Is Inclusivity an Indicator of Quality of Care for Children With Autism in Special Education?

abstract

OBJECTIVE: To determine whether the proportion of time spent in an inclusive educational setting, a process indicator of the quality of schooling for children with autism, improves key outcomes.

METHODS: Patients were 484 children and youth educated in special education with a primary diagnosis of autism in the National Longitudinal Transition Study-2. These individuals were ages 20 to 23 in 2007. We used propensity-score inverse probability of treatment weights to eliminate the effect of multiple confounders. A causal interpretation of the effect of inclusivity on key educational and functional outcomes still depends on a critical assumption, that inclusivity is not confounded by remaining, omitted confounders.

RESULTS: Compared with children with autism who were not educated in an inclusive setting (n = 215), children with autism who spent 75% to 100% of their time in a general education classroom (n = 82) were no more likely to attend college (P = .40), not drop out of high school (P = .24), or have an improved functional cognitive score (P = .99) after controlling for key confounders.

CONCLUSIONS: We find no systematic indication that the level of inclusivity improves key future outcomes. Research on educational and functional outcomes for children with autism can benefit from data on large samples of children educated in real-world settings, such as the National Longitudinal Transition Study-2, but more nuanced indicators should be developed to measure the quality of special education for children with autism. Pediatrics 2012;130:S179–S185
Children and youth on the autism spectrum frequently receive both health and education services, but schools remain the primary provider of services. Regardless of the source of care, measuring the quality of those services remains a challenge. In both education and health care, quality is often measured with process measures rather than outcome measures. Process measures generally involve some aspect of provider-client interaction believed to improve outcomes. For example, an indicator of quality of care for mental health services might involve follow-up outpatient visits after an inpatient admission.

In much the same way, special education uses process measures to measure and ensure treatment quality. By law, children in special education must have an Individualized Education Plan that sets goals for each child and a treatment plan to attain those goals. Another requirement under the Individuals with Disabilities Education Act is that children be educated in the least restrictive or inclusive setting (ie, a general education classroom). Inclusive education is defined as educating a child in a general education classroom with a focus on bringing the services to the child.

Process measures of quality, either in health care or education, presume that services delivered consistently with these measures produce improved outcomes. To a large degree, however, these linkages are presumed rather than demonstrated. A small number of studies have considered the link between inclusivity and key child outcomes, such as academic achievement or social skills. Most of these studies focus on younger children; fewer have assessed the experiences of adolescents. For a review of this literature, see Ferraioli and Harris.

In general, this literature has a range of methodological problems, such as very small samples of children in a small number of communities. More fundamentally, random assignment to treatment groups is difficult or impossible. As a result, researchers are left to find comparable groups of children in settings of varying inclusivity. Of course, one would expect children in different educational settings to differ in other ways, and this possibility is particularly likely given the heterogeneity of children on the autism spectrum. The spectrum includes children with autistic disorder, Asperger syndrome, and pervasive developmental disorder not otherwise specified, who experience varying degrees of severity. As a result, some of the research studies lack a comparison group entirely or make only a very limited effort to form a comparison group by matching children who do and do not experience inclusion. Fisher and Meyer, for example, use only 2 variables, age and a baseline measure of functioning, to match treatment and comparison groups. That the studies often involve a single community makes this task even more difficult; local variation in special education funding and policies represent at least 1 source of variability in educational placement unrelated to the child’s condition.

METHODS

Data

The National Longitudinal Transition Study-2 (NLTS2) is a 10-year study of youth with disabilities who were receiving special education services in public or state-supported special schools. The NLTS2 uses a nationally representative sample of youth in special education who were between the ages of 13 and 16 on December 1, 2000. 434 of these individuals had a primary diagnosis of autism based on parent report and were included in these analyses. The study collected data biannually in 5 waves from 2001 to 2009. The current study uses the Wave 2 data, collected in 2003, for characteristics of the school program and home environment and Wave 4 data, collected in 2007, for the outcome measures. (More details about the study can be found at http://www.nlts2.org/)

Waves and Instruments

This study uses data collected using 2 instruments: a parent telephone interview and a school program questionnaire. To assess academic performance, the NLTS2 also collected data using a direct assessment of a student’s abilities by a trained on-site professional, other than the student’s own teacher. We did not use these data directly in our study but treated an inability to participate in that assessment as an indicator of functional impairment.

Covariates

The choice of covariates represents potential confounders: variables that influence outcomes as well as the exposure, inclusivity. The parent interview provided data on explanatory variables including the severity of the youth’s disability, the level of family support for education, and demographic information. Four measures of functioning were included: the number of domains affected by the disability, a functional cognitive scale, a social skills scale, and whether the youth was able to be evaluated using the direct assessment. The first of these, the number of domains affected, ranged from 0 to 7 and included vision; hearing, expressive communication; receptive language; bidirectional communication; use of arms, hands, legs, and feet; and general health.

The functional cognitive scale measures a combination of parent-reported cognitive, sensory, and motor skills used in performing daily activities (such as counting change). Parents rated their child on a scale of 1 (“not at all well”) to 4 (“very well”) for each of
these skills. The rating for each skill was added to create the functional cognitive scale, which ranged from 4 (not at all well for any of the skills) to 16 (very well for all of the skills). The social skills scale ranged from 0 to 18 and measured the youth's ability to interact with family and friends. A scale ranging from 0 to 12 was used to measure family support for education at school and included the frequency with which the parent attended school meetings, school or class events, or volunteered at the school. A scale ranging from 1 to 9 was used to measure family support for education in the home and included the frequency with which the parent helped the youth with homework and talked with the youth about his or her school experience. Demographic characteristics included the youth's race and the parent's level of education. Parents were also asked to rate student persistence. Parents were asked how often youth kept working at something until it is finished, even if it takes a long time. Response categories were never, sometimes, or very often.

Exposure: Inclusivity

The primary exposure of interest in this analysis was the proportion of time the youth spent in a general education classroom. The school program questionnaire collected data on the courses that each student took during the 2003 school year and whether each course was taken in a general education or special education classroom. The number of courses taken in a general education classroom was divided by the total number of courses taken to calculate the proportion. The proportion of time spent in an inclusive setting was then categorized as 0%, 1% to 74%, or 75% to 100% of courses taken in a general education classroom.

Outcomes

Three outcomes were assessed in this analysis by using Wave 4 data: not dropping out of high school, any college attendance, and the cognitive functional scale at Wave 4. Not dropping out of high school was chosen instead of high school graduation because not all youth would be expected to graduate from high school by Wave 4. (Wave 5 data are not yet available.) Youth were coded as not dropping out if the parent reported that they graduated, received a certificate or General Educational Development certificate, or were still in high school at the time of Wave 4 data collection. Any college attendance was based on parent report of whether the youth attended any type of postsecondary school in the previous 2 years, including postsecondary classes to earn a high school degree, a 2-year or 4-year college, or postsecondary vocational school. The functional cognitive scale at Wave 4 was calculated in the same way as the scale used for Wave 2, described previously.

Propensity Score Methodology

In an observational study, the link between an exposure of interest and an outcome represents an association. Moving from that association to drawing causal inference depends on a key assumption. Researchers commonly assume "ignorability" or the absence of unobserved confounding. Under this assumption, the outcomes for individuals at a level of inclusivity represent a counterfactual for what other, comparable children would experience had they had that same level of inclusivity (rather than what they actually experienced). Ignorability assumes away unobserved differences over and above any covariates used to adjust comparisons of individuals at different levels of exposure. Ignorability essentially assumes that the exposure, inclusivity in this study, is randomly assigned among subgroups of participants sharing the same set of observed characteristics. Is this assumption plausible? It is impossible to fully test this assumption empirically, but at least a necessary condition for plausibility is that one select the correct covariates and omits incorrect ones. "Correct" in this sense means potential confounders: variables that influence both inclusivity and the outcomes of interest.

Analyses grounded in ignorability generally involve comparisons of outcomes across levels of exposure adjusted for the covariates selected. For example, an analyst might regress the outcome on exposure and the covariates selected. A second condition for causal inference in this case, even if ignorability is correct, is that the mechanics of regression (or other methods) work correctly. By this we mean that the adjustment mechanism fully "balances" the distribution of the covariates across levels of exposure. In a regression context, achieving balance involves specifying the functional form of the regression model correctly.

An alternative methodology for adjusting comparisons across levels of exposure for covariates involves propensity scores. These are the predicted probability of exposure, and they represent a convenient summary of the covariates. The propensity score can be used to calculate adjusted between-group means in a variety of ways, such as matching. Propensity score–based methods assume ignorability but have advantages over regression, such as producing estimates of the effect of the exposure with a clear interpretation and checking covariate balance. Propensity scores can be used...
in analyses that take various forms. We use inverse probability of treatment weights. Unlike matching, for example, this methodology easily generalizes beyond 2 levels of exposure. The weights are calculated as 1 over the probability of the exposure actually received. These weights can be incorporated in the analyses like survey weights; they represent pseudopopulations where the covariates and exposure are no longer related. As discussed earlier, we model inclusivity as an ordered category. To generate predicted levels of inclusivity, we used a multinomial logit model. (We might have used an ordered logit. Such a model would have been more parsimonious; it involves 1 regression coefficient for each covariate [half as many as the multinomial logit]).

Handling of Missing Data
Table 1 demonstrates that the data suffer from fairly extensive missing data. In analyzing missing data, one conducts separate analyses of the imputations (in our case 5) and combines the estimates using Rubin’s rules. The SEs of the resulting estimates reflect the uncertainty in each imputation-specific estimate as well as variation across imputations in the estimates. The latter captures the uncertainty stemming from the fact that the data are missing.

RESULTS
Descriptive Statistics
Table 1 describes the sample. The table reports the covariates (section A), the level of inclusivity experienced (section B), and outcomes of interest (section C). One can see that the vast majority of youth with autism are male. Roughly 6 in 10 participated in direct assessments as part of the study. The average child had his or her conditioned identified early (age 2).

The table also provides information on the exposure, inclusivity. One can see that nearly half (45%) spent no time in the regular classroom; 17% spent three-quarters or more of their school day in regular classroom settings. Presumably, this variation reflects the child’s characteristics as well as “supply side” factors, such as the range of special education services offered in the school.

Generating Propensity Scores
Table 2 presents the results of the multinomial logit. One can see that
### TABLE 2 Multinomial Logit Model Predicting Level of Inclusivity

<table>
<thead>
<tr>
<th>Outcome Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3 presents the unadjusted and adjusted levels of 3 outcomes across the levels of the inclusivity variable. Looking across the 3 outcomes, one can see that the effect of adjusting for the covariates narrows the variation across levels of exposure. For not dropping out of high school, those spending no time in inclusive settings are least likely to continue in high school; the gap between those youth and those spending all of their time in inclusive settings is 17 percentage points. Adjusting for the covariates narrows this gap to 8 percentage points. For the second outcome, college attendance, the gap between the highest and lowest categories is enormous: 57 percentage points. Those with moderate inclusivity fall in between, as one would expect. Adjusting for the covariates closes this gap to 14 percentage points. This estimate is sizable, but the null hypothesis of no effect cannot be rejected because of the high imprecision associated with the estimate. For the third outcome, the score on the functional cognitive scale, the between-group difference is largely unchanged by adjusting for the covariates. For all 3 outcome measures, adjusting for confounders removes the effect of inclusive education on the outcome. Compared with children who were not educated in an inclusive setting, children who spent 75% to 100% of their time in a general education classroom were no more likely to attend college (P = .40), not drop out of high school (P = .24), or have an improved functional cognitive score (P = .99) after controlling for key confounders.</td>
</tr>
</tbody>
</table>

---

**TABLE 3 Associations of Inclusivity With Key Outcomes**

<table>
<thead>
<tr>
<th>% of Time Spent in an Inclusive Setting</th>
<th>Not Dropping Out</th>
<th>Some College Attendance</th>
<th>Functional Cognitive Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td>Unadjusted</td>
</tr>
<tr>
<td>0% (Reference category)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% to 74%</td>
<td>0.06</td>
<td>0.01</td>
<td>0.23</td>
</tr>
<tr>
<td>Effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>−0.04</td>
<td>−0.04</td>
<td>−0.07</td>
</tr>
<tr>
<td>P value</td>
<td>0.38</td>
<td>0.38</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>75% to 100%</td>
<td>0.17</td>
<td>0.08</td>
<td>0.57</td>
</tr>
<tr>
<td>Effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>−0.07</td>
<td>−0.07</td>
<td>−0.08</td>
</tr>
<tr>
<td>P value</td>
<td>0.12</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>P value for joint significance</td>
<td>0.05</td>
<td>0.68</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Observations</td>
<td>434</td>
<td>434</td>
<td>434</td>
</tr>
</tbody>
</table>

---

Note: the “effects” for the dichotomous outcomes are the so-called marginal effects, that is the adjusted mean difference in the predicted probability of the outcome. For the third, the effect is the type regression coefficient. In general, our analyses suggest that inclusivity does not improve educational or functional outcomes for children with autism. Any remaining biases would seem to be in the direction of...
overstating effects: that better functioning youth were, all else equal, still more likely to be in inclusive settings. In that case, the apparent effect of inclusivity on high school completion is an overestimate. There is, however, no real way to test this possibility. It is entirely possible that unobserved differences biased the estimated effects of inclusivity toward 0. The best one can do is qualitatively judge the plausibility of ignorability based on the list of included covariates and what one knows about the processes determining exposure (inclusivity).

A strength of the study is that the list of covariates included in the analyses is more extensive than that used in previous research. It is striking that adding these covariates did not change the relationship between inclusivity and the outcomes very much. One interpretation of these findings is as an assessment of criterion validity of these covariates. If these measures cannot predict features of special education involvement (like inclusivity), then perhaps these measures are not as strong as believed.

Of course, another possibility is that placement in special education is essentially randomly assigned. This possibility is alarming. Perhaps youth find their way into placement based on school and district characteristics, such as funding, unrelated to their own needs and goals. We know from other data that schools, districts, and states differ enormously in their funding and eligibility requirements for special education. If funding influenced children’s outcomes directly and fostered inclusivity, then funding would represent a confounder. In that case, one would expect the “effect” of placement to be inflated: children in affluent areas receive more inclusive services and perform better as well.

Given that inclusivity should reflect youth characteristics, observational analyses of this type are challenging. In fact, we are assuming in essence that the schools do not meet their obligation to these youth. If they did, confounding by both observed and unobserved variables would be so severe that the analyses would be impossible.

Even if the ignorability assumption is valid, it remains true that the effect measured here is that of a rather amorphous “treatment.” Still, although the measure of inclusivity is crude, it is one of the main measures of the quality of special education. The link between this measure and important outcomes is weak. It remains true, however, that inclusivity well-implemented and supported might have substantial benefits; however, refining the exposure of interest in this way raises the challenge of causal inference in general, and the plausibility of ignorability in particular. The processes determining the availability of supportive services, for example, may share determinants with the outcomes of interest, creating the potential for further confounding.

The study illustrates the challenges of understanding the effect of real-world services and treatments, especially those involving a rather small heterogeneous group. We used data that have many strengths relative to those used in previous research. NLTS2 data are nationally representative and include relatively many children and youth with autism. Nonetheless, the data lack key measures specific to the characteristics and education of children with autism. Such information, for example, would allow better differentiation among the heterogeneous children identified by their parents as having a diagnosis of autism. Progress in understanding the quality of education for these children may very well depend on the development of datasets for children with autism that are enriched by measures developed for that population, include careful description of their learning environment, and sufficient size to illuminate variation in the experiences of these children within and between communities. Community-level variation in school policies and strategies may represent natural experiments that provide instrumental variables or regression discontinuities that offer the potential for valid causal inference in the absence of ignorability. A fuller understanding of inclusivity and other potential measures of educational quality may have to wait for both better data and methods.

REFERENCES
2. Foster EM. Service use under the continuum of care: do followup services forestall hospital readmission? Health Serv Res. 1999;34(3):715–736
Is Inclusivity an Indicator of Quality of Care for Children With Autism in Special Education?
E. Michael Foster and Erin Pearson

*Pediatrics* 2012;130;S179
DOI: 10.1542/peds.2012-0900P

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/130/Supplement_2/S179.full.html