



University of Massachusetts Amherst

Department of Resource Economics

Working Paper No. 2009-3

<http://www.umass.edu/resec/workingpapers>

Estimation of Treatment Effect of Asthma Case Management Using Propensity Score Methods

Sylvia Brandt¹, Sara Gale², and Ira Tager³

Abstract:

Objective: To estimate the treatment effect from participating in an asthma intervention that was part of the National Asthma Control Program.

Study Setting: Data on children who participated in asthma case management (N=270) and eligible children who did not participate in case management (N=2,742) were extracted from a claims database.

Study Design: We created 81 measures of health care utilization and 40 measures of neighborhood characteristics that could be related to participation in the program. The participation model was selected using the cross-validation-based Deletion Substitution and Addition (DSA) algorithm. We used optimal full matching for the vector of Mahalanobis' distances and propensity scores to estimate the difference between participants and non-participants in the probability of a range of asthma outcomes.

Principal Findings: Compared to non-participants, participants were more likely to have vaccinations for pulmonary illness, use controller medications, and have a refill for rescue medication. There was no statistically significant difference in the number of nebulizer treatments or ED visits between the two groups. We find that the asthma program had no significant effect on overall asthma control.

Conclusion: We are not able to discern whether the lack of an effect in overall control is due to the effectiveness of the program, heterogeneity of effects or barriers outside the program's control. We discuss how current programs could be modified to better inform future research and program design.

Keywords: Asthma, treatment effect, health intervention, propensity scores

JEL Classification: I1, D13, H51, C31, C80, C81, C83

¹Sylvia Brandt, Department of Resource Economics
University of Massachusetts, 219A Stockbridge Hall
80 Campus Center Way, Amherst, MA 01003-9246
E: brandt@resecon.umass.edu P: 413-545-5722 F: 413-545-5853

²Sara Gale, Department of Epidemiology
University of California, School of Public Health, Berkeley California
E: sagale@berkeley.edu

³Ira Tager, Department of Epidemiology
University of California, School of Public Health, Berkeley California
E: ibt@berkeley.edu

**Estimation of Treatment Effect of Asthma Case Management
Using Propensity Score Methods**

Sylvia Brandt, Sara Gale, and Ira Tager

Contact Author:

Sylvia Brandt
219 Stockbridge Hall
University of Massachusetts, Amherst
Amherst, MA 01003
413-545-5722
Brandt@resecon.umass.edu

Sara Gale and Ira Tager
Department of Epidemiology
School of Public Health
University of California, Berkeley

Acknowledgements:

We thank Dr. Arthur Chen and Elizabeth Edwards at Alameda Alliance for Health and Adam Davis of the East Bay Lung Association for their advice and support. This work would not have been possible without the staff and program participants of *Oakland Kicks Asthma*. We appreciate the statistical assistance from Long Ngo and Tad Haight. This research is supported by the Centers for Disease Control and Prevention (CDC) Control of Asthma in American Cities Project (CAACP), Cooperative Agreement Number U59 EH923264.

Estimation of Treatment Effect of Asthma Case Management Using Propensity Score Methods

Objective: To estimate the treatment effect from participating in an asthma intervention that was part of the National Asthma Control Program.

Study Setting: Data on children who participated in asthma case management (N=270) and eligible children who did not participate in case management (N=2,742) were extracted from a claims database.

Study Design: We created 81 measures of health care utilization and 40 measures of neighborhood characteristics that could be related to participation in the program. The participation model was selected using the cross-validation-based Deletion Substitution and Addition (DSA) algorithm. We used optimal full matching for the vector of Mahalanobis' distances and propensity scores to estimate the difference between participants and non-participants in the probability of a range of asthma outcomes.

Principal Findings: Compared to non-participants, participants were more likely to have vaccinations for pulmonary illness, use controller medications, and have a refill for rescue medication. There was no statistically significant difference in the number of nebulizer treatments or ED visits between the two groups. We find that the asthma program had no significant effect on overall asthma control

Conclusion: We are not able to discern whether the lack of an effect in overall control is due to the effectiveness of the program, heterogeneity of effects or barriers outside the program's control. We discuss how current programs could be modified to better inform future research and program design.

Key words: asthma, treatment effect, health intervention, propensity scores

1. Introduction

The most prominent example of public investment in asthma programs is the CDC's National Asthma Control Program. Created in 1999, the program has spent almost \$208 million (CDC, 2007)¹ on two intervention models - the Inner-City Asthma Intervention and the Controlling Asthma in American Cities Project (CAACP).² Each model attempts to achieve clinical-like results in a community setting similar to that of the National Cooperative Inner-City Asthma Study (NCICAS). NCICAS was a family-focused program targeted at inner-city children with moderate to severe asthma that included over one-thousand families randomized into intervention and control groups. A master's-level asthma educator met with caregivers and children in a combination of group classes and individualized meetings. The NCICAS intervention was publicized as "a scientifically proven asthma intervention program" (NIH, 2001), despite the fact that the assessment of the program found no statistically significant differences in utilization of health care between the intervention and control groups. While there was a statistically significant difference between the intervention and control groups in the maximum number of symptom days³ in year one of the study, after two years, differences were no longer significant.

During the design and implementation of the National Asthma Control Program, there was little attention paid to assessing the effectiveness of the interventions that were funded. While some programs have summarized process outcomes and qualitative assessments (for a

¹ In 2003, the CDC spending on asthma programs (\$36.9 million) was greater than for autism (\$10.8 million) but less than for diabetes (\$63 million) or lead prevention programs (\$42.0 million). The projects funded in 2003 included 7 tracking programs, 49 interventions and 39 community partnerships. For more information see <http://www.cdc.gov/asthma/aag07.htm#more>.

² The Inner-City Asthma Intervention was located in 23 sites. The programs all had asthma educators who were based in health organizations that treated low-income inner-city children. Controlling Asthma in American Cities Project (CAACP) was also based in low-income urban areas, but each intervention was locally developed and varied in the design and components. CAACP was implemented in seven cities.

³ The number of days with asthma symptoms over a two-week period averaged over six observations during the year.

review see the 2006 supplement to *Annals of Allergy, Asthma and Immunology*), there are few data on actual health outcomes and, therefore, no rigorous evaluation of these investments to determine if they are, in fact, the best use of available funds to reduce morbidity in childhood asthma. At present, whether these interventions in urban communities can shift behavior from treatment of asthma attacks to prevention of attacks through risk-reducing investments, and whether costs can be reduced and quality of life can be improved, remains an open question.

In this paper, we examine *Oakland Kicks Asthma* (OKA), a case management program that is part of the CAACP. Oakland, CA is among the most diverse cities in the United States, with 29% of children living in poverty (U.S. Census, 2006). The annual rate of asthma-related hospitalizations for children ages 0-14 is approximately twice that for the state (36.63 versus 18.73 per 10,000 residents/year).⁴ The *Oakland Kicks Asthma* (OKA) case management program was jointly developed by the American Lung Association of the East Bay and researchers at the UC Berkeley School of Public Health.⁵ Children were referred by physicians, the local Medicaid provider or identified by surveillance surveys administered in the middle and high schools of the Oakland Unified School District. Over three to five months, an asthma case manager visited each participating household 3-4 times to develop an asthma management plan and assist with its implementation. The program was free to all families.

⁴ The 95% confidence intervals are (34.59 to 38.72) versus (18.22 to 19.04) per 10,000 residents (Stockman et al., 2003).

⁵ The asthma program, data collection tools and all protocols for protection of personal data were approved by the Committee for the Protection of Human Subjects, UC, Berkeley.

2. Methods

Matching Participants and Non-participants

To determine the real effectiveness of the program, it is necessary to compare those outcomes to the outcomes realized by a comparison group that did not participate in the program. This is a non-trivial problem, because households do not randomly sort themselves into otherwise-identical treatment and control groups. Failure to correct, as completely as possible, for this non-random sorting will produce biased estimates of the treatment effect and a biased assessment of the program. We address this issue by matching participants to non-participants using the estimated propensity score.⁶

For each child who is eligible to participate in OKA, there are two possible decisions, participation and non-participation, denoted $T_i=1$ and $T_i=0$. Individuals differ on pre-intervention characteristics that are associated with participation (x_i). The pre-intervention variables fall into three different domains: health care use (e.g., count of pre-treatment ED visits), socio-demographics (e.g., race), and provider and plan data (e.g., AAH health care plan type). These domains are denoted u_i, s_i, p_i . The propensity score can be written as:

$$e(x_i) \equiv pr(T_i | X_i = u_i, s_i, p_i)$$

One caveat of propensity scores is that results may be sensitive to misspecification of the selection equation, especially in cases where there are quadratic terms (Drake, 1993). The economics and public health literature provides little guidance to construct our participation equation; thus, we rely on the cross-validation-based Deletion Substitution and Addition (DSA) algorithm to select a participation model. The DSA is a data-adaptive selection procedure based on cross-validation that relies on deletion, substitution, and addition moves to search through a

⁶ A cornerstone in evaluation methodology across disciplines is the seminal work by Rosenbaum and Rubin (1983, 1985).

large space of possible polynomial models (Moore *et al.*, 2008). The criterion for model selection is minimization of a loss function (empirical and cross-validated residual sum of squares) rather than tests of significance of parameters. The result is a vector of pre-treatment variables that includes interactions and polynomials $\tilde{x} \in X$. The DSA procedure is publicly available as an R package at <http://www.stat.berkeley.edu/~laan/Software/>. After using the DSA algorithm to select the model that best predicts the empirical distribution of participation, conditional on pre-treatment variables, propensity scores are calculated by logistic regression of participation on the selected vector \tilde{x}_i .

We use the most flexible method to match participants to non-participants on propensity scores, optimal full matching for the vector of Mahalanobis' distances and propensity scores (Haviland *et al.*, 2007 and Hansen and Klopfer, 2006).⁷ and ⁸ Full matching improves our statistical power over pair matching or any other subset matching (Rosenbaum, 1989), which is critical in our case where we have a limited number of observations on participants but a substantially larger number of non-participants. This matching proceeds in two steps. First, the Mahalanobis distances are defined as the difference between the vectors of pre-treatment variables that includes the logit of the propensity score for each participant (\tilde{x}_i) and non-participant (\tilde{x}_j) pair, weighted by the inverse of the covariance matrix of the pre-treatment variables and the logit of the propensity score, C:

⁷ There are several options for using the score to create comparison groups; see D'Agostino (1998) for a summary of three approaches. While pair matching has a long history in the literature, full matching was introduced relatively later (Rosenbaum, 1991).

⁸ The advantage of using the Mahalanobis distance is that it takes into account the covariance among the variables in calculation of distances; however, binary elements that are rare can overly influence matching based on Mahalanobis distances (Rosenbaum and Rubin, 1983). Inclusion of the propensity score in the vector of covariates can reduce bias in these situations essentially by balancing some deviations between individual covariates with closer matches on the composite of those covariates as reflected in the propensity score (Rubin and Thomas, 2000). Given that we know outcomes such as hospitalizations are rare, we include the propensity score in the vector of pre-treatment variables.

$$d_{i,j} \equiv (\tilde{x}_i - \tilde{x}_j)^T C^{-1} (\tilde{x}_i - \tilde{x}_j).$$

The differences or “distance” matrix, D , is formed with participants as rows and non-participants as columns and the Mahalanobis distance for each pair in the appropriate cell. In any case where the difference in propensity score between participant and non-participant is greater than 0.25 of the standard deviation of the logit of the observed propensity scores, we assign a penalty of 500 units⁹ to prevent matching.

In the second step, once the matrix D is constructed, optimal full matching uses an algorithm based on a minimum-cost network flow to create k to m matches of participants and non-participants. Essentially, the matching creates sets based on the values in the matrix D , and allows for a set to contain any number of participants or controls. The optimal match algorithm considers the group costs associated with a particular matched set. For instance, a match might be made on the second- or third-closest distance rather than the first because another set would be worse off if the closest match were made.

Comparison of Health Outcomes

We use health care encounters as markers of asthma exacerbations (undesirable health care encounters) and of managed or controlled asthma (appropriate health care encounters). First, we estimate the mean effect of participation on the probabilities of two binary measures of preventative health care and three binary measures of treatment of asthma exacerbations. In these five binary measures we account for the matching of participants to non-participants, using a

⁹ This penalty serves the same function as a caliper (Haviland et al., 2007) where the standard caliper is a quarter of a standard deviation of the differences in estimated propensity scores (Rosenbaum and Rubin, 1985).

conditional logistic regression to measure the effect of participation in the program within a matched group. However, individual measures of health care utilization give an incomplete picture of a child's overall asthma status; ideally, we would compare the evolution of health status between participants and non-participants. To do this, we create an indicator that captures the transition between normal and negative asthma states over time, and evaluate the effect of program participation on this indicator (H=1,2,3,4). We used Proc GLIMMIX to implement conditional logistic regression for this multinomial outcome. Matched sets were treated as random intercepts. The reference group was subjects who had stable outcomes before and after the intervention. This indicator is discussed in more detail below.

3. Study Data

Our data come from the claims database of Alameda Alliance for Health (AAH), a not-for-profit Medicaid umbrella organization that manages and provides health care services for low-income families in Alameda County. To construct our sample of participants and non-participants (controls), we examined all AAH plan members who had a health claim with a primary diagnosis code for asthma (ICD-9 of 493.x) and were between the ages of 6 and 20 at the time of service. We divided this set of children into those who had a record of participation in OKA (n=282) and non-participation (n=8,307). Figure I describes the criteria for inclusion in either the treatment or comparison group, and how many individuals were excluded at each step. We randomly assigned each potential control a "start date" based on the actual distribution of program start dates of participants (see Johnson et al., 2005)¹⁰ to create comparable pre-treatment

¹⁰ For instance, if 10 out of 270 children (4%) started OKA in February 2005, we randomly assigned 110 out of 2,742 potential controls (4%) the same start date.

and post-treatment periods. Because asthma symptoms often show seasonal patterns (Gergen et al., 2002), we consider outcomes at six-months and twelve-months after the intervention.

<<Insert Figure I here>>

Ideally, we would have complete symptom information on participants and non-participants. In the absence of these data for non-participants, we created an inventory of health care utilization measures based on the billing codes for asthma-relevant health service encounters (summarized in Table A.1 in the appendix) that was as comprehensive as possible. With the exception of home medical equipment, allergy testing, allergy immunology, and participation in an asthma program other than OKA, variables were defined as both a binary measure (0=no encounter, 1=any encounter) and a count of encounters. In addition to health care encounters, we identified asthma-related prescription drug claims and distinguished between rescue medications (for use in asthma exacerbation) and controller medications (preventative medication). An important marker of a significant asthma exacerbation in children is the use of systemic/oral glucocorticoids (e.g., Prednisone), although these are not limited to asthma treatment; a prescription for glucocorticoids was included in the measure of steroid use only if it was in conjunction with a health care encounter with an asthma diagnosis. As with health care utilization, for each class of medication (rescue, control, systemic steroid), we created both binary and count variables. In addition, we created indicators of excessive or adequate use based on recommended doses and prescription volumes.¹¹ Lastly, we created a medication intensity scale following Schatz et al (2006).¹² The result is a set of 72 measures of utilization and prescription use prior to participation in OKA.

¹¹ Information on doses per canister was compiled using a Medline search on each prescription inhaler. Documentation available from authors by request.

¹² To create medication intensity scale, we ran a stepwise logistic regression of number of prescriptions on an indicator for a visit to the emergency department for asthma. Cut-points in the number of prescriptions were defined

To evaluate an individual's interaction with health providers, we reviewed the primary care provider (PCP) and specialty codes for all asthma-related claims. To assess continuity of care, we created counts of the total number of different primary care providers for all asthma claims within six and twelve months prior to the program start date, as well as an indicator variable for no PCP assigned. We also created counts of encounters with each of three types of providers (pulmonologist, allergist, or any respiratory specialist). Lastly, we created an indicator for whether the child had ever had a claim with an allergy diagnosis. This set of individual characteristics included a total of nine variables.

The eligibility files for AAH members provided socio-demographic data that included race/ethnicity, language, age, type of health plan and group, number of others included in the health plan, and number of adults on the health plan. To account for measurement bias introduced by gaps in service, we created a marker for those children who had gaps that totaled more than 65 days in the year prior to or after the program start date. Twenty-eight children in the treatment group (10%) and 252 children in the potential control group (9%) had gaps in insurance coverage. We accounted for the potentially missing data that a gap in service might incur by inclusion of this variable in selection of the participation model using the DSA. There were eleven socio-demographic measures in total.¹³

Finally, to include community level variables, we geocoded our study subjects to the 2000 US Census tract level to describe each child's neighborhood composition. We considered forty measures that covered age distribution, country of birth, family structure (e.g., percentage married), employment status, and income characteristics of the neighborhood.

by breaks in the odds ratio for ED visit (see Appendix C for regression results). This approach to defining medication intensity has been found to be predictive of future utilization.

¹³ A complete list of the procedure and billing codes used to create each variable is available from the authors upon request.

In Table I, we describe the basic socioeconomic characteristics of the treatment and non-participant groups. Compared to the potential control group, the treatment group had slightly more Hispanic participants (and more Spanish speakers) and slightly fewer Black and Asian participants. This difference is perhaps due to recruiting efforts in a local clinic targeting Spanish-speaking families. The only other notable difference is that a substantially higher proportion of participants had an allergy diagnosis than non-participants. The small difference in the age distribution is explained by the program eligibility requirements.

<<Insert Table I here>>

As shown in Table II, the participant and non-participant groups differed in patterns of health care utilization. Relative to eligible AAH members who had an asthma diagnosis, OKA participants were more likely to have had an ED visit, outpatient visits, or a prescription for systemic steroids. A higher proportion of participants fell into the high medication intensity categories for both rescue and controller medications. Participants were somewhat more likely to have had an inpatient hospitalization. Interestingly, participants also were more likely to have had preventative care such as flu or pneumonia vaccines or pulmonary testing. As a group, participants appear to be heavier users of health care than non-participants.

<<Insert Table II here>>

4. Construction of the Comparison Group

The first step in constructing the comparison between program participants and non-participants is to estimate the propensity scores (propensity to participate in the program) of all individuals based on pre-treatment characteristics. The complete vector of all measures of pre-treatment characteristics consisted of 132 elements. We used the DSA to identify the most informative variables in each domain: health care use, socio-demographics, provider and plan

characteristics and neighborhood composition. We allowed for up to fifteen terms with a maximum of three interactions and a maximum power of three for each model. First we ran the DSA separately for each domain of explanatory variables. Next we created a subset of explanatory variables that were in any of the models that minimized the loss function four or more times out of all the runs of the data (we used 5 splits of the data for each of the 10 seeds). Finally we ran the DSA on this subset of variables. The result is a model for program participation based on a subset of pre-treatment characteristics, $\tilde{x} \in X$, with the maximal predictive power across the entire set of all such models.

<<Insert Table III here>>

Table III lists the measures for each domain that maximized predictive power as well as their odds ratios from the logistic regression for participation. Measures of utilization, allergy diagnosis, and number of different primary care providers have a positive impact on the probability of participation;¹⁴ the number of adults in the health plan has a negative impact. One characteristic of note is that the relationship between utilization and probability of participation is not linear in the variables; this highlights the importance of doing an exhaustive search over possible models.

While a behavioral model of the participation decision is interesting, the main importance of the participation regression is derivation of the propensity scores used to match participants and non-participants. Figure II shows that the propensity scores for the two groups had little overlap, which makes it more important to use full matching to make full use of the

¹⁴ Having an emergency room visit or urgent outpatient visit increased the probability of participation most likely because it served to highlight the negative outcomes associated with unmanaged asthma. The positive relationship between having controller medication and an allergy diagnosis and participation could be interpreted in two ways. Either those children with more severe asthma or comorbidities are more likely to participate, or that individuals with higher propensity to seek out health services are those who are more likely to participate. The number of primary care providers is probably capturing many factors that could affect continuity of care.

available data. Optimal full matching (Haviland et al., 2007)¹⁵ yielded 247 matched sets, matching 266 out of 270 people in the treatment group to 2,739 out of 2,742 in the control group.¹⁶ The overlap in propensity scores improved significantly, as illustrated in the second frame of Figure II.

<<Insert Figure II here>>

We also conducted t-tests for each of the pre-treatment variables selected by the DSA algorithm. Because the relationship between participants and non-participants in each matched group is not one-to-one, we construct the sum of differences between each pair of individuals (i and j) within each group and weight this by the inverse of the number of matches within that group.

$$\bar{x} = \frac{\sum_{k=1}^K (\tilde{x}_i - \tilde{x}_{j,k})}{K}.$$

The test statistics are reported in Table IV. With the exception of the number of control medications in the previous six months and number of adult family members with AAH insurance, the treatment and control groups do not differ significantly. Together, the overlap in the propensity scores of the two groups and the results of the t-tests suggest that optimal full matching has substantially reduced the bias in comparing outcomes across the two groups¹⁷.

<<Insert Table IV here>>

¹⁵ Using the optmatch package in R.

¹⁶ The matched sets' quartile case : control ratios (.25, .5, .75, 1.0) are 1:13, 1:4, 1:1, 8:1; one matched set contained eight treated participants and one control. Hence, the total number of matched sets (247) is less than the total of matched treated participants (266).

¹⁷ In addition, we wanted to confirm that our analysis of program outcomes, described in the following section, was not sensitive to the choice of matching technique. Therefore, we tried several optimal matching combinations including 1:1, 1:4, 1:k without a penalty, and optimal matching with only propensity score distances and a caliper, and repeated our regressions of treatment outcomes each time. In each case, the regressions of treatment outcomes had odds ratios of similar magnitude and direction, regardless of the matching technique. Because optimal full matching is preferred on statistical grounds, we limit our presentation to only those results, but the results are qualitatively equivalent for all matching techniques.

5. Evaluation Results

First, we use traditional bivariate analysis to measure the impact of the program on preventative care and treatment for exacerbations. Changes in preventative care are important intermediary steps to reduce the costly outcomes associated with certain consequences of asthma exacerbations (e.g., emergency room visits, hospitalization). Table V shows the results of five conditional logistic regressions for matched participant-control groups.¹⁸

<<Insert Table V here>>

Compared to non-participants, participants were more likely to have engaged in two measures of risk-reducing behaviors: vaccinations for pulmonary illness (influenza or pneumonia) and using controller medications.¹⁹ The odds ratios between participants and non-participants are 2.95 for vaccinations (1.82 to 4.81, 95% CI) and 1.53 for controller medication (1.07 to 2.19, 95% CI). These are all promising indicators, and this increase in risk-reducing behaviors should translate into a reduction in adverse outcomes (asthma exacerbations). However, out of three indicators of treatment for asthma exacerbations in the six months following the program - the number of prescriptions for rescue medications, any nebulizer treatment, and any ED visit - only one is statistically different between the two groups at the 95% confidence level. The odds of a participant having a prescription filled for a rescue inhaler are 1.49 times greater than for non-participants (1.07 to 2.07). It is unclear how to interpret the increase in prescriptions for rescue inhalers. Most children with asthma should have a rescue inhaler on hand in case of an emergency. OKA may have made more families aware of this fact

¹⁸ Note that conditional logistic regression for matched case-control groups is equivalent to fixed-effects logit for panel data; because of the health nature of the study, we will use the term most commonly used in biostatistics.

¹⁹ We decided not to consider pulmonary testing because without a detailed chart review there is no way to accurately categorize this type of encounter as being due to an exacerbation or use as a preventative measure.

and encouraged them to have an adequate number of prescriptions²⁰ - or OKA participants may have suffered a higher level of exacerbations, forcing them to refill their prescriptions more often.

Given the complex nature of asthma, the goal of a program such as OKA should not be simply to reduce consumption of specific health care services. More broadly, the goal should be to improve the overall state of a child's asthma. To reflect this priority, we created an indicator for change (or lack thereof) in asthma state after program participation. A child is considered to be in a negative asthma state if he or she has had a health care encounter to treat an asthma exacerbation or has had excessive use of rescue medication; otherwise he or she is in a stable asthma state. This definition is based on the Global Initiative on Asthma (GINA) and is similar to indices used in other studies of asthma exacerbations (see Price and Briggs, 2002 and Stempel et al., 2005). Since there are two possible states before and after participation, H_i has four possible values, as illustrated in Figure III.²¹ Our approach is conservative, because a person is coded as being in a negative state if he or she has any encounter that indicates an asthma exacerbation. We use a strict definition for asthma control because we do not want to overestimate a treatment effect.

<<Insert Figure III here>>

We then estimated a conditional multinomial logit using Proc Glimmix where the outcome is the individual's score of the index. Table VI reports the probabilities of the four possible

²⁰ For example, a child may need to have an inhaler at home and school. If prior to OKA the child did not have an inhaler at one of those places, or had an expired inhaler, then adding an inhaler after OKA is actually a risk-reducing behavior.

²¹ The four possible values are strongly and logically associated with changes in direct health care costs discussed in the previous section. Recall that the mean decrease in costs for all OKA participants was \$133. Participants who worsened had a mean *increase* of \$166, while those who improved had a mean *decrease* of \$377. Those who remained stable over both periods had a mean increase of \$2, and those who remained in the negative state had a decrease of \$65. Because these numbers are not relative to the control group, they do not reflect the impact of the program itself. However, they show the benefit to be gained by moving children from the negative to the stable state.

combinations of pre- and post-treatment states for the participants and non-participants. As previously discussed, asthma is a chronic disease with episodic symptoms. With perfect asthma control, an individual would stay in the stable state over time. In reality, asthma severity and risk reducing behaviors vary over time and individuals will move between stable and negative states. The statistics in Table VI illustrate how in the absence of an intervention a comparable group of children with asthma are distributed between stable and negative states over the two time periods relative to the intervention. The greatest share of non-participants (49%) were in a stable asthma state before intervention and stayed in that state after the intervention, and 20% of non-participants worsened over the two periods. A smaller proportion of children who participated in the intervention started and remained in the stable states (30%), and 19% of participants worsened over the periods. There was no statistically significant difference in the conditional probability of worsening (moving into a negative state given a stable state initially) between the participants (0.39, 95% CI 0.28 to 0.52) and non-participants (0.29, 95% CI 0.27, 0.32). Likewise there is no statistically significant difference between participants and non-participants in the conditional probability of remaining stable given a stable initial state. A higher proportion of participants started in a negative state than non-participants, but again the difference is not statistically significant. However, there is a statistically significant difference in the conditional probability of improving given that the initial state was negative. The conditional probability of improving was 0.39 (0.3874, 0.3922) for participants and 0.49 (0.4867, 0.5021) for non-participants.

<<Insert Table VI here>>

6. Discussion

Our evaluation of *Oakland Kicks Asthma* demonstrates the challenges of validating the case management approach. We determined that program participants are more likely than non-participants to use two types of recommended preventative health care - control medications and vaccines for respiratory infections - which could be due to greater awareness that asthma can be managed. However, we found no reduction in treatments for asthma exacerbations, and an increase in filling prescriptions for rescue medications. This increase may have been due to families updating expired prescriptions or obtaining new prescriptions to have rescue medication available at school, as suggested by discussions with OKA program managers, which would be a positive outcome, but this cannot be verified using the claims data available. A reduction in nebulizer treatments and ED visits, the other markers of asthma exacerbations, would be difficult to find in any case. First, these are uncommon events to begin with. Second, any long-term benefits of improved preventative care would only appear over a longer time frame.²² For these reasons, even a successful program could lead to an increase in utilization in the short term, before any long-term decrease in utilization becomes visible.

Alternatively a program could be effective for some segments of the target population but ineffective for others, making its impact statistically insignificant for the group of participants as a whole. Schuck and Zeckhauser (2006) characterize the inefficiencies (both economic and political) generated by not targeting social programs to those for whom the programs will do the most good.²³ However, additional research can help us better understand these differences and thereby improve programs and outcomes.

²² This limitation is evident in previous studies as well (see Evans et al., 1999; Harish et al., 2001; Krieger et al., 2005, and Morgan et al., 2004).

²³ Schuck and Zeckhauser differentiate between those individuals who derive little benefit either relative to the resources consumed or relative to others in the program (bad best) from those individuals who actually impose

It is also plausible that the program did improve some outcomes, but that we were not able to measure those changes using our administrative data. The issue of measurement is complicated because the strength of the research design is the ability to compare the participants to non-participants who are similar in all other relevant dimensions. Typically these data are taken from administrative or insurance records. Measuring other changes (e.g., quality of life) would require surveying both participants and non-participants.

Finally, the program may have faced structural challenges beyond its potential reach that prevented it from achieving its desired outcomes. One possible barrier is failure of physicians to follow best practice. Despite efforts to publicize and encourage physicians to implement existing asthma management guidelines, there is significant empirical evidence that physician compliance is low. A second potential barrier is household dysfunction that prevents effective management (Perry 2008).

7. Conclusion

The National Cooperative Inner City Asthma Study has given rise to a major wave of public investment in asthma intervention programs that use case management to affect household behaviors. We used a rigorous methodology to evaluate the *Oakland Kicks Asthma* program based on optimal matching that incorporated propensity score matching to construct a control group to compare to the group of program participants. In short, our evaluation finds that OKA has no significant effect on participants that distinguishes them from the control group, whether we look at utilization or transitions between negative and stable asthma states. This analysis raised a number of questions that it could not answer, because there are several competing

externalities on other participants (bad apples). We are primarily interested in avoiding bad bets because most of the asthma programs are focused on small groups or individuals.

explanations for not finding statistically significant changes in outcomes. Asthma programs funded by the CDC share a common approach to data collection and program evaluation – i.e., one that focuses on administrative measures rather than actual outcomes – therefore these uncertainties cannot be addressed even with the data aggregation over all of the various programs that are part of CAACP.

To determine whether this type of asthma management intervention is effective, public health researchers need to be able to rule out several competing explanations for not finding statistically significant changes in outcomes: (1) the post-intervention time period may be too short to see changes; (2) utilization data miss relevant changes; (3) even if the program has no significant impact on average, it may have a positive impact on a segment of participants; (4) structural barriers beyond the scope of the program may limit its impact; or (5) the program is not effective. Which of these possible explanations are appropriate requires a broad range of data about both participants and non-participants over an adequately long time-frame. In addition to improving the accuracy of the evaluation, these data would enable us to construct a richer model of the causal relationships between program attributes, participant behaviors, and outcomes. Then asthma intervention programs could be shaped to increase overall effectiveness and to target those participants who are most likely to benefit.

We recommend three changes to how these national programs are implemented. First, resources must be invested in collecting long-run, detailed data on both participants and non-participants thus improving the measurement of changes on the individual or household-level otherwise not captured in administrative data. Second, more research is needed to identify individual or household characteristics that increase the probability of a positive outcome due to the program. Third, the program can itself be targeted at those who are likely to derive a benefit

from participation. For example, one could design a two-stage program that provides a low-cost, group education program for any eligible household and then a second, more intensive and more costly, individualized intervention for the more committed families. This program design does reduce the total number of participants in the personalized program, but can increase the proportion that will see benefits and therefore increase the program's return on investment. Such a shift in the collection and usage of program data will both increase the public health benefits of the CDC's asthma programs and make it possible to shift resources toward those intervention models that are proven to succeed.

References

- Centers for Disease Control and Prevention. 2007. *CDC-funded asthma activities by state and type of funding: 2007 at a glance*. Available at: <http://www.cdc.gov/asthma/aag07.htm>. Accessed on May 2, 2008.
- D'Agostino, R. B. 1998. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 17 (19):2265-2281.
- Drake, C. 1993. Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics* 49 (4):1231-1236.
- Evans, R., 3rd, P. J. Gergen, H. Mitchell, M. Kattan, C. Kerckmar, E. Crain, J. Anderson, P. Eggleston, F. J. Malveaux, and H. J. Wedner. 1999. A randomized clinical trial to reduce asthma morbidity among inner-city children: results of the National Cooperative Inner-City Asthma Study. *Journal of Pediatrics* 135 (3):332-338.
- Gergen, P. J., H. Mitchell, and H. Lynn. 2002. Understanding the seasonal pattern of childhood asthma: results from the National Cooperative Inner-City Asthma Study (NCICAS). *Journal of Pediatrics* 141 (5):631-636.
- Hansen, B. B., and S. O. Klopfer. 2006. Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics* 15 (3):609-627.
- Harish, Z., A. C. Bregante, C. Morgan, C. S. J. Fann, C. M. Callaghan, M. A. Witt, K. A. Levinson, and W. B. Caspe. 2001. A comprehensive inner-city asthma program reduces hospital and emergency room utilization. *Annals of Allergy Asthma & Immunology* 86 (2):185-189.
- Haviland, A., D. S. Nagin, and P. R. Rosenbaum. 2007. Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological Methods* 12 (3):247-267.

- Johnson, A., G. Berg, E. Fleegler, and M. Sauerbrun. 2005. A matched-cohort study of selected clinical and utilization outcomes for an asthma care support program. *Disease Management* 8 (3):144-154.
- Krieger, J. W., T. K. Takaro, L. Song, and M. Weaver. 2005. The Seattle-King County Healthy Homes Project: A randomized, controlled trial of a community health worker intervention to decrease exposure to indoor asthma triggers. *American Journal of Public Health* 95 (4):652-659.
- Moore, K., R. Neugebauer, F. Lurmann, J. Hall, V. Brajer, S. Alcorn, and I. Tager. 2008. Ambient ozone concentrations cause increased hospitalizations for asthma in children: An 18-year study in Southern California. *Environmental Health Perspectives* In press. [Available online 6 March 2008].
- Morgan, W. J., E. F. Crain, R. S. Gruchalla, G. T. O'Connor, M. Kattan, R. I. Evans, J. Stout, G. Malindzak, E. Smartt, M. Plaut, M. Walter, B. Vaughn, and H. Mitchell. 2004. Results of a home-based environmental intervention among urban children with asthma. *New England Journal of Medicine* 351 (11):1068-1080.
- National Institutes of Health. Feb 5, 2001. Grants awarded to implement Inner-City Asthma Intervention. *NIH News*. Available at: <http://www3.niaid.nih.gov/news/newsreleases/2001/asthmagrants.htm>. Accessed on May 7, 2008.
- Perry, C. D. 2008. Does treating maternal depression improve child health management? The case of pediatric asthma. *Journal of Health Economics* 27 (1):157-173.
- Price, M. J., and A. H. Briggs. 2002. Development of an economic model to assess the cost effectiveness of asthma management strategies. *Pharmacoeconomics* 20 (3):183-194.
- Rosenbaum, P. R. 1989. Optimal matching for observational studies. *Journal of the American Statistical Association* 84 (408):1024-1032.
- Rosenbaum, P. R. 1991. A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society Series B-Methodological* 53 (3):597-610.

- Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1):41-55.
- Rosenbaum, P. R., and D. B. Rubin. 1985. Constructing a control-group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39 (1):33-38.
- Rubin, D. B., and N. Thomas. 2000. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* 95 (450):573-585.
- Schatz, M., R. S. Zeiger, W. M. Vollmer, D. Mosen, A. J. Apter, T. B. Stibolt, A. Leong, M. S. Johnson, G. Mendoza, and E. F. Cook. 2006. Development and validation of a medication intensity scale derived from computerized pharmacy data that predicts emergency hospital utilization for persistent asthma. *American Journal of Managed Care* 12 (8):478-484.
- Schuck, P. and R. Zeckhauser. 2006. *Targeting in social programs*. Washington, D.C.; Brookings Institute.
- Stempel, D. A., T. P. McLaughlin, R. H. Stanford, and A. L. Fuhlbrigge. 2005. Patterns of asthma control: a 3-year analysis of patient claims. *Journal of Allergy and Clinical Immunology* 115 (5):935-939.
- Stockman, J., J. Von Behren, N. Shaikh, and R. Kreutzer. 2003. California asthma facts. *Environmental Health Investigations Branch, California Department of Health Services*. Available at: <http://www.ehib.org/papers/CaliforniaAsthmaFacts010503.pdf>. Accessed on May 2, 2008.
- U.S. Census Bureau. 2006 *American Community Survey, Oakland, CA*. Available at: <http://factfinder.census.gov>. Accessed on April 2008.

Figure I

Identification of treatment and control groups

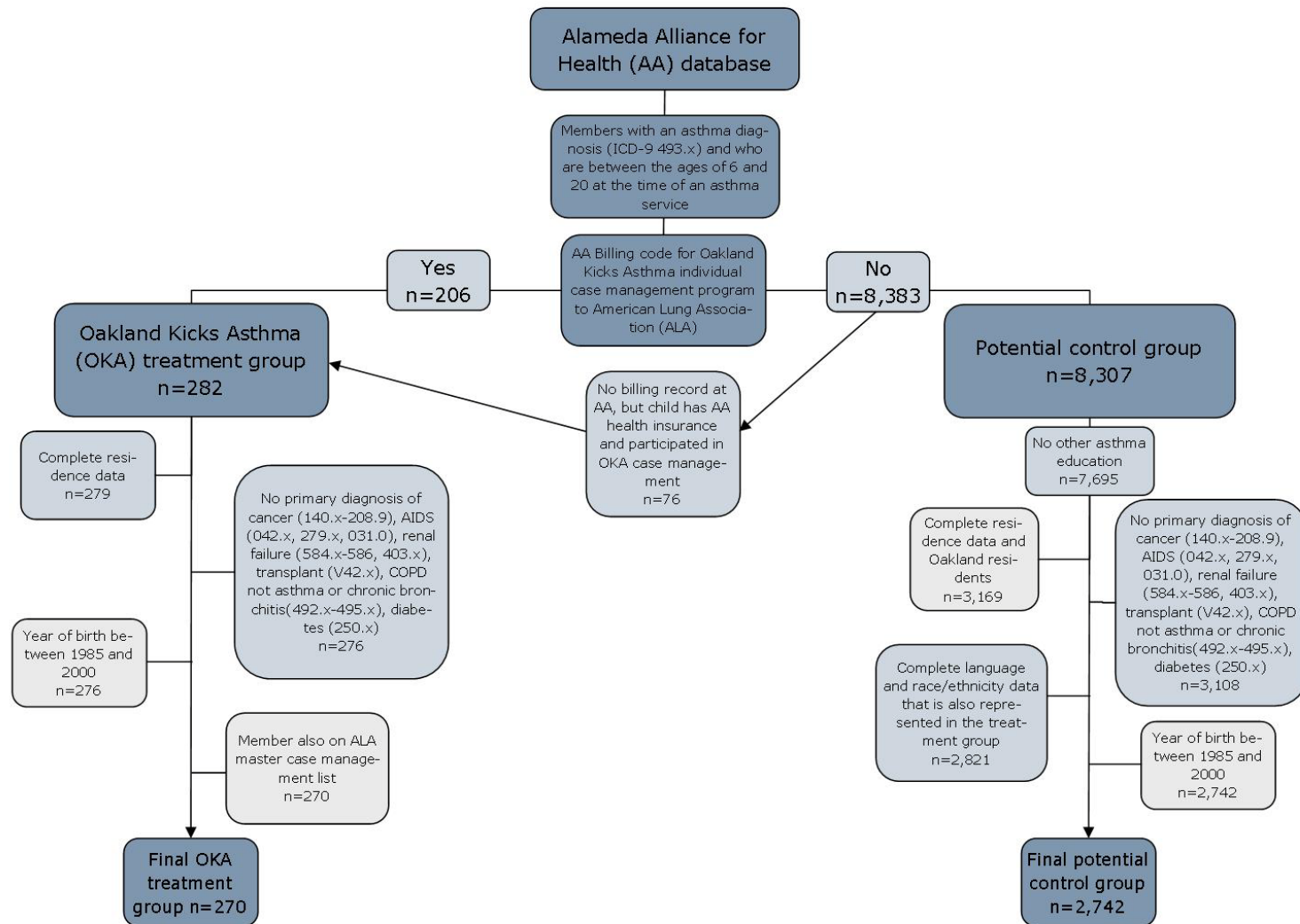


Figure II: Box plot of propensity score before and after matching

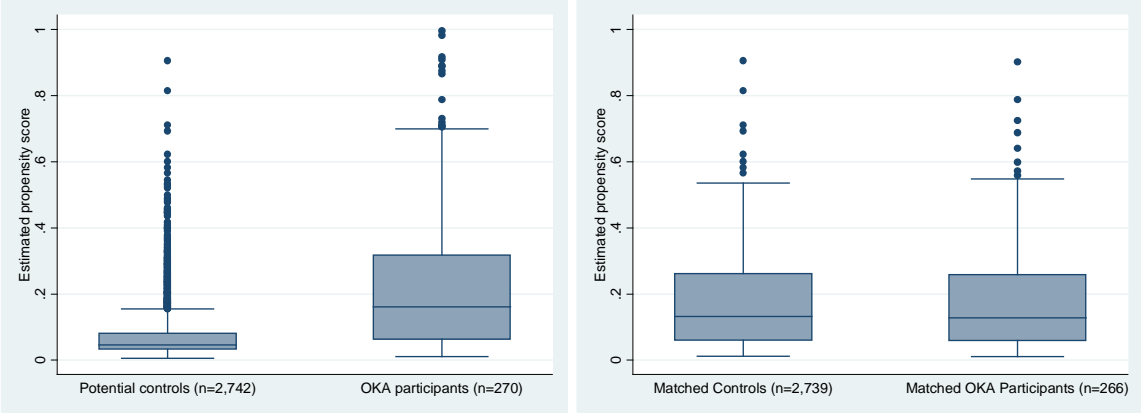


Figure III: Transitions between asthma states

	Post intervention	
Prior intervention	<u>Stable state post</u>	<u>Negative state post</u>
Stable state	No change stable (H=1) <i>Reference group</i>	Worsened (H=2)
Negative state	Improved (H=3)	No change negative (H=4)

Table I. Demographics of study participants (n=3,012)

	Treatment group n (%)	Control group n (%)
Race/ethnicity		
Black	139 (51)	1,515 (55)
Hispanic	96 (36)	736 (27)
Asian	35 (13)	491 (18)
Language		
English	155 (57)	1,778 (65)
Asian group	25 (9)	371 (14)
Spanish	88 (33)	593 (22)
Missing/unknown	2 (1)	0 (0)
Gender		
Male	147 (54)	1,446 (53)
Female	123 (46)	1,296 (47)
Age group at study start*		
2-11	103 (38)	1,234 (45)
12-16	148 (55)	1,024 (37)
17-22	19 (7)	484 (18)
Allergy diagnosis		
No	103 (38)	1,664 (61)
Yes	167 (62)	1,078 (39)
N	270	2,742

*The treatment group age range is 6-19, while the potential control group range is 2-22. Participants in both groups were born between 1985 and 2000, but the control group had a greater age at the study start due to the random assignment of start dates. Age was not selected by the DSA for being associated with OKA participation.

**Table II: Health care use by treatment and control groups 1 year prior to enrollment
HEDIS adapted variables**

Asthma-related health care	Treatment group n (%)	Control group n (%)
Emergency department or urgent care visits		
0	210 (78)	2,568 (94)
≥1	60 (22)	174 (6)
Inpatient hospitalization		
0	255 (94)	2,709 (99)
≥1	15 (6)	33 (1)
Outpatient or MD office visits		
0	108 (40)	1,831 (67)
≥1	162 (60)	911 (33)
Flu or pneumonia vaccine		
0	226 (84)	2,509 (92)
1 or 2	43 (16)	233 (8)
Prednisone prescriptions		
0	218 (81)	2,545 (93)
≥1	52 (19)	197 (7)
Pulmonary tests		
0	196 (73)	2,464 (90)
≥1	74 (27)	278 (10)
Rescue prescriptions		
0	61 (23)	1,468 (54)
1-8	175 (65)	1,208 (44)
>8	34 (13)	66 (2)
Control prescriptions		
0	104 (39)	1,997 (73)
1-2	65 (24)	403 (15)
>2	101 (37)	342 (12)
N	270	2,742

Table III. Propensity score model (with OKA participation as the outcome)

Pre-treatment characteristics	Odds Ratio, Standard Error, [95% CI]
Utilization and prescriptions	
# of control medications in the previous 6 months*	1.14, 0.04, [1.06, 1.22]
Any control medication in the previous year*	1.55, 0.27, [1.10, 2.19]
# of ED or urgent care visits for asthma in the previous 6 months (squared)	4.51, 1.90, [1.97, 10.30]
# of ED or urgent care visits for asthma in the previous 6 months (cubed)	0.56, 0.12, [0.36, 0.86]
# of outpatient visits for asthma in the previous year* (squared)	1.02, 0.01, [1.01, 1.03]
Interaction with health care providers	
# different primary care physicians listed for a child in previous 6 months	5.89, 2.87, [2.27, 15.31]
# different primary care physicians listed for a child in 6 months (squared)	0.39, 0.17, [0.17, 0.90]
Individual ever had an allergy diagnosis	1.75, 0.25, [1.32, 2.31]
Socio-economic and demographic	
# adult family members with AAH health insurance	0.76, 0.07, [0.64, 0.91]
% households in a study participant's census tract with income between 25-50k** (cubed)	1.00, 0.00, [1.00, 1.00]

*HEDIS defined variable

**Rounded to the nearest hundredth

Table IV. T-test results of covariate balance based on weighted distribution among matched sets

Variables	Mean difference	t-statistic
# outpatient visits for asthma in the previous year* (squared)	-0.45	-1.01
# ER or urgent care visits for asthma in the previous 6 months (squared)	0.00	-0.02
# control medications in the previous 6 months*	-0.13	-2.18
# different primary care physicians listed for a child in previous 6 months	0.02	1.17
# adult family members with AAH health insurance	-0.09	-1.83
Allergy diagnosis (proportion)	0.02	0.77
# different primary care physicians listed for a child in 6 months (squared)	0.02	0.89
% households in a study participant's census tract with income between 25-50k (cubed)	331.22	1.08
Any control medication in the previous year*	-0.04	-2.51
# ER or urgent care visits for asthma in the previous 6 months (cubed)	0.00	0.03

*HEDIS defined variable

Table V. Bivariate odds ratios for outcomes from conditional logistic regression after OKA case management treatment

Outcome	Observations (n), matched sets	OR	Likelihood Ratio	90% CI	95% CI	Standard Error
Preventative Care						
# of control medication 6 months post	2359, 172 sets	1.53	-761.28	[1.13, 2.07]	[1.07, 2.19]	0.28
Any flu or pneumonia vaccine 1 year post	1507, 110 sets	2.95	-316.73	[1.97, 4.45]	[1.82, 4.81]	0.73
Treatment of exacerbation						
# of rescue medication 6 months post	2511, 183 sets	1.49	-1154.9	[1.13, 1.96]	[1.07, 2.07]	0.25
Any nebulizer administration 6 months post	1462, 66 sets	1.93	-230.84	[1.09, 3.42]	[0.98, 3.81]	0.67
Any ED or urgent care center visit for asthma 6 months post	1501, 70 sets	1.73	-259.98	[1.00, 3.00]	[0.90, 3.33]	0.58

Table VI: Probabilities of asthma states

Prior to intervention → Post intervention	Participants (95% CI)	Non-participants (95% CI)
Stable state → Stable state	0.30 (0.18, 0.45)	0.49 (0.44, 0.54)
Stable state → Negative state	0.19 (0.17, 0.20)	0.20 (0.20, 0.20)*
Negative state → Negative state	0.31 (0.23, 0.38)	0.16 (0.13, 0.19)
Negative state → Stable state	0.20 (0.15, 0.24)	0.15 (0.13, 0.18)

* Due to rounding.

These probabilities are derived from the conditional multinomial logit estimated using Proc Glimmix in SAS.

Appendix

Table A1: Utilization measures

Utilization measures	ICD-9 procedure codes	Binary Only
Inpatient admissions (HEDIS defined*)	99221-99223, 99231-99233, 99238-99239, 99251-99255, 99261-99263, 99291-99292, 99356-99357	
Emergency Department (HEDIS defined)	99281-99285, 99288	
Outpatient visits (HEDIS defined)	99201-99205, 99211-99215, 99217-99220, 99241-99245, 99271-99275	
Inpatient admission	HEDIS codes plus 99235- 99236	
Treatment codes	90772, 90774, 90782, 94002-94004, 94640, 94642, 94644-94645, 94660, 94662, 94665, 31500, 83735, 4025F, J1100, J1030, J7506, J7510, J7611, J7613-J7614, J7616, J7618- J7621, J7625- J7626, J7644, X5528, X5530, X5554 , X6004, X6008, X6014, X6036, X6038, X6328, X6580, J7510, E0570, E0570	
Inpatient treatment	Any treatment code combined with inpatient hospital as place of service	
Emergency Department or Urgent Care visit	HEDIS codes plus 90760, 96150-96151, 99062, 99064-99065, 99236, , 99382-99384, 99394-99393, 450, Z7502, Z7610	
Emergency Department or Urgent Care treatment	Any treatment code combined with emergency room or urgent care center as place of service	
Outpatient or physician's office visit	94664, 96150-96151, 99234, 99348, 99371-99374, 99382-99385, 99392-99395	
Outpatient or physician's office treatment	Any treatment code combined with physician office or outpatient center as place of service	
Flu or pneumonia vaccine	90655-90660, 90669, 90724, 90732	
Nebulizer use	94640	
Pulmonary testing	92551-92552, 94010-94016, 94060, 94070, 94150, 94160, 94200, 94375, 94620-94621, 94680, 94681, 94690, 94750, 94760-94762, 94799, 95012, 95070, 95071, 82803, 82805, 31622, 31646, 3023F, 3025F, 3027F, 3028F, 3035F, 3037F, 71010-71555, A4614,	
Home medical equipment	9944A, E1390, E0570, E0431, A9900, A7003, E0550, A7005, A7004	X
Allergy testing	94200, 94240, 95004, 95010, 95015, 95024, 95027, 95044, 95070, 95075, 86001, 86003, 86005, 0168T	X
Allergy immunotherapy	95115, 95117, 95144-95149, 95165, 95170, 95199	X
Other asthma education (not OKA)	96150, 96151, 96155	X
High severity visits to outpatient facility or doctor's office	99354, 99355, 99058, 99060, 99220, 99235, 99236	
Ambulance service	X0036, A0800, 99289, X0030, A0434	

* Source: Healthcare Effectiveness Data and Information Set (HEDIS).

Note: We felt that the HEDIS criteria were overly restrictive for our purposes, and we relaxed the HEDIS criteria to generate a second set of amended-HEDIS measures for ED visit, inpatient admission, and outpatient/doctor's office

visit. Our adapted HEDIS definitions generally were more inclusive and contained more procedure and point of service codes than HEDIS. With the exception of home medical equipment, allergy testing, allergy immunology, and participation in asthma program other than OKA, variables were defined as both a binary measure of any encounter (0= no encounter, 1=any encounter) and a count of encounters. All variables are measured at 6 month and one year intervals.

Table A2: Medication intensity scale

Rescue prescriptions 1 year prior to enrollment for eligible cases and controls	OR for an ED visit 1 year after treatment	SE	95% CI
>0	1.18	0.20	[0.85, 1.64]
>1	1.25	0.22	[0.88, 1.76]
>2	1.17	0.24	[0.79, 1.75]
>3	1.17	0.28	[0.74, 1.85]
>4	1.19	0.32	[0.70, 2.01]
>5	1.33	0.38	[0.76, 2.33]
>6	1.71	0.51	[0.96, 3.06]
>7	1.84	0.61	[0.96, 3.53]
>8	2.64	0.90	[1.35, 5.13]
>9	2.70	1.01	[1.29, 5.62]
>10	2.97	1.18	[1.36, 6.50]
>11	2.33	1.15	[0.89, 6.12]
>12	2.71	1.35	[1.02, 7.19]
>13	2.37	1.50	[0.69, 8.18]
>14	2.88	1.85	[0.82, 10.14]
>15	4.50	3.02	[1.20, 16.78]

This is an application of Schatz et al.'s work (2006) to estimate categories of medication use where there is a change in the odds of an ED visit. We created variables in Table II for rescue medication use (0, 1-8, and >8) where there were larger OR differences. We wanted to differentiate between children who were properly using their rescue medication and those who were not. We also included these categories in the DSA to model program participation.