



# HHS Public Access

Author manuscript

*Am J Epidemiol.* Author manuscript; available in PMC 2018 May 22.

Published in final edited form as:

*Am J Epidemiol.* 2018 March 01; 187(3): 592–603. doi:10.1093/aje/kwx262.

## Demographic and Operational Factors Predicting Study Completion in a Multisite Case-Control Study of Preschool Children

**Chyrise B. Bradley\***,

Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

**Erica N. Browne,**

Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

**Aimee A. Alexander,**

National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia

**Jack Collins,**

Autism Research Program, Kaiser Permanente Northern California Division of Research, Oakland, California

**Jamie L. Dahm,**

Wendy Klag Center for Autism and Developmental Disabilities, Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

**Carolyn G. DiGuseppi,**

Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado

**Susan E. Levy,**

Center for Autism Research, Children's Hospital of Philadelphia and Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

**Eric J. Moody,**

Department of Psychiatry, University of Colorado Anschutz Medical Campus, Aurora, Colorado

**Laura A. Schieve,**

National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia

**Gayle C. Windham,**

---

For permissions, [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

\*Correspondence to Chyrise B. Bradley, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, CB #8126, Chapel Hill, NC 27599-8126 ([chyrise@unc.edu](mailto:chyrise@unc.edu)).

Conflict of interest: none declared.

California Department of Public Health, Environmental Health Investigations Branch, Richmond, California

**Lisa Young**, and

School of Nursing, University of Pennsylvania, Philadelphia, Pennsylvania

**Julie L. Daniels**

Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

## Abstract

Participant attrition can limit inferences drawn from study results and inflate research costs. We examined factors associated with completion of the Study to Explore Early Development (2007–2011), a multiple-component, case-control study of risk factors for autism spectrum disorder in preschoolers, conducted in California, Colorado, Georgia, Maryland, North Carolina, and Pennsylvania. Participants ( $n = 3,769$ ) were asked to complete phone interviews, questionnaires, an in-person evaluation, and biologic sampling. We examined whether participant demographic and administrative factors predicted completion using mixed-effects logistic regression models. Completion of individual key study components was generally 70% or higher. However, 58% of families completed all per-protocol data elements (defined a priori as key study components). Per-protocol completion differed according to mother's age, race, educational level, driving distance to clinic, number of contact attempts to enroll, and number of telephone numbers provided (all  $P < 0.05$ ). Case status was not associated with completion, despite additional data collection for case-confirmation. Analysis of a subset that completed an early interview revealed no differences in completion by household factors of income, primary language spoken, number of adults, or number of children with chronic conditions. Differences in completion by race and education were notable and need to be carefully considered in developing future recruitment and completion strategies.

## Keywords

attrition; child development; completion; epidemiologic research design; recruitment of research subjects

---

For complex, multistep research protocols, retaining participants is challenging, yet study attrition can affect the validity of results as well as study costs. While many studies present attrition rates according to demographic characteristics, the impact of other intrinsic factors of study operations is not typically presented (1–4). Exploring operational factors that influence the completion of multiple-component, epidemiologic studies could inform future study designs, sampling plans, and logistical efforts in ways that might improve retention.

Most studies that have examined a range of potential influences on study completion were intervention/treatment trials or assessments of longitudinal studies involving several waves of data collection on a cohort. These studies have assessed associations between study attrition and participant health status, general life stress, health behaviors, intelligence quotient, ease in contacting participants, illness-related costs, residential mobility, travel

distance for treatment, incentive structure, race/ethnicity, age, income, education, and marital status (5–18). There is a dearth of contemporary information about factors associated with completion of shorter, especially multiple-component, studies involving families of children with special health-care needs.

The Study to Explore Early Development (SEED) Phase 1 was a multisite, observational case-control study of the causes and correlates of autism spectrum disorder (ASD) that used numerous recommended strategies for enhancing the participants' experience and personal engagement in the study process (19). SEED employed various strategies to promote study completion (Web Table 1, available at <https://academic.oup.com/aje>). The study required children, ages 2–5 years, and their parents or other primary caregivers to complete multiple data-collection components that were offered in sequence over approximately 4–12 months. The data-collection protocol included telephone interviews, self-administered questionnaires, and in-person evaluations with developmental assessments (20). We investigated associations between SEED study completion and participant demographic characteristics, administrative aspects of recruitment, and operational data-collection procedures that might be associated with completion of this multiple-component study. Our goal was to identify factors that could improve study efficiency and reduce attrition in similar epidemiologic studies in the future.

## METHODS

### Data source

Details of the recruitment and enrollment processes and data-collection components for SEED have been reported elsewhere (20). Briefly, the study included catchment areas within 6 states: California, Colorado, Georgia, Maryland, North Carolina, and Pennsylvania. The geographic size of each catchment area was determined by the minimum births per annum expected to allow achievement of recruitment goals, thus reflecting its population density. Recruitment and data collection for this phase of SEED occurred during 2007–2011. Eligible children were born between September 1, 2003, and August 31, 2006; were aged 24–68 months at enrollment; and had birth and current residence in one of the 6 catchment areas, a legal guardian available to consent, and a consistent caregiver from 6 months of age (or younger) who spoke English (all sites) or Spanish (California and Colorado only). Each site enrolled children representing 3 groups: children with ASD; children with other, non-ASD developmental delay (DD); and children from the general population (POP). Children who were potentially eligible for the study were identified from health-care and special education sources and from a random sample of birth records in each catchment area. For all groups, invitation materials were mailed to the child's home. Each study site employed various follow-up strategies for nonresponders, including follow-up phone calls and repeat mailings. The study was approved by the institutional review board at the Centers for Disease Control and Prevention as well as that of each participating site.

### Study procedures and protocol

During the enrollment phone call, the mother (98%) or other primary caregiver (2%) was screened for eligibility, asked to provide all available contact phone numbers, and asked for

verbal consent for the study. During this call, caregivers were also asked if the child had been formally diagnosed with ASD and were administered a brief ASD screening instrument, the Social Communication Questionnaire (21). Additionally, families were informed that they would receive feedback about the developmental assessments done as part of the study. This initial screening and the source through which the child was invited were used to determine which of 2 protocols the family would be asked to complete, the main protocol or the potential-ASD protocol (20).

Components of the main protocol were requested of all families and offered in sequence (Figure 1). Components included: a comprehensive telephone interview (Primary Caregiver Interview (PCI)) covering family sociodemographic characteristics, mother's reproductive history and pregnancy with the index child, and child's early development; 2 sets of questionnaires about the family's health and index child's development; consent for abstraction of mother's and child's medical records; an in-person evaluation of the child, including a developmental assessment, physical measurements, and a dysmorphology examination; biological sampling; and a diet and stool diary for the index child (offered only to those who attended an in-person evaluation). Those with a prior ASD diagnosis or Social Communication Questionnaire score  $\geq 11$  were considered at risk for ASD (22); they were thus assigned to the potential-ASD protocol and asked to complete additional study components. The potential-ASD protocol was designed to determine final case classification; it included 2 additional child development questionnaires and additional, in-person developmental assessments (1 with the child and 2 interviews for the caregiver). Collectively, all components of the main protocol took up to 6.5 hours to complete. The additional components of the potential-ASD protocol brought the time commitment to 11 hours.

Table 1 and Figure 1 detail the main and potential ASD data-collection protocols and sequence. Some flexibility was allowed in the sequence and mode of collection to enhance study completion (Web Table 1 lists SEED strategies for enhancing completion). Further, although components were generally offered to families in this order, subsequent components were offered without necessarily having completed previously offered components. The data-collection protocol required site staff to pursue completion of every offered component (20), but families were permitted to refuse any component. Efforts to collect each data component ceased after explicit participant request or 8 unsuccessful contact attempts or after 4 cancellations or 2 "no shows" of a scheduled telephone or in-person appointment.

### Definition of study completion

While most of SEED's primary research questions can be addressed without completion of every study component, all sites worked to achieve full completion with all enrolled families. Per-protocol completion was defined to assist sites in monitoring progress (Table 1) and ensured primary research questions could be addressed sufficiently. Per-protocol completion included completion of the PCI, a specific subset of the questionnaires, consent for review of at least 1 medical record, the assigned developmental assessment(s), the dysmorphology examination, and collection of biological samples (either buccal cells or

blood) from the enrolled child and biological mother. While per-protocol completion included most study components, certain components were considered supplemental— notably, paternal instruments and specimens (often not possible to obtain) and child diet and stool diaries (offered only to those seen in person due to the complexity of completion instructions).

### Demographic and administrative factors

We examined several factors for which data were collected at enrollment and during the PCI for their association with study completion. Because the biological mother was the primary caregiver in 98% of SEED families, demographic factors included in this analysis focused on maternal characteristics. Demographic factors obtained from birth certificate data included mother's race, Hispanic ethnicity, and education. Mother's age at study enrollment was calculated from her age at the child's birth and categorized for analysis. Hispanic ethnicity data were missing for 458 mothers (12%) mainly due to a change in recording method for vital records data at one site. Data on Hispanic ethnicity were acquired from the PCI for 368 of the missing 458. The PCI also provided the data on total household income, primary language spoken at home, number of caregiving adults in the household, and number of other children with any chronic health or developmental conditions born to the enrolled child's mother.

We examined administrative factors documented in the web-based study tracking system, including the number of phone contact attempts needed to enroll the family, number of phone numbers provided for the primary contact, and assignment to protocol group (main (DD or POP) or potential ASD). We also examined driving distance to the assessment clinic. ArcGIS, version 9.3.1 (ESRI, Redlands, California), was used to geocode residence and clinic addresses for each site. Driving distances to clinics were calculated in kilometers. Three sites had more than 1 clinic location. Because sites did not track which clinic each family attended, the distance to the nearest clinic was used. One site offered all families the option to conduct the in-person evaluation and blood sample collection in participants' homes or at a study clinic. Nearly all their families selected the home option. Thus the distance to clinic for families enrolled at that site was set to 0 km. Distance to clinic was categorized into quartiles.

### Statistical analyses

Multivariable mixed-effects logistic regression models were used to examine the association between administrative and demographic factors and per-protocol completion. A random effect for site was included in the model to adjust for correlation of demographic and administrative factors within each site. The primary model included all enrolled participants and factors available at enrollment. Because the PCI was requested early in the protocol and completed by a majority of enrollees, and because additional information on potential predictors was available from the PCI, a second model was used to assess additional factors associated with per-protocol completion among those who completed at least the PCI. In addition, models were strati-fied by mother's race, education, and protocol group assignment (potential ASD, main DD, or main POP) to evaluate how administrative factors, adjusted for other covariates, differed between those subgroups. Two-sided *P* values of <0.05 were

considered statistically significant. Analyses were conducted with Stata, version 12.0 (StataCorp LP, College Station, Texas).

## RESULTS

### Characteristics of enrollees

SEED enrolled 3,769 families, distributed across the 6 enrollment sites. Table 2 summarizes demographic and administrative characteristics of enrolled families by site. Overall, 71% of mothers were white and 12% were Hispanic. Forty-four percent of mothers had some college or a Bachelor's degree, and 28% had an advanced degree. The median number of call attempts needed to enroll a family was 3 (interquartile range, 1–4). Thirty-four percent of families required >3 calls, with the maximum needing 25 calls to enroll. More families in North Carolina and Pennsylvania lived >39 km from a study clinic (38% and 32%, respectively, versus 11% in California and Colorado). Table 3 summarizes additional characteristics available for those who completed the PCI. Among those families, 21% had annual household income under \$30,000, 90% spoke English at home, 78% had 2 adults in the household, and 19% had at least 1 child (not enrolled in the study) with a chronic health or developmental condition.

### Associations with study completion

All enrolled families completed the Social Communication Questionnaire, 79% completed the PCI, 74% completed Questionnaire Packet 1, 68% completed Questionnaire Packet 2, 70% provided consent for medical record abstraction, 72% completed the in-person evaluation, and 76% provided biologic samples for mother and child. Altogether, 2,204 (58%) achieved per-protocol completion (Table 4). Of the 1,565 families who did not complete per protocol, 785 (50%) completed the PCI, 518 (33%) completed the in-person evaluation, and 475 (30%) completed both the PCI and in-person evaluation.

Odds of per-protocol completion are presented in Table 5. In the primary model (including all enrollees), the odds of completion were significantly lower when the biological mother was black (odds ratio (OR) = 0.61, 95% confidence interval (CI): 0.49, 0.75) and significantly higher when the biological mother had some college or an undergraduate degree (OR = 1.52, 95% CI: 1.25, 1.83) or advanced degree (OR = 1.97, 95% CI: 1.57, 2.47). Odds of completion were lowest among the youngest mothers (ages 18.0–24.9 years: OR = 0.61, 95% CI: 0.41, 0.91). Several administrative factors were also associated with per-protocol completion. A caregiver needing >3 call attempts to enroll was much less likely to complete the study compared with those who required only 1 call (OR = 0.68, 95% CI: 0.57, 0.83). Provision of multiple contact phone numbers was associated with higher per-protocol completion. Distance of >39 km to nearest clinic was associated with lower per-protocol completion. Additionally, although a mixed effect model was used to adjust for site clustering, because 1 site had a noticeably higher proportion of black or African-American participants, we tested the primary model with that site excluded and found the same result: significantly lower odds of completion when the biologic mother was black (OR = 0.57, 95% CI: 0.44, 0.73).

The odds of completing the PCI were significantly higher if the biological mother had more than a high school education and if the family provided more than 1 contact number or required <3 call attempts to enroll (Web Tables 2 and 3). However, the demographic and administrative factors influencing per-protocol completion for the subgroup that completed at least the PCI ( $n = 2,989$ ) were generally similar to the primary model (Table 5). One notable difference in this subgroup was that POP protocol families were more likely to achieve per-protocol completion than potential-ASD protocol families (OR = 1.27, 95% CI: 1.01, 1.58). No additional factors collected in the PCI were associated with per-protocol completion.

In models stratified by race, odds of per-protocol completion were associated with education and number of contact phone numbers, both for families where the biological mother was black and those where the mother was white. When we stratified by protocol assignment group, we found that race, education, and number of phone numbers were associated with per-protocol completion for all groups. However, older maternal age was associated with completion only among families in the DD and POP protocols, and distance to nearest study clinic was associated with a lower likelihood of completion only among families in the main POP protocol (Web Table 4).

Odds of completing some individual study components are presented in Web Tables 2 and 3. Families where the mother had at most a high school education, who required >3 calls to enroll, and who provided only 1 contact phone number were significantly less likely to complete each study component (all  $P < 0.001$ ). Those living  $\geq 9$  km to the nearest clinic were less likely to complete the in-person evaluation (OR = 0.73, 95% CI: 0.56, 0.94). Families in the main DD (OR = 0.78, 95% CI: 0.65, 0.93) and POP (OR = 0.83, 95% CI: 0.69, 0.99) protocol groups were less likely to provide blood samples. Odds of completing each study component were lower when the biological mother was black or African American, and were significantly lower for providing a medical record release (OR = 0.68, 95% CI: 0.55, 0.86) and for providing a blood sample from mother or child (OR = 0.72, 95% CI: 0.59, 0.89).

## DISCUSSION

SEED employed numerous recommended strategies for enhancing study completion (19) and achieved completion near 70% or higher on individual components of the multiple-component SEED protocol (Table 4). While 76% of families met the biological-sample requirement for both mother and child, the proportion of participants providing blood (52%) was much lower than the proportion providing buccal samples (72%). Although there was, generally, high completion of individual components of the study, only 58% of enrolled families met the definition of per-protocol completion. Completion varied by study component but was highest for the PCI, followed closely by provision of a biological sample (primarily buccal cells) and completion of the first set of questionnaires. Notably, the PCI came early in the protocol sequence yet—similar to per-protocol completion factors—higher education, fewer calls to enroll, and provision of more contact phone numbers were associated with completion, although lower maternal age and distance to nearest clinic trended toward association with completion (Web Tables 2 and 3). In general, higher

maternal education, white maternal race, provision of multiple family contact phone numbers, and fewer calls required for enrollment were associated with greater odds of per-protocol completion.

SEED attempted to circumvent potential problems related to lower education, such as reading difficulty (by providing the option to complete questionnaires via telephone) or inflexible employment situations (by providing flexible appointment times). However, consistent with previous reports (12, 15, 23–27), we found that those who were younger or less educated were less likely to provide complete data. Thus, controlling for logistical factors did not explain age or education associations with completion.

As in other studies, minority race was associated with lower study completion rates (5, 13, 18). The need to retain representative proportions of a population by race/ethnicity may require overrecruiting minorities and carefully monitoring study completion within racial/ethnic subgroups to minimize differential attrition. However, it is important to note that SEED families where the mother was black or African American were less likely to comply with providing a medical-record release or providing blood samples from mother or child. Epidemiologic studies seeking generalizability across racial groups might need to consider strategies to make these components more acceptable or consider omitting these types of components to ensure more even participation by race.

Also consistent with previous research, participants' responsiveness to initial contact attempts and willingness to provide multiple phone numbers at enrollment were associated with study completion (8, 14, 18, 26–30). Our experience suggested that 3 call attempts were sufficient to enroll about two-thirds of individuals, and these individuals were more likely to complete study protocols. Operational planning for studies seeking to enroll families of young children might need to consider the balance of committed staff time for enrollment efforts against the likelihood of participant completion.

Contrary to previous literature (4, 27, 31) indicating that study burden (length, complexity, survey design) is associated with protocol completion, we found that the potential ASD group, who had an additional 4.5 hours of study components to complete compared with the main DD and POP groups, were just as likely to complete per protocol. In SEED, families with a child in the potential-ASD protocol might have been especially motivated to complete the protocol. Families were informed at enrollment that they would receive feedback about their child's developmental assessment. This could have encouraged completion of the in-person evaluation, particularly in situations where long waiting periods for developmental assessments were present in clinical settings.

The burden of farthest travel distance to the nearest study clinic ( 39 km) was associated with lower likelihood of completion as compared with a distance of <12 km, despite provision of higher financial incentives for this component compared with other components. This was particularly true for families in the main POP protocol group; this group was generally composed of children who were typically developing with no symptoms of ASD and whose families may have been less likely to perceive benefit from the study. However, they were only slightly less likely to complete. Notably, variability in travel



distance of 12–38 km was not associated with completion, suggesting some degree of travel is not a barrier to completion of an in-person evaluation.

Predictors of completion for the group who completed at least the PCI were similar to that of all enrollees, except that number of call attempts to complete enrollment was no longer predictive of per-protocol completion. These similar patterns indicate similar commitment to complete the study between these 2 groups. Nearly 80% of enrollees completed the PCI, and none of the additional factors collected during the PCI were associated with completion. This leaves open the question of what factors most influence participants' willingness to complete the protocol when a certain level of compliance can be achieved.

Studies have indicated that poor health of the enrolled subject is usually predictive of attrition in longitudinal studies of adults (5, 10, 15, 32). However, we found no reports of either the effect of chronic health conditions of a child in the family, or of single-adult households, on completion of child development research protocols. We anticipated that having more than 1 child with chronic medical or behavioral conditions, living in a household with a single adult caretaker, or low household income would decrease the likelihood of providing complete study data, but none of these factors were associated with completion after controlling for other demographic factors, such as race and maternal education. In models stratified by race, again, none of these factors were associated with completion. Families in child-development studies may be motivated to overcome barriers to participation such as complex life circumstances, particularly if they are seeking developmental evaluations that are challenging to obtain from clinical or educational sources (33). Additionally, we provided monetary incentives for completing study components. Among families who completed at least the PCI, the lack of association between study completion and household income suggests that the observed association between mother's education and completion may not necessarily reflect an economic issue.

Identifying factors that influence who completes a study can inform the selection of the most effective recruitment and retention strategies (12, 29, 32, 34). SEED is a unique study with a complex data-collection protocol, making it difficult to compare with other published studies. For more than half of the children enrolled in this phase of SEED, all data-collection components designated as per-protocol were completed. Completion of individual components was even higher. This is important to underscore because completion rates for individual SEED analyses vary, as they might for other similar studies of causes and correlates of diseases or disorders in children. For example, the primary data-collection instruments needed for most SEED analyses of ASD risk factors are the PCI and in-person evaluation, which were completed by 71% of enrollees. Future epidemiologic studies of children might consider prioritizing specific types of data collection that may have higher compliance within their target population.

## Limitations

While we examined a substantial set of factors in association with study completion, there are other potentially relevant factors that we were unable to examine. Reports from pediatric intervention trials suggest that retention is associated with whether or not mothers worked outside the home or the child had begun formal schooling, stressful life events such as loss

of job by income-earning parent(s) or the death or declining health of an adult family member, and mental or physical health problems of the mother (14, 24, 26, 35–37), which SEED was unable to evaluate. SEED staff did not ask participants why they did not complete specific study components, so we were unable to formally assess how participants balanced the burden of a demanding study protocol with motivation to complete this study or even specific components of the study. We were also unable to evaluate additional administrative factors that varied by site, but were not systematically tracked, such as deviations from standard protocol order, appointment reminder practices, or length of time to follow-up after a missed appointment. While SEED was unable to account for these factors, other studies may be able to address or account for these in their operational planning.

To the best of our knowledge, SEED is the first multisite, epidemiologic study of early childhood development to examine a variety of factors associated with completion of a protocol that includes such an array of data-collection components collected at separate time points. Previous studies have examined fewer factors associated with completion and/or limited types of data components. Additionally, many of the previous reports on pediatric study completion are clinical trials or intervention studies that provide some potential personal benefit to participants, or longitudinal surveys that do not include multifaceted protocols. Nevertheless, many of our findings are similar to these other types of studies.

Better understanding of the myriad factors that might influence completion of specific types of studies can improve planning of future, similar studies and help maximize participant retention. Results from our analysis of SEED suggest that most caregivers are willing to complete phone interviews and many are also willing to complete questionnaires, buccal swab sampling, and in-person visits. However, there are racial and educational differences in completion rates, regardless of initial responsiveness to enrollment. We recommend that racial and educational differences be monitored and accounted for with well-planned recruitment and completion strategies. Additionally, SEED findings suggest that certain factors—such as maternal race, education, or distance from a study clinic—may not be equally associated with completion of all types of data components in all subgroups, which may have implications for planning future studies of child development.

## Acknowledgments

This work was supported by the Centers for Disease Control and Prevention (cooperative agreements U10DD000180, U10DD000181, U10DD000182, U10DD000183, U10DD000184, and U10DD000498). Pennsylvania birth certificate data were supplied by the Bureau of Health Statistics and Research, Pennsylvania Department of Health, Harrisburg, Pennsylvania.

We thank those involved in the recruitment of families and the project staff, particularly Patricia Leonard and Joy Kloetzer, who proposed that those who will complete the study are those who respond quickly to enrollment contacts.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Additionally, the Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions.

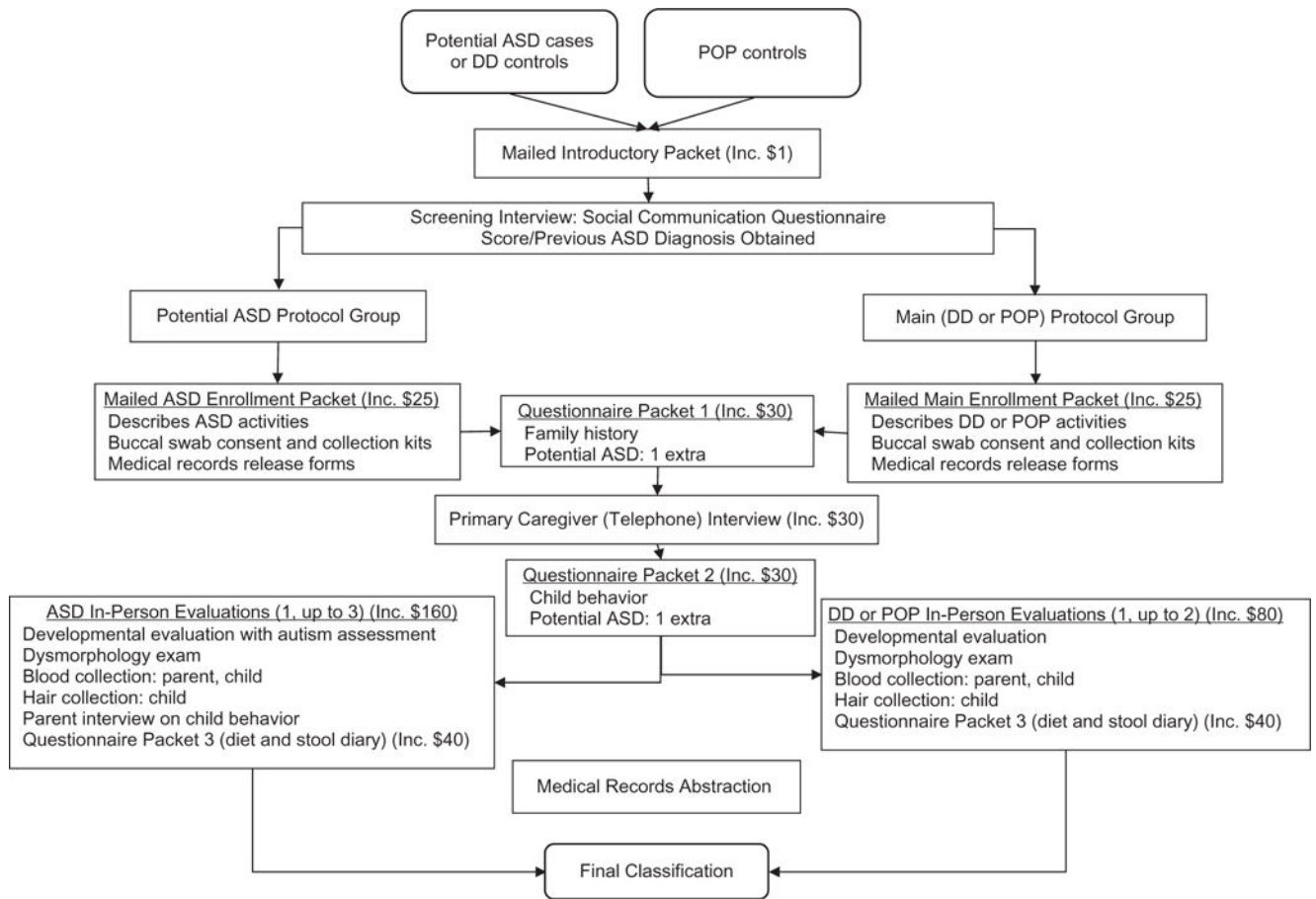
## Abbreviations

<b>ASD</b>	autism spectrum disorder
<b>CI</b>	confidence interval
<b>DD</b>	developmental delay
<b>OR</b>	odds ratio
<b>PCI</b>	Primary Caregiver Interview
<b>POP</b>	population controls
<b>SEED</b>	Study to Explore Early Development

## References

- Ahern K, Le Brocque R. Methodological issues in the effects of attrition: simple solutions for social scientists. *Field Methods*. 2005; 17(1):53–69.
- Demark-Wahnefried W, Bowen DJ, Jabson JM, et al. Scientific bias arising from sampling, selective recruitment, and attrition: the case for improved reporting. *Cancer Epidemiol Biomarkers Prev*. 2011; 20(3):415–418. [PubMed: 21382981]
- von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Prev Med*. 2007; 45(4):247–251. [PubMed: 17950122]
- Frankel LL, Hillygus DS. Looking beyond demographics: panel attrition in the ANES and GSS. *Polit Anal*. 2014; 22(3):336–353.
- Aylward GP, Hatcher RP, Stripp B, et al. Who goes and who stays: subject loss in a multicenter, longitudinal follow-up study. *J Dev Behav Pediatr*. 1985; 6(1):3–8. [PubMed: 3882762]
- Beaver KM. Intelligence and selective attrition in a nationally representative and longitudinal sample of Americans. *Pers Individ Dif*. 2013; 55(2):157–161.
- Eckholm O, Gundgaard J, Rasmussen NK, et al. The effect of health, socio-economic position, and mode of data collection on non-response in health interview surveys. *Scand J Public Health*. 2010; 38(7):699–706. [PubMed: 20851845]
- Cotter RB, Burke JD, Loeber R, et al. Predictors of contact difficulty and refusal in a longitudinal study. *Crim Behav Ment Health*. 2005; 15(2):126–137. [PubMed: 16470508]
- DeVita DA, White MC, Zhao X, et al. Determinants of subject visit participation in a prospective cohort study of HTLV infection. *BMC Med Res Methodol*. 2009; 9:19. [PubMed: 19284565]
- Fullerton B, Erler A, Pohlmann B, et al. Predictors of dropout in the German disease management program for type 2 diabetes. *BMC Health Serv Res*. 2012; 12:8. [PubMed: 22233930]
- Swan-Kremeier LA, Mitchell JE, Twardowski T, et al. Travel distance and attrition in outpatient eating disorders treatment. *Int J Eat Disord*. 2005; 38(4):367–370. [PubMed: 16254872]
- Young AF, Powers JR, Bell SL. Attrition in longitudinal studies: who do you lose? *Aust NZ J Public Health*. 2006; 30(4):353–361.
- Fischer EH, Dornelas EA, Goethe JW. Characteristics of people lost to attrition in psychiatric follow-up studies. *J Nerv Ment Dis*. 2001; 189(1):49–55. [PubMed: 11206665]
- Green SM, Navratil JL, Loeber R, et al. Potential dropouts in a longitudinal study: prevalence, stability, and associated characteristics. *J Child and Fam Stud*. 1994; 3(1):69–87.
- Eaton WW, Anthony JC, Tepper S, et al. Psychopathology and attrition in the epidemiologic catchment area surveys. *Am J Epidemiol*. 1992; 135(9):1051–1059. [PubMed: 1595691]
- Hewitt CE, Kumaravel B, Dumville JC, et al. Assessing the impact of attrition in randomized controlled trials. *J Clin Epidemiol*. 2010; 63(11):1264–1270. [PubMed: 20573482]

17. Petersen JC, Pirraglia PA, Wells MT, et al. Attrition in longitudinal randomized controlled trials: home visits make a difference. *BMC Med Res Methodol.* 2012; 12:178. [PubMed: 23176384]
18. Siddiqi A, Sikorskii A, Given CW, et al. Early participant attrition from clinical trials: role of trial design and logistics. *Clin Trials.* 2008; 5(4):328–335. [PubMed: 18697847]
19. Hunt JR, White E. Retaining and tracking cohort study members. *Epidemiol Rev.* 1998; 20(1):57–70. [PubMed: 9762509]
20. Schendel DE, DiGuseppi C, Croen LA, et al. The Study to Explore Early Development (SEED): a multisite epidemiologic study of autism by the Centers for Autism and Developmental Disabilities Research and Epidemiology (CADDRE) network. *J Autism Dev Disord.* 2012; 42(10):2121–2140. [PubMed: 22350336]
21. Rutter, M., Bailey, A., Lord, C. SCQ: Social Communication Questionnaire. Los Angeles, CA: Western Psychological Services; 2003.
22. Wiggins L, Reynolds A, Rice C, et al. Using standardized diagnostic instruments to classify children with autism in the study to explore early development. *J Autism Dev Disord.* 2015; 45(5):1271–1280. [PubMed: 25348175]
23. Zook PM, Jordan C, Adams B, et al. Retention strategies and predictors of attrition in an urban pediatric asthma study. *Clin Trials.* 2010; 7(4):400–410. [PubMed: 20571137]
24. Haring R, Alte D, Volzke H, et al. Extended recruitment efforts minimize attrition but not necessarily bias. *J Clin Epidemiol.* 2009; 62(3):252–260. [PubMed: 18834716]
25. Dickinson HO, Rapp M, Arnaud C, et al. Predictors of drop-out in a multi-centre longitudinal study of participation and quality of life in children with cerebral palsy. *BMC Res Notes.* 2012; 5:300. [PubMed: 22704327]
26. Anderson JW, Fass R, van der Horst C. Factors associated with early study discontinuation in AACTG studies, DACS 200. *Contemp Clin Trials.* 2007; 28(5):583–592. [PubMed: 17395549]
27. Cotter RB, Burke JD, Loeber R, et al. Innovative retention methods in longitudinal research: a case study of the Developmental Trends Study. *J Child Fam Stud.* 2002; 11(4):485–498.
28. Robinson KA, Dennison CR, Wayman DM, et al. Systematic review identifies number of strategies important for retaining study participants. *J Clin Epidemiol.* 2007; 60(8):757–765. [PubMed: 17606170]
29. Morrison TC, Wahlgren DR, Hovell MF, et al. Tracking and follow-up of 16,915 adolescents: minimizing attrition bias. *Control Clin Trials.* 1997; 18(5):383–396. [PubMed: 9315423]
30. Koss E, Peterson B, Fillenbaum GG. Determinants of attrition in a natural history study of Alzheimer disease. *Alzheimer Dis Assoc Disord.* 1999; 13(4):209–215. [PubMed: 10609669]
31. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol.* 2007; 17(9):643–653. [PubMed: 17553702]
32. Goldberg M, Chastang JF, Zins M, et al. Health problems were the strongest predictors of attrition during follow up of the GAZEL cohort. *J Clin Epidemiol.* 2006; 59(11):1213–1221. [PubMed: 17027433]
33. Vohra R, Madhavan S, Sambamoorthi U, et al. Access to services, quality of care, and family impact for children with autism, other developmental disabilities, and other mental health conditions. *Autism.* 2014; 18(7):815–826. [PubMed: 24353274]
34. Yancy AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. *Annu Rev Public Health.* 2006; 27:1–28. [PubMed: 16533107]
35. Williams PL, Van Dyke R, Eagle M, et al. Association of site-specific and participant-specific factors with retention of children in a long-term pediatric HIV cohort study. *Am J Epidemiol.* 2008; 167(11):1375–1386.
36. Warnick EM, Gonzalez A, Weersing R, et al. Defining dropout from youth psychotherapy: how definitions shape the prevalence and predictors of attrition. *Child Adolesc Ment Health.* 2012; 17(2):76–85.
37. Karlson CW, Rapoff MA. Attrition in randomized controlled trials for pediatric chronic conditions. *J Pediatr Psychol.* 2009; 34(7):782–793. [PubMed: 19064607]



**Figure 1.**

Standard flow of study components for the Study to Explore Early Development, United States, 2007–2011. Potential cases of autism spectrum disorder (ASD) and developmental delay (DD) controls (non-ASD developmental disabilities) were identified from educational or clinical service providers. Population controls (POP) were identified from birth records. Incentive (Inc.) was given as check or gift card (varied by site operations), except with the Introductory Packet. The Introductory Packet incentive was a beach ball worth approximately \$1. To be assigned to the potential-ASD protocol group, children had a previous ASD diagnosis and/or a Social Communication Questionnaire score of  $\geq 11$ . To be assigned to the DD or POP control groups, children had a Social Communication Questionnaire score of  $<11$  and remained associated with the source group from which they were identified. Final classification as a case of ASD or as a DD or POP control was based on the outcome of the developmental evaluation.

Table 1

Data-Collection Components According to Inclusion in Per-Protocol Completion Measure, Study to Explore Early Development, 2007–2011

Component <sup>a</sup>	Included in Per-Protocol Completion Measure? <sup>b</sup>
SCQ administered to primary caregiver via phone at enrollment	Not applicable
Medical records release: prenatal care, labor and delivery, neonatal care, and pediatric care (Consent and medical provider release forms included in initial enrollment packet, but medical records abstraction typically occurred much later in the study.)	
Provision of release to review at least 1 medical record from mother or child <sup>c</sup>	Yes
Questionnaire Packet 1 (mailed to primary contact and self-administered/caregiver-administered or completed with study staff assistance via phone or in person)	
Maternal and family medical history forms: maternal medical history, family autoimmune history, and child gastrointestinal function questionnaire	Yes
Paternal forms <sup>d</sup> : paternal medical history form and paternal occupational history form	No
Child development form: early development questionnaire (potential-ASD protocol families only)	Yes
PCI about family sociodemographic factors, maternal reproductive history and pregnancy health and behaviors, and early child development (administered via phone to mother and/or other primary caregiver)	Yes
Questionnaire Packet 2 (mailed to primary contact and self-administered/caregiver-administered, or completed with study staff assistance via phone or in person)	
Child behavioral development forms: Child Behavior Checklist, Carey Temperament (or Behavioral Styles) Scales (depending on child's age), Social Responsiveness Scale (preschool or child version, depending on child's age), and sleep habits questionnaire	Yes
Services and treatment questionnaire (potential-ASD protocol families only)	Yes
Maternal forms: Social Responsiveness Scale (adult version)	No
Paternal forms <sup>d</sup> : Social Responsiveness Scale (adult version)	No
In-person evaluation(s) (1 or more in-person visits in clinical setting or child's home; included several components involving child and primary caregiver and/or biological parents)	
Developmental assessments	Yes (as applicable)
Mullen Scales of Early Learning (potential-ASD and main (POP and DD) protocol families)	
Vineland Adaptive Behavior Scale–II (potential-ASD protocol families <sup>e</sup> )	
Autism Diagnostic Interview–Revised (potential-ASD protocol families only)	
Autism Diagnostic Observation Scale (potential-ASD protocol families only)	
Dysmorphology examination <sup>f</sup>	Yes
Biological samples (buccal swabs self/parent-collected; blood and hair specimens collected in person by study staff)	
Child: buccal swabs and/or blood	Yes
Mother: buccal swabs and/or blood	Yes
Father: buccal swabs and/or blood <sup>d</sup>	No
Child: hair sample	No
Questionnaire Packet 3 <sup>g</sup> (provided to primary caregiver as a single booklet during final in-person visit with instructions to complete and return to study site)	
Child 3-day diet and 7-day stool diary	No

Abbreviations: ASD, autism spectrum disorder; DD, developmental delay; PCI, Primary Caregiver Interview; POP, population controls; SCQ, Social Communication Questionnaire.

<sup>a</sup>Data-collection components listed in the usual chronological order of offer to study respondents except buccal swabs. Buccal swab self-collection kits were mailed to families at enrollment. If not received by the time of in-person evaluation, they were requested then. Completion and return of forms included in Questionnaire Packets 1 and 2 occurred throughout the data-collection period. In some instances, multiple follow-up calls were made to remind participants to complete the forms, and/or forms were completed with assistance from staff via phone or during the in-person evaluation.

<sup>b</sup>Per-protocol completion measure was developed for operational purposes (as opposed to being tied to any specific scientific study objectives). However, inclusion of selected components would contribute most to analyses of main study research questions. This measure was used throughout the study to assist sites in monitoring progress. Completion of every study component was pursued equally, but participants could refuse any component and continue with other aspects of the study. For most scientific analyses, the actual number of children who could be included is substantially higher than the number considered complete for all components of per-protocol measure.

<sup>c</sup>Medical records component was considered complete if 1 or more medical provider release forms were returned to project staff. However, in some instances medical provider releases were provided, but medical records could not be abstracted because records could not be obtained from the provider or incomplete records were sent by the provider.

<sup>d</sup>Paternal forms and samples were not always possible to collect for reasons other than refusal (e.g., father deceased, father not in touch with mother/primary caregiver, father unknown).

<sup>e</sup>Vineland Adaptive Behavior Scale II could be administered to main (DD or POP) families if child's performance on composite score of Mullen Scales of Early Learning fell below the standard.

<sup>f</sup>Dysmorphology examination component was considered complete if the examination form was completed. However, in a small number of instances, examination photos could not be obtained from child. These children could not be included in analyses requiring dysmorphology data even though they had completed the examination.

<sup>g</sup>Questionnaire Packet 3 consisted of 3-day diet and 7-day stool diaries, combined into a single booklet. This packet was given only to families who were seen for the in-person evaluation because the instructions for recording diet and stool quality information required in-person interaction with caregivers.

**Table 2**  
Demographic and Administrative Factors for Enrolled Families, According to Enrollment Site, Study to Explore Early Development, United States, 2007–2011

Demographic or Administrative Factor	Enrollment Site					Total (n=3,769), % <sup>a</sup>
	California (n = 651), % <sup>a</sup>	Colorado (n = 589), % <sup>a</sup>	Georgia (n = 723), % <sup>a</sup>	Maryland (n = 560), % <sup>a</sup>	North Carolina (n=666), % <sup>a</sup>	
Family demographic factors						
Mother's race						
White	70	95	54	71	76	64
Black or African American	5	2	41	26	21	27
Other	22	3	5	2	3	7
Missing	2	0	<1	1	0	2
Mother's Hispanic origin						
Hispanic	31	23	6	10	4	6
Missing	1	0	11	<1	0	<1
Mother's education						
Up to a high school diploma	28	31	23	30	23	25
Some college or Bachelor's degree	44	48	51	44	51	23
Advanced degree	26	20	22	25	26	51
Missing	3	1	4	1	<1	1
Mother's age at enrollment, years						
18.0–24.9	3	3	3	5	5	4
25.0–34.9	29	40	32	37	37	33
35.0–44.9	59	53	56	53	54	57
45.0	9	4	8	5	4	7
Administrative factors						
No. of call attempts to enroll family						
1	23	34	21	33	15	29
2–3	37	39	43	48	45	32
4	39	28	36	18	40	39
No. of contact phone numbers provided						
1	23	34	21	33	15	29
2–3	37	39	43	48	45	32
4	39	28	36	18	40	39



Demographic or Administrative Factor	Enrollment Site							Total (n=3,769), % <sup>a</sup>
	California (n = 651), % <sup>a</sup>	Colorado (n = 589), % <sup>a</sup>	Georgia (n = 723), % <sup>a</sup>	Maryland (n = 560), % <sup>a</sup>	North Carolina (n = 666), % <sup>a</sup>	Pennsylvania (n = 580), % <sup>a</sup>		
1	42	64	60	45	32	66	51	
2	51	32	36	49	52	30	42	
3	7	5	4	6	16	4	7	
Protocol group assignment								
Potential ASD	41	44	41	44	40	52	43	
Main, DD	26	23	28	24	30	20	26	
Main, POP	33	33	31	33	29	28	31	
Distance to nearest study clinic, km								
<12.0	28	24	100 <sup>b</sup>	31	12	32	39	
12.0–22.9	29	33	0	23	13	15	18	
23.0–38.9	27	28	0	19	32	15	20	
39.0	11	11	0	19	38	32	18	
Missing	5	4	0	8	5	7	5	

Abbreviations: ASD, autism spectrum disorder; DD, developmental delay; POP, population controls.

<sup>a</sup>Totals may not sum to 100% due to rounding.

<sup>b</sup>Site generally conducted visits in participants' homes.

Table 3

Demographic Information Collected From Those Who Completed the Primary Caregiver Phone Interview for Enrolled Families, According to Enrollment Site, Study to Explore Early Development, United States, 2007–2011

Demographic Factor	Enrollment Site							Total (n = 2,989), % <sup>a</sup>
	California (n = 462), % <sup>a</sup>	Colorado (n = 550), % <sup>a</sup>	Georgia (n = 585), % <sup>a</sup>	Maryland (n = 425), % <sup>a</sup>	North Carolina (n = 540), % <sup>a</sup>	Pennsylvania (n = 427), % <sup>a</sup>		
Total household income before taxes (previous 12 months), \$ <sup>b</sup>								
<30,000	14	22	23	22	20	23	21	
30,000–70,000	16	29	23	23	30	26	25	
70,000–110,000	22	27	23	26	29	19	24	
>110,000	34	21	28	27	18	30	26	
Missing	14	2	3	3	2	1	4	
Language spoken at home								
English	71	84	94	98	97	95	90	
Spanish	16	13	2	<1	1	<1	5	
Other	12	3	4	2	2	5	5	
Missing	<1	<1	<1	0	<1	<1	<1	
No. of adults in the household (previous 12 months)								
1	7	7	15	15	12	13	12	
2	73	85	76	76	81	79	78	
3	16	7	9	9	7	8	9	
Missing	3	<1	<1	0	1	<1	1	
No. of other siblings with special needs								
0	87	71	85	79	86	70	80	
1	10	22	10	16	11	20	15	
2	1	5	4	4	3	9	4	
Missing	2	2	2	1	0	1	1	

<sup>a</sup>Totals may not sum to 100% due to rounding.

<sup>b</sup>Instrument was designed with overlapping income ranges. Ranges are presented as offered to respondents.

**Table 4**  
Study Data-Collection Components Completed for Enrolled Families, According to Enrollment Site, Study to Explore Early Development, United States, 2007–2011

Component <sup>a</sup>	Enrollment Site							Total (n = 3,769), %
	California (n = 651), %	Colorado (n = 589), %	Georgia (n = 723), %	Maryland (n = 560), %	North Carolina (n = 666), %	Pennsylvania (n = 580), %		
Social Communication Questionnaire	100	100	100	100	100	100	100	100
Medical record release <sup>b,c</sup>	62	90	69	64	71	65	70	70
Buccal swabs								
Child <sup>a</sup>	70	79	70	74	80	67	73	73
Mother <sup>a</sup>	70	79	69	74	80	68	73	73
Father <sup>a,d</sup>	58	68	48	53	58	46	55	55
Questionnaire Packet 1 <sup>b</sup>	69	87	73	68	78	69	74	74
Primary Caregiver Interview	71	93	81	76	81	74	79	79
Questionnaire Packet 2 <sup>b</sup>	65	76	70	62	71	63	68	68
In-person evaluation(s) <sup>e</sup>	66	91	78	66	70	62	72	72
Blood samples								
Child	55	76	42	55	50	45	53	53
Mother	58	78	56	63	63	49	61	61
Father <sup>d</sup>	26	54	21	34	23	22	29	29
Biological samples for per-protocol <sup>f</sup>	73	88	71	76	80	69	76	76
Questionnaire Packet 3 <sup>a,g</sup>	43	67	29	43	51	36	44	44
Completed all components of per-protocol measure <sup>h</sup>	55	68	58	54	65	51	58	58

<sup>a</sup>Data-collection components listed in the usual chronological order of offer to study respondents. Buccal swab self-collection kits were mailed to families at enrollment. If not received by the time of in-person evaluation, they were requested then.

<sup>b</sup>Completion and return of forms included in packets occurred throughout the data-collection period. In some instances, multiple follow-up calls were made to remind participants to complete the forms, and/or forms were completed with assistance from staff via phone or during the in-person assessment visit.

<sup>c</sup>Medical records component was considered complete if 1 or more medical provider release forms were returned to project staff. However, in some instances medical provider releases were provided, but medical records could not be abstracted because records could not be obtained from the provider or incomplete records were sent by the provider.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

<sup>p</sup>Paternal forms and samples were not always possible to collect for reasons other than refusal (e.g., father deceased, father not in touch with mother/primary caregiver; father unknown).

<sup>q</sup>In-person evaluation considered attended if Mullen Scales of Early Learning or dysmorphology examination was complete. Dysmorphology examination component was considered complete if the examination form was completed. However, in a small number of instances, examination photos could not be obtained from the child. These children could not be included in other analyses requiring dysmorphology data even though they had completed the examination

<sup>r</sup>Biological samples required to meet per-protocol measures included buccal swabs and/or blood from both child and mother.

<sup>s</sup>Questionnaire Packet 3 consisted of 3-day diet and 7-day stool diaries, combined into a single booklet. This packet was offered only to families who were seen for the in-person evaluation because instructions for recording diet and stool quality information required in-person interaction with caregivers.

<sup>t</sup>Per-protocol completion measure was developed for operational purposes (as opposed to being tied to any specific scientific study objectives). However, inclusion of selected components would contribute most to analyses of primary study research questions. This measure was used throughout the study to assist sites in monitoring progress. Completion of every study component was pursued equally, but participants could refuse any component and continue with other aspects of the study. For most scientific analyses, the actual number of children who could be included is substantially higher than the number considered complete for all components of per-protocol measure.

Table 5

Demographic and Administrative Factors Associated With the Per-Protocol Completion Measure Using Multivariable Mixed-Effect Logistic Regression Models, Study to Explore Early Development, United States, 2007–2011

Demographic or Administrative Factor	Study Completion Among All Participants <i>n</i> = 3,769		Study Completion Among Subset Who Completed PCI <i>n</i> = 2,989	
	OR <sup>a</sup>	95% CI	OR <sup>a</sup>	95% CI
Mother's race				
White	1.00	Referent	1.00	Referent
Black or African American	0.61	0.49, 0.75	0.53	0.40, 0.69
Other	0.89	0.67, 1.18	0.83	0.56, 1.24
Mother's Hispanic origin				
Non-Hispanic	1.00	Referent	1.00	Referent
Hispanic	0.94	0.74, 1.19	0.94	0.64, 1.37
Mother's education				
Up to high school diploma	1.00	Referent	1.00	Referent
Some college or Bachelor's degree	1.52	1.25, 1.83	1.31	1.01, 1.69
Advanced degree	1.97	1.57, 2.47	1.73	1.26, 2.36
Mother's age at enrollment, years				
18.0–24.9	0.61	0.41, 0.91	0.67	0.39, 1.15
25.0–34.9	0.82	0.69, 0.95	0.74	0.60, 0.91
35.0–44.9	1.00	Referent	1.00	Referent
45.0	0.75	0.56, 1.01	0.76	0.52, 1.11
No. of call attempts needed to enroll family				
1	1.00	Referent	1.00	Referent
2–3	0.97	0.81, 1.17	1.09	0.87, 1.37
4	0.68	0.57, 0.83	0.89	0.70, 1.12
No. of contact phone numbers provided				
1	1.00	Referent	1.00	Referent
2	1.36	1.17, 1.58	1.21	0.99, 1.48
3	1.65	1.23, 2.32	1.29	0.90, 1.87
Protocol group assignment				
Potential ASD	1.00	Referent	1.00	Referent

Demographic or Administrative Factor	Study Completion Among All Participants $n = 3,769$			Study Completion Among Subset Who Completed PCI $n = 2,989$		
	OR <sup>a</sup>	95% CI	P Value <sup>b</sup>	OR <sup>a</sup>	95% CI	P Value <sup>b</sup>
Main, DD	0.89	0.74, 1.06	0.20	0.99	0.79, 1.24	0.92
Main, POP	1.08	0.91, 1.28	0.39	1.27	1.01, 1.58	0.04
Distance to nearest clinic, km						
<12.0	1.00	Referent		1.00	Referent	
12.0–22.9	0.82	0.66, 1.03	0.09	0.85	0.64, 1.12	0.25
23.0–38.9	1.05	0.85, 1.34	0.58	1.07	0.80, 1.43	0.64
39.0	0.72	0.58, 0.92	0.01	0.76	0.56, 1.02	0.06
Total household income before taxes (previous 12 months), \$ <sup>c</sup>						
<30,000				1.00	Referent	
30,000–70,000				0.98	0.73, 1.30	0.88
70,000–110,000				1.10	0.78, 1.55	0.57
>110,000				0.78	0.54, 1.11	0.16
Language spoken at home						
English				1.00	Referent	
Spanish				1.40	0.78, 2.50	0.26
Other				0.85	0.52, 1.38	0.51
No. of adults in the household (previous 12 months)						
1				0.89	0.66, 1.20	0.46
2				1.00	Referent	
3				1.13	0.81, 1.58	0.47
No. of other siblings with special needs						
0				1.00	Referent	
1				0.99	0.77, 1.29	0.99
2				1.29	0.82, 2.03	0.27

Abbreviations: ASD, autism spectrum disorder; CI, confidence interval; DD, developmental delay; OR, odds ratio; PCI, Primary Caregiver Interview; POP, population controls.

<sup>a</sup>Adjusted for all other variables in the table.

<sup>b</sup>P value from multivariable mixed-effect logistic regression model.

<sup>c</sup>Instrument was designed with overlapping income ranges. Ranges are presented as offered to respondents.