able to identify or correct other potential errors in the birth data. Also, we required exact matches for the covariate matching method. Less stringent methods, such as matching on categories or ranges of values may result in more successful matching, and therefore, less bias.

Nonetheless, this study demonstrates propensity matching can control for many confounders in Down syndrome research, overcoming the limitations of covariate matching methods. With propensity matching, 100% of the newborns with Down syndrome were successfully matched, the confounding variables were controlled, and a less biased estimate of effect size resulted.

Implications

The study findings have implications for practitioners, policy makers and researchers by illustrating the importance of controlling for confounders. Failure to identify and control for possible confounders will likely result in under or overestimates of the true relationship. For example, in this study, Down syndrome was associated with lower birthweight in newborns. However, after removing the effects due to various maternal and newborn factors, there was no significant association between Down syndrome and birthweight. This suggests that Down syndrome may not uniquely contribute to birthweight, but instead has an impact through another factor such as maternal age or weight gain. This finding could change both prenatal health care and future research in Down syndrome by focusing on factors related to prenatal or maternal weight gain. Practitioners, policy makers, and researchers should consider the role of confounders when evaluating the potential impact of research findings.

In addition, research findings in developmental disabilities research may be impacted by method selection when controlling for confounders. Covariate matching methods had multiple practical limitations but most importantly, introduced serious bias into the estimation of the effect of Down syndrome on newborn birthweight. This finding highlights the importance of selecting matching methods, such as propensity matching, that restrict the final matched sample and introduce bias.

Propensity matching was used in this study to identify a matched comparison group from a large, pre-existing, administrative dataset. However, propensity scores can also be used to

control for confounders in other study types. Confounder control may be especially important for developmental disability research because there are often multiple potential confounders and small samples. For example, in a two group study design, the propensity score could be included as covariate, providing statistical control for a group of potential confounders. Representation of multiple confounder variables with a single score provides confounder control without sacrificing degrees of freedom and reducing statistical power. However, propensity methods are not applicable to all study designs (e.g. prospective assignment to treatment groups) because propensity scores can only be computed on complete datasets. Future studies should test for propensity score method advantages in other types of study designs.

Acknowledgments

This work was supported by a National Institute of Child Health and Development program grant (P30 HD15052) to the Vanderbilt Kennedy Center for Research on Human Development. Birth (2001) and Hospital Discharge (2001–2003) were provided to the Vanderbilt Kennedy Center under contract NC-06-00849-00 with the Tennessee Department of Health, Office of Policy, Planning and Assessment.

I thank the following for assistance with this manuscript: Robert Hodapp for significant guidance in manuscript preparation, Trent Rosenbloom for review and comments, and Richard Urbano for data linkage.

References

- Abrams B, Selvin S. Maternal weight-gain pattern and birth-weight. *Obstetrics and Gynecology* 1995;86(2):163–169. [PubMed: 7617344]
- Blackford JU. Statistical issues in development epidemiology and developmental disabilities research: confounding variables, small sample size, and numerous outcome variables. *International Review of Research in Mental Retardation* 2007;33:93–120.
- Canfield MA, Honein MA, Yuskiv N, Xing J, Mai CT, Collins JS, et al. National estimates and race/ethnic-specific variation of selected birth defects in the United States, 1999–2001. *Birth Defects Research Part a-Clinical and Molecular Teratology* 2006;76(11):747–756.
- Chapman DA, Scott KG, Stanton-Chapman TL. Public health approach to the study of mental retardation. *American Journal on Mental Retardation* 2008;113(2):102–116. [PubMed: 18240872]
- Cleary-Goldman J, Malone FD, Vidaver M, Ball RH, Nyberg DA, Comstock CH, et al. Impact of maternal age on obstetric outcome. *Obstetrics and Gynecology* 2005;105(5):983–990. [PubMed: 15863534]
- Cochran WG, Rubin DB. Controlling bias in observational studies: A review. *Sankhya: The Indian Journal of Statistics (Series A)* 1973;35:417–446.
- Cohen J. A Power Primer. *Psychological Bulletin* 1992;112(1):155–159. [PubMed: 19565683]
- Cronk CE. Growth of children with down's syndrome: Birth to age 3 years. *Pediatrics* 1978;61(4):564–568. [PubMed: 149290]
- D'agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 1998;17(19):2265–2281. [PubMed: 9802183]
- Efron B. 1977 Rietz Lecture - Bootstrap methods - Another look at the jackknife. *Annals of Statistics* 1979;7(1):1–26.
- Hill JL, Brooks-Gunn J, Waldfogel J. Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology* 2003;39(4):730–744. [PubMed: 12859126]
- Hodapp RM, Glidden LM, Kaiser AP. Siblings of persons with disabilities: Toward a research agenda. *Mental Retardation* 2005;43(5):334–338. [PubMed: 16131230]
- Hook EB. Rates of chromosome-abnormalities at different maternal ages. *Obstetrics and Gynecology* 1981;58(3):282–285. [PubMed: 6455611]
- Iezzoni LI. Assessing quality using administrative data. *Annals of Internal Medicine* 1997;127(8):666–674. [PubMed: 9382378]

- Joffe M, Rosenbaum PR. Invited commentary: Propensity scores. *American Journal of Epidemiology* 1999;150(4):327–333. [PubMed: 10453808]
- Luke B. The changing pattern of multiple births in the united-states - Maternal and infant characteristics, 1973 and 1990. *Obstetrics and Gynecology* 1994;84(1):101–106. [PubMed: 8008301]
- Luo ZC, Wilkins R, Kramer MS. Disparities in pregnancy outcomes according to marital and cohabitation status. *Obstetrics and Gynecology* 2004;103(6):1300–1307. [PubMed: 15172868]
- Most DE, Fidler DJ, Laforce-Booth C, Kelly J. Stress trajectories in mothers of young children with Down syndrome. *Journal of Intellectual Disability Research* 2006;50:501–514. [PubMed: 16774635]
- Normand SLT, Landrum NB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology* 2001;54(4):387–398. [PubMed: 11297888]
- Parsons, LS. Reducing bias in a propensity score matched-pair sample using greedy matching techniques. Paper presented at the Proceedings of the 26th Annual SAS Users Group International Conference; 2001.
- Piper JM, Mitchel EF Jr, Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *Am J Epidemiol* 1993;137(7):758–768. [PubMed: 8484367]
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70(1):41–55.
- Rosenbaum PR, Rubin DB. The bias due to incomplete matching. *Biometrics* 1985;41(1):103–116. [PubMed: 4005368]
- Rosenbaum PR, Rubin DB. Constructing a control-group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 1985;39(1):33–38.
- Seidman DS, Everhadani P, Stevenson DK, Slater PE, Harlap S, Gale R. Birth-order and birth-weight reexamined. *Obstetrics and Gynecology* 1988;72(2):158–162. [PubMed: 3260664]
- Seltzer MM, Abbeduto L, Krauss MW, Greenberg J, Swe A. Comparison groups in autism family research: Down syndrome, fragile X syndrome, and schizophrenia. *Journal of Autism and Developmental Disorders* 2004;34(1):41–48. [PubMed: 15098956]
- Sernyak MJ, Desai R, Stolar M, Rosenheck R. Impact of Clozapine on Completed Suicide 2001;158:931–937.
- Stoneman Z. Siblings of children with disabilities: research themes. *Mental Retardation* 2005;43(5):339–350. [PubMed: 16131231]
- Tennessee Birth Records. [Electronic database]. City: Tennessee Department of Health, Office of Health Statistics; 2001.
- Tennessee Hospital Discharge Data System. [Electronic database]. City: Tennessee Department of Health, Office of Health Statistics; 2001–2003.
- Urbano RC, Hodapp RM. Divorce in families of children with Down syndrome: A population-based study. *American Journal on Mental Retardation* 2007;112(4):261–274. [PubMed: 17559293]
- Witt WP, Kasper JD, Riley AW. Mental health services use among school-aged children with disabilities: The role of sociodemographics, functional limitations, family burdens, and care coordination. *Health Services Research* 2003;38(6):1441–1466. [PubMed: 14727782]

Figure 1.
Maternal age as a confounder for the relationship between Down syndrome and birthweight.