ECHO AUTISM
Statistical Analysis Plan

1 Signature Page

_____

Kristin Sohl, MD
Thompson Center for Autism and other Neurodevelopmental Disorders,
University of Missouri
Study Chair

_____

Micah Mazurek, PhD
University of Virginia
Study Co-Chair

_____

Robert A. Parker, ScD
MGH Biostatistics Center

Harvard Medical School
Study Statistician

2 Table of Contents

3 Abbreviations and Definitions

AIR-P	Autism Intervention Research Network on Physical Health
ASD	Autism Spectrum Disorder
ATN	Autism Treatment Network
DCC	Data Coordinating Center
ECHO	Extension for Community Healthcare Outcomes
HRSA	Health Resources and Services Administration
IRB	Institutional Review Board
PCP	Primary Care Provider (includes non-physicians e.g. Nurse Practitioners)
SAP	Statistical Analysis Plan
UNM	University of New Mexico

4 Overview

4.1 Brief Description of Study

ECHO Autism was intended to assess rigorously the impact of a 12-session telemedicine training program on participating PCP's knowledge, clinical behavior, and self-efficacy in the screening and care of children with ASD. Each session is referred to as an "ECHO clinic".

The study involved 10 sites (each referred to as an "ECHO Autism Hub"), each running a 12-session training program using a common curriculum and core lecture, with each site expected to recruit 15 PCPs. Sites were randomized in a stepped-wedge design with 5 clusters (2 sites per cluster) and a staggered start over a 1-year period. Staggering the start allowed for some control for potential temporal trends, as well as allowing the core team to focus on working with each site to ensure smooth startup of the training program at each site.

Outcomes are measured at baseline (T1), during the intervention (T2, approximately 3 months after the start of the intervention), and after the end of the intervention (T3). An

additional measurement (T4) was made 3 months after the end of the intervention to assess whether deterioration occurs after clinic participation ends.

All PCP participants received the intervention.

IRB approval was obtained by each ECHO Autism Hub and the central Data Coordinating Center (DCC).

The study was funded by the Autism Intervention Research Network on Physical Health (AIR-P) which is supported by the Health Resources and Services Administration (HRSA).

4.2 Scope of Analysis

This is intended to be the SAP for the primary analysis for the study. Thus, this document includes:

- primary analysis for each study endpoint; and
- pre-defined sensitivity and exploratory analyses of study endpoints.

In the case of discrepancies between this document and the study protocol, this document has priority on all issues related to the analysis of the study.

Revisions to the SAP in this amendment were made to (a) clarify specific analysis issues; (b) allow for contingencies when there was concern that primary analyses were not possible (e.g. convergence issues of models); and (c) to correct previous documentation errors.

5 Study Objectives and Outcome Measures

5.1 Study Objectives

To determine whether participation in a collaborative telehealth intervention will result in improved learning, clinical practice behavior and efficacy among primary care providers (PCPs).

Hypothesis 1: Following participation in ECHO Autism, PCPs will demonstrate significant improvements in ASD knowledge as assessed by pre- to post-intervention knowledge tests in ASD screening and identification and assessment and treatment of medical co-morbidities.

Hypothesis 2: Following participation in ECHO Autism, PCPs will demonstrate significant improvements in clinical practice/behavior as assessed by pre- to post-intervention chart reviews in ASD screening (co-primary outcome) and treatment of medical co-morbidities, in particular, sleep problems and constipation (co-primary outcome).

Hypothesis 3: Following participation in ECHO Autism, PCPs will demonstrate significant improvements in self-efficacy in ASD screening and identification and treatment of medical co-morbidities.

5.2 Outcome Measures

The primary outcome measures for the study, as specified in Hypothesis 2 are:

- clinical practice / behavior based on
  - screening of children at well child visits, described in Section 6.3.1.1; and
  - treatment of co-occurring medical conditions at visits of children with ASD, described in Section 6.3.1.2

The specified secondary endpoints are:

- an ASD Knowledge quiz (Hypothesis 1), described in Section 6.3.2.1; and
- a self-efficacy assessment (Hypothesis 3), described in Section 6.3.2.2.

Although not formally stated in a hypothesis, an outcome measure related to treatment of co-occurring medical conditions in children with ASD is:

- number of co-morbidities reported in children with ASD, described in Section 6.3.3.

Although not formally stated in a hypothesis, an outcome measure related to self-efficacy is:

- perceived barriers to care for children with ASD, described in Section 6.3.4.

6 Study Methods

6.1 Inclusion-Exclusion Criteria and General Study Population

6.1.1 Participant Inclusion Criteria

All the following inclusion criteria must be met for a participant for the study:

- Current practice as a primary care provider (PCP).
- Currently providing care for children.
- Professional training in: general pediatrics, family medicine, advance practice nursing (i.e. nurse practitioner or physician assistant).
- Active medical license in the state of practice.
- Patient population is at least 50% underserved.

6.1.2 Participant Exclusion Criteria

Any of the following would preclude an individual from participating in the study:

- Trainee status (e.g., medical student, intern, resident, or other pre-professional trainee).
- Subspecialist (e.g., psychiatrists, neurologists, developmental and behavioral pediatricians).
- Practicing within the same practice as another PCP Participant (i.e., only one PCP participant from any given practice may be enrolled as a research participant in the study).

6.2 Schedule of Assessments

After informed consent, PCP participants provided demographic and practice information (Section 6.4.1) before the start of the study.

Each PCP participant was supposed to complete the battery of provider-completed measures at four timepoints: Baseline/Pre-Intervention (T1), Mid-Intervention (T2), Post-Intervention (T3), and Follow-up (T4). The duration of the ECHO intervention will be 6 months. The target time point for the T2 assessment is between the 6th and 7th ECHO sessions. The T3 assessment should occur within 4 weeks of completion of the final ECHO session. The Follow-up assessment (T4) should be conducted between 9 and 10 months after the start of the ECHO program. Provider completed measures at these time points are: (a) ASD Knowledge Quiz (Section 6.3.2.1); (b) ASD Self-Efficacy (Section 6.3.2.2); and (c) Perceived Barriers to Care for Children with Autism in Primary Care (Section 6.3.4).

Chart reviews were planned to be done in the same time frame for T1, T3, and T4. Because it would not be feasible to do the chart review in the two weeks for 15 participants, the T2 review will include charts from the 30 or 60 days before the 7th ECHO session for all participants, or earlier if the clinic is visited before the 7th ECHO session.

PCP participants answer a satisfaction questionnaire (Section 6.4.3) at the end of the training program.

6.3 Study Assessments: Outcome Measures

6.3.1 Primary Endpoints: Clinical Practice / Behavior

6.3.1.1 Screening Practice (co-primary endpoint)

Clinical Practice/Behavior was assessed at T1, T2, T3, and T4 by review of a subset of charts from each PCP's practice. Four subsets of charts will be reviewed, with a limit of 25 charts in any group. The groups are:

1. Charts for all children seen for 9-month well-child visits in the 30 days prior to the date of chart review.
2. Charts for all children seen for 18-month well-child visits in the 30 days prior to the date of chart review.
3. Charts for all children seen for 24-month well-child visits in the 30 days prior to the date of chart review.
4. Charts for all children seen for 30-month well-child visits in the 30 days prior to the date of chart review.

If more than 25 well-child visits at a specific age are available for chart review, the most recent 25 well-child visits at a specific age will be reviewed.

Because of the timing and feasibility of doing all chart-reviews in the 2-week interval between the 6th and 7th ECHO clinics, the 30-days was either (a) the 30-days prior to the date of the 7th ECHO clinic; or (b) the 30-days prior to the date of the visit scheduled for chart reviews, if the visit occurred prior to the 7th clinic.

These chart reviews assess the adequacy of screening for each child at each visit. The screening practice is summarized over the four sets of charts as total number screened appropriately / total number of charts reviewed and then converted to a percentage.

For the 9 US sites, adequate screening, as defined by US guidelines consider the use of any general developmental screening tool as appropriate screening for the 9- and 30-month visits. For the 18- and 24-month visits, an ASD specific screen must have been used for the child to be considered correctly screened for Autism.

A different guideline is used in Canada, so that adequate screening was defined differently for the Canadian site. The recommended screening practice in Canada uses a general developmental screening tool at 12- and 18-month well-child visits. Only visits at those times were reviewed at the Canadian site, and children were considered appropriately screened if a general developmental screening tool was administered.

For analysis purposes, the results of each individual chart reviewed (screened or not screened appropriately) is used in the analysis rather than the summary over all charts for a PCP.

PCPs having no well-child visits at baseline would have baseline results imputed if appropriate (Section 8.4.1).

6.3.1.2 Treatment of Co-morbidities in Children with ASD (co-primary endpoint)

Clinical Practice/Behavior was assessed at T1, T2, T3, and T4 by review of charts for all children with ASD in the 60 days prior to the data of the chart review.

The score is based on treating reported conditions appropriately. Any of a range of treatments was considered appropriate for each condition. Charts without any reported conditions are excluded in this analysis. The total number of appropriately treated conditions among the total number of conditions reported is converted to a percentage. PCPs having no ASD visits at baseline would have baseline results imputed if appropriate (Section 8.4.1).

6.3.2 Prespecified Secondary Endpoints

6.3.2.1 ASD Knowledge

ASD knowledge was assessed at T1, T2, T3, and T4 using a 33-item test developed specifically for the current study. The original test was developed and piloted with a group of 14 PCP participants, questions with very low difficulty were removed and/or reworded (e.g., if at least 90% of participants answered correctly at pre-test), and additional questions were included to ensure that all content was adequately covered. The revised version was then piloted in a second sample of nine PCPs. The test assesses knowledge in the areas of ASD screening/identification, psychiatric co-morbidities, medical co-morbidities, and management of additional ASD-specific needs.

This test is scored (maximum of 1 point per question) and the score is then converted to a percent, with 100% representing no errors. Scoring is based on the total number of

correct answers, among all 33 questions. Any missing answers are counted as incorrect responses.

There are several specific considerations for the scoring:

- Questions 24, 25, 28, and 32 ask to check all answers that apply: these are scored as zero if any incorrect option is selected, and if no incorrect options are selected then each correct selection is given an appropriate fraction of a point, e.g. for question 28 there are three answers (out of the four options) which should be selected, if the incorrect option is not selected each correct option is given 1/3 of a point;
- An ambiguity was identified in question 7 after the study started, and either possible answer (annually, as the child has already been on the medication for three years; or baseline, 6 months, annually, which was the intended answer for the schedule of testing from initial use) is scored as correct; and
- Because of differences in Canadian and US screening practices, questions 22, 23, and 25 are not scored for the Canadian site, and the percent is based on the total score divided by 30.

The baseline data suggested that the overall score ranged between 20-80% indicating that this can be analyzed as a continuous variable.

6.3.2.2 ASD Self-Efficacy

ASD self-efficacy was assessed at T1, T2, T3, and T4 using a questionnaire developed for a previous ECHO Autism pilot study. The questionnaire is comprised of 57 items across five domains: 1) ASD screening and identification (7 items), 2) ASD referral and resources (9 items), 3) assessment and treatment of medical comorbidities (19 items), 4) assessment and treatment of psychiatric comorbidities (13 items), and 5) additional (9 items). Participants report the degree to which they are confident in their ability to provide effective care in each domain. Items are rated on a 6-point Likert-type scale (ranging from 1= “no confidence” to 6 = “highly confident/expert”).

Items are recoded to a 0 ("no confidence") to 5 ("highly confident/expert") and then summed for a total score and the five sub-scale scores. These scores are then normalized to a percentage, by dividing by 5 x number of relevant items (57 for the total scale, 7, 9, 19, 13, and 9 for the subscale if no missing data).

A subscale score is set to missing if more than 20% of the questions in that subscale are not answered. The total score is set to missing if 6 or more questions are missing or if any subscale score is missing. Only the total score of the ASD self-efficacy scale will be examined in this analysis.

The structure of the self-efficacy questionnaire means that the total score is the number of points achieved out of a total of 285 possible points. This number is high enough that it would be safe to consider this a normally distributed variable.

6.3.3 Additional Outcome Measure Related to Treatment of Comorbidities: Number of Reported Co-morbidities in Children with ASD

Related to the number of co-morbidities correctly treated, but not stated in a formal hypothesis, is the number of reported co-morbidities needing treatment in children with ASD. The summary measure is defined as the number of comorbidities reported among the 4 possible comorbidities for a child.

For analysis purposes, the number of co-morbidities reported for each chart reviewed will be used.

PCPs having no ASD visits at baseline have baseline results imputed if appropriate (Section 8.4.1).

6.3.4 Additional Outcome Measure Related to Self-Efficacy: Perceived Barriers to Care for Children with Autism in Primary Care

This was assessed at T1, T2, T3, and T4 by participant response to a 9-item checklist with two additional open-ended questions. Descriptive analysis will give the proportion of participants with each specific barrier at each time point and summarize the additional questions.

For analysis purposes, the total number of specific barriers checked will be calculated for each PCP. The first of the open-ended questions ("Other/specify") will be included as a 10^{th} barrier in the count of number of barriers checked.

6.4 Additional Measures

6.4.1 Demographic and Practice Information

This Information was collected at baseline using a demographic questionnaire. Providers report the following information: age, gender, race, ethnicity, zip code of practice, patient population (volume, patient characteristics), years of practice, provider type, and previous training in ASD.

6.4.2 Amount of Training

This was abstracted from the CME sign-in sheets for all participants. The amount of training will be analyzed as the percentage attendance for available sign-in sheets.

6.4.3 Satisfaction

This was assessed at completion of the intervention (T3) using a 12-item survey developed for a previous ECHO Autism pilot study. The survey includes 10 questions assessing overall satisfaction with participation in the ECHO Autism clinic (rated on a 5-point Likert-type scale, 1="Strongly agree" to 5="Strongly disagree"), and two questions asking for overall comments and suggestions.

Overall benefit of the program is defined as the proportion of participants who answer question 1 ("Participation in ECHO Autism improved my ability to care for children with autism in my practice") with a response of "Strongly Agree" or "Agree".

6.4.4 Intervention Fidelity Evaluations

The fidelity of the ECHO intervention was assessed using a 25-item observer-rated form assessing fidelity of implementation including: training flow, facilitator engagement of participants, and other indicators of adherence. Each item is measured on a 5-point Likert scale from 1="Strongly disagree" to 5 = "Strongly agree". The measure was developed by the UNM ECHO Team to ensure that facilitators adhere to the model. Fidelity will be assessed at 2 randomly selected Clinics for each ECHO Autism Hub.

Fidelity of an ECHO clinic will be determined as the percent of questions answered as "agree" or "strongly agree" among all the questions completed. Questions which were not completed are ignored in this calculation as the auditor considered some items not applicable on a specific call. Adequate fidelity is defined to be 80% or higher fidelity.

6.5 Data Monitoring through Study

6.5.1 Routine Monitoring and Quality Control

The study data was subject to routine checks on a monthly basis. Each month, query reports were run on the data and individual query reports were sent to sites. Query reports noted incomplete or missing forms, illogical or inconsistent data, range checks for values and dates, and other issues as needed. Sites were required to address all queries on the report and return them to the DCC data manager in a timely manner and resolve issues with the DCC data manager as necessary. Additionally, monthly completion reports were generated through the course of the study, which noted the number of forms that were complete, incomplete, and missing at each site. This was used to monitor data collection progress during the study. A report tracking enrollment was also generated on a monthly basis. Data monitoring activities were unblinded, but there was no systematic summary of results and data was used only to monitor for quality issues and track data collection.

6.5.2 DSMB Monitoring

The study was monitored by the standing AIR-P network DSMB annually. The DSMB recommended continuation of the study without modification at each annual review.

7 Sample Size

When the study was planned, the sample size was determined based on feasibility / practicality considerations, given the number of sites (10) and what was considered a feasible enrollment target (15 PCPs per site) for the program with an expectation of minimal dropout. Based on these assumptions, results of simulations suggested that the study would have a reasonable chance of detecting clinically important changes and was described in the protocol as follows:

> Given the complexity of the proposed analysis, power calculations were based on simulations. The data generation process allowed for random effects for center, PCP within center, and nominal period. There was no time trend in the

data, although a potential time trend as a fixed effect was included in the model. Simulations were done for 10 randomly selected seeds (from several different websites and different random number tables), 1000 simulations per seed. The data generating process allowed for approximately a 50% intra-class correlation for the PCP within group effect, reflecting the possibility that the impact of ECHO would be correlated within each center, even with good fidelity to the intervention program. Simulations allowed for varying numbers of patients per PCP practice.

If there are on average 5 patients per PCP (e.g. 5 autistic patients seen in the last 60 days), we would have over 90% power to detect an increase of 15% in appropriate co-morbidity management (alpha=0.025, two-sided). If there are 15 patients per PCP on average (e.g. 15 patients with well child visits in the past month), we would have over 90% power to detect an increase of 10% in autism screening (alpha=0.025, two-sided). If the number of patients per PCP was higher, then we would have over 90% power for even smaller differences. Results were consistent for the different seeds.

8 General Analysis Considerations

8.1 Statistical Standards

8.1.1 Statistical Software

All analyses will be done in the latest version of SAS available at the Massachusetts General Hospital Biostatistics Center, the Data Coordinating Center for the AIR-P and graphs prepared in R.

The version currently available at time of SAP preparation is version 9.4 for SAS and version 3.5 for R.

8.1.2 Reporting and Scoring Conventions

Given the planned sample size (150 participants), percent will be reported in whole numbers (no decimal places), rounded as needed.

For continuous variables, results will be reported to the same precision as the raw data, generally without decimal places.

Scoring instructions have been given for each instrument separately above.

8.1.3 Summary Statistics

Summary statistics for categorical variables will be counts and percents. The percents for valid responses will be based on non-missing responses (e.g. if the variable has responses for Yes and No, the total of the two percents will be 100 even if some data is missing). The percent for missing data will be calculated based on the total number of responses to that questionnaire.

8.1.4 Basic Statistical Analyses

8.1.4.1 Analyses Between Groups

All statistical analyses will be adjusted for site.

Bivariate statistical analyses (e.g. between completers and non-completers, Sections 9.4 and 9.5) will use Van Elteren's test for continuous variables and Cochran-Mantel-Haenszel statistics for categorical variables.

8.1.4.2 Analyses for Differences by Site

This analysis will use Kruskal-Wallis test for continuous variables and Fisher's exact test (extension) for categorical variables.

8.1.5 Blinded Data Review

As all participants were known to be receiving the intervention, blinded data review was not possible. No summary of data collected after baseline was done except as noted in Section 12 (operational issues). No analyses of data after baseline were done until after the initial version of the SAP was finalized.

8.2 Analysis Populations

There will be two analysis populations for the data collected from baseline through six months:

a) an **efficacy analysis** will be limited to the group of PCPs who complete the six-month training program (**completers**). Completion is defined by:

i) having chart reviews completed at six months; and

ii) completing at least one of the participant surveys at six months.

Note that completion does not imply that the PCP attended a specific minimum number of sessions.

b) an **effectiveness analysis** will use the **total population** enrolled without exclusions.

There will be a single analysis population for analyses involving data collected at nine months:

c) long-term impact at 9 months will be assessed using the completers group.

8.3 Covariates and Subgroups

Exploratory analyses will assess the impact of:

- demographic variables on outcome
- practice information on outcome
- fidelity of intervention on outcome classified as sites with all clinics meeting the fidelity standard in Section 6.4.4 vs. sites with one or more clinics failing to meet the fidelity standard;
- amount of training on outcome

No formal summary by subgroups is planned.

8.4 Missing Data Imputation

8.4.1 Imputing Missing Data for Baseline Measures

Since baseline values are included in the basic modeling analysis of the study (Section 10.3), baseline data for a primary outcome measure (listed in Section 6.3.1) for a PCP will be imputed using multiple imputation if there is data at three- and six-months for the outcome measure for that PCP.

Such missing data would occur if there are no well-child visits or no ASD child visits at baseline for a PCP. Imputation for a primary outcome will only be done if it allows us to include at least 5% more individuals in the analysis for a primary outcome measure. Several PCPs had very low volume offices, so that before embarking on multiple imputation we are requiring that there be sufficient information to be gained to make the additional complexity worthwhile.

Data will be imputed from the distribution of site specific values of the baseline data. Imputation will not be done for other measures.

8.4.2 Imputing Data for Missing Three and Six Month Visits

For the missing three- and six-month time visits we will use baseline data, if available. We recognize that this is an extremely conservative assumption and biases the effectiveness but believe it is appropriate in this project for several reasons.

1. Little improvement is likely for participants dropping out before the midpoint visit because of minimal training. Thus, the baseline would be a reasonable estimate of practice at the missing time points.

2. It is less clear whether there would be deterioration (if a midpoint value is available) for a participant discontinuing the program between the three- and six-month visits. Data review found that only about 3% of the participants discontinued after the three-month chart review/survey and before the six-month chart review/survey. As there are concerns about deterioration after participation ends (being assessed in the deterioration analysis), it is more conservative to use the baseline value at six months for participants missing only this data point.

8.4.3 Imputing Data for Analyses of Nine Month Visits

No data will be imputed for any time point in the analysis of the nine-month time point. Imputation of baseline data at the six and nine months would reduce any observed changes.

8.5 Interim Analyses

No interim analyses were done in this study.

8.6 Multiplicity Considerations

The protocol specified that a multiplicity adjustment would be made for the primary analysis of the co-primary endpoints. For all other analyses, $P < 0.05$ (two-sided) will be considered statistically significant.

9 Summary of Study Data

9.1 Participant Disposition

The study did not collect data on all potential participants contacted for the study. Only the number of participants recruited and their disposition by study interval (T2, T3, T4) will be described.

9.2 Protocol Deviations

The following protocol deviations have been reported:

Deviation Type	Number	Description
Eligibility	1	One participant was found to be ineligible due to overlap in patient population with another provider. This participant was dropped from the study prior to completing any baseline forms.
Study procedure	5	Five participants were unable to complete chart reviews.
Visit scheduling	20	Two providers did not complete the baseline surveys and were dropped from the study. Five providers were given access to the survey forms early for one timepoint, and completed surveys earlier than scheduled. Two deviations for one provider were reported for enrollment and completion of baseline surveys after the first ECHO clinic. Eleven providers completed provider surveys after the survey collection window had ended.
Missed visit	9	Nine deviations across six providers were reported for incomplete or missing provider surveys.
Total	35	

None were considered serious enough to exclude available data from the analysis.

9.3 Demographic and Practice Information

Demographic and practice information will be summarized for the entire population. Demographic and practice information will also be summarized separately for completers and non-completers including assessing the statistical significance of differences between the two groups to assess the representativeness of the group completing the study.

Variables are listed in Section 6.4.1.

9.4 Baseline Outcome Measures

Baseline outcome measures will be summarized for the entire population. Baseline variables will also be summarized separately for completers and non-completers including assessing the statistical significance of differences between the two groups to assess the representativeness of the group completing the study, and separately by site to assess comparability of site.

Variables to be summarized are described in Section 10.1

9.5 Participant Attendance

The amount of training will be summarized (Section 6.4.2). In addition to the standard summary statistics, the cumulative proportion of those participating in X% or more of sessions (where X varies from 0 to 100) will be calculated.

9.6 Satisfaction

Summaries of the distribution of each satisfaction question will be done across all sites. In addition, a list of all the individual responses to the specific questions will be prepared.

These results will also be summarized separately for each site in the overall report. Sites will be referred to by letter (e.g. "A", "B", … , "J") which will be assigned randomly and will not be based on study cohort.

To preserve anonymity of the site data, proportions will be grouped or rounded so that individual sites cannot be identified based on the proportions in the overall report.

Site specific results will be supplied to each site for quality improvement purposes.

9.7 Intervention Fidelity

Intervention fidelity will be calculated for the entire program and separately for each site using all available intervention fidelity evaluations using both the fidelity score and the proportion of clinics with adequate fidelity (both defined in Section 6.4.4).

These results will also be summarized separately for each site in the overall report. Sites will be referred to by letter (e.g. "A", "B", … , "J") which will be assigned randomly and will not be based on study cohort.

To preserve anonymity of the site data, proportions will be grouped or rounded so that individual sites cannot be identified based on the proportions in the overall report.

Site specific results will be supplied to each site for quality improvement purposes.

10 Analysis of Outcomes

10.1 Outcome Measures

The table below lists for each outcome variable (a) the raw data used in the analysis for the outcome variable; (b) the distribution / link to be used in the model; and (c) the baseline summary for the participant included in the model. Alternate approaches are identified if models for the prespecified outcomes fail to estimate parameters or results appear inconsistent with the summary data (See Section 10.6)

Variable	Raw Data Used in Analysis	Distribution and Link to be Used	Baseline Value for Participant Used
Co-Primary Endpoints			
Proportion screened (Section 6.3.1.1)	Screening (Yes/No) for each chart abstracted Alternate: total number screened/ number of charts reviewed	Binary/Logit Alternate: Binomial/Logit	Proportion of all children screened at T1. Range: 0-1
Proportion of reported co-morbidities correctly treated (Section 6.3.1.2)	Number of comorbidities correctly treated for all reported comorbidity for a child. Children with no reported comorbidities are not included in the analysis. Alternate: total correctly treated / total number of comorbidities identified	Binomial/Logit Alternate: Binomial/Logit	Proportion of comorbidities correctly treated over all ASD children seen at T1. Range: 0-1
Prespecified Secondary Outcome Measures			
ASD knowledge (Section 6.3.2.1)	Percent score, as described in Section 6.3.2.1	Normal / Identity Alternate: Beta / Logit	Percent score at T1. Range 0-100

ASD self-efficacy (Section 6.3.2.2)	Percent score, as described in Section 6.3.2.2	Normal/Identity Alternate: Beta/Logit	Percent score at T1. Range 0-100
Additional Secondary Outcome Measures			
Number of co-morbidities reported (Section 6.3.3)	Number of co-morbidities reported for each child with ASD out of the 4 possible comorbidities per child. Alternate: total number identified / total possible	Binomial/Logit Alternate: Binomial/Logit	Mean number of comorbidities over all ASD children seen at T1. Range: 0-4.
Perceived barriers to care (Section 6.3.4)	Total number of specific barriers checked / total number possible, as described in Section 6.3.4 Alternate: total number of specific barriers	Binomial/Logit Alternate: Poisson/Log	Proportion at T1. Range 0-1 Alternate: Number at T1. Range: 0-10

10.2 Descriptive Summary of Outcome Measures

The observed data (without imputation) will be summarized for each outcome measure at each time point.

10.3 Efficacy Analyses

10.3.1 Primary Efficacy Analyses

The primary efficacy analysis will be done in the completer population.

After imputation of missing data as described in Section 8.4, a generalized linear mixed model analysis will be used to predict the outcome (listed in Section 10.1, with details of how each variable is calculated and baseline is calculated) with the following fixed effects:

- period (cohort), a continuous variable from 1-5;
- time point (treated both as categorical variables [coded as "baseline", "3 months" and "6 months"] and as a continuous variable [coded as 0, 0.5, 1.0]; see below);

and the following random effects:

- site; and

- participant.

After the initial analysis, a final decision will be made as to whether the effect of time point should be treated as a categorical or a continuous variable. If the results suggest that there is a substantial benefit to treating time point as a categorical variable in at least one of the two co-primary outcome measures, then it will be retained as a categorical variable for all outcome measures; otherwise it will be treated as a continuous variable. For the purposes of this analysis, a substantial benefit is defined as a statistically significant improvement using a likelihood ratio test when time is treated as a categorical variable rather than a continuous variable. The decision in the primary efficacy analysis will be used in all other analyses.

Technical note: This model will use PROC GLIMMIX, and two random statements, one fitting a random intercept for site (RANDOM SITE) and one fitting a random intercept for PCP (RANDOM /subject = PCP). Note that although multiple imputation (PROC MIANALYZE) should work with PROC GLIMMIX results, there are reports of computational problems arising. Should such problems arise in the analysis of the data from the study, the use of multiple imputation will be reconsidered.

10.3.2 Sensitivity Efficacy Analyses

Because of differences in screening practices at one site, the primary efficacy analysis of screening will be repeated with the data from this site removed to ensure robustness of conclusions.

10.3.3 Exploratory Analyses

Exploratory analyses will explore whether demographic or practice variables (Section 6.4.1), or the amount of training (Section 6.4.2) are predictors of screening practice or treatment of comorbidities. Such analyses will be done by adding the relevant demographic or practice variable to the model in Section 10.3.1, to determine the statistical significance of the factor as a main effect.

We will also explore whether there are differences in screening rates by well-child visit by incorporating a term for the well-child visit in the model specified in Section 10.3.1. This will exclude the data from the site with different screening practices.

No exploratory analyses are planned for the other endpoints.

10.4 Effectiveness Analysis

The effectiveness analysis will repeat the efficacy analysis (Section 10.3) for the total population.

10.5 Analysis of Long-Term Impact

This analysis will estimate how much of a change occurs at 9 months. It will extend the model included in the efficacy analysis (section 10.3.1) using the following predictors as fixed effects in the generalized linear mixed model framework:

- period (cohort), a continuous variable from 1-5;
- time point (categorical value coded as “baseline”, "3 months", "6 months" and "9 months"));

and the following random effects:

- site; and
- participant.

Deterioration will be measured by the contrast of the "9 month" estimate and the "6 month" estimate.

10.6 Consistency of Outcome Results

If modeling results appear inconsistent with the summary data then the alternative model specified will also be used, and both sets of results will be presented to help the study team better understand the results of the study.

10.7 Graphical Presentation of Outcome Results

Using the efficacy analysis population, spaghetti plots for proportion of children correctly screened, proportion of co-morbidities correctly addressed, and number of co-morbidities identified per charge will be plotted for

a) site averages; and

b) individual participants.

Given the number of participants completing the study, results for the individual participants will be plotted separately for sites with over 8 completers and sites with under 8 completers will be pooled together.

11 Summary of Changes to the Protocol and / or SAP

The protocol specified numerous early thoughts on the analysis and the following changes have been made to the analysis plans:

a) The primary outcome analysis has been changed as follows:

- Rather than using ASD patient received / did not receive appropriate co-morbidity management, the number of correctly treated co-morbidities, among the identified co-morbidities will be used as the outcome variable, as described in

Section 6.3.1.2. This endpoint contains more information than the original yes/no variable.

- Study period is treated as a fixed effect rather than a random effect in the model.
- Data from T1, T2, and T3 are used in the primary analysis rather than only T1 and T3, as this should provide more information.
- There is no attempt to quantify the amount of treatment an individual PCP received in the basic modeling analyses. The amount of treatment (as proportion of sessions completed) will be examined in an exploratory analysis.

b) No attempt is being made to incorporate the precise timing of the T2 chart review in the analysis. This is related to the decision not to attempt to quantify treatment in the primary analysis.

c) Similarly, we are not planning to use the specific sessions that a PCP attended in the analysis. We are, however, planning to use the amount of training as a potential predictor in exploratory analyses.

d) As we are not quantifying the amount of treatment in the primary analysis, no interactions of treatment with other variables will be considered in the analysis.

e) No attempt will be made to analyze how the number of ASD patients changes over time. It was recognized while preparing the SAP that this metric is not an immediate outcome from ECHO AUTISM.

f) There will be no comparison of results (e.g. satisfaction) across centers, although each site will receive its own data as a quality improvement measure.

12 Operational Problems During the Study and Impact on the Analysis

As with all clinical studies, various problems arose during the study. Only those that are relevant to the analysis are mentioned here.

a) The initial plan was that each cohort would open early in the start month and have two ECHO clinics each month. Because of IRB or recruitment issues, two sites started in the month after the scheduled start for their cohort. This is ignored in the analysis.

b) During the study the team learned that Canadian guidelines for screening for ASD were not consistent with US guidelines. In the primary analysis (Section 10.3.1) PCP screening practice at the Canadian site is included in the analysis, and screening graded based on consistency with the Canadian guidelines. A sensitivity analysis (Section 10.3.3) excludes the site from the analysis of the screening outcome measure.

c) One site withdrew from the ATN during the course of the study. All data collection at the site was completed before the site withdrew from the network. There was a special data review of all their data to ensure data collection was complete, which included data post-baseline. No analyses compared results over time during this data review. Thus, the original SAP was prepared prior to any formal analysis of post-baseline data. The study statistician was aware of convergence / modeling problems at the time of the revision, but did not have access to data after baseline or results of the analyses.

d) It was originally anticipated that hub coordinators would visit each PCP practice four times, once at each of the four timepoints. Because of the travel involved, multiple sites

requested permission to combine chart reviews so that fewer visits were needed to clinics. Ultimately, the core team allowed sites to abstract two sets of charts at the same time (T1 and T2; T3 and T4). The core team decided that this was unlikely to introduce bias into the study as the period for each of the chart reviews was clear and charts would continue to be available if the site was visited at a later time.

These operational problems are not expected to impact the conclusions from the study.