



HEALTH CARE AND HUMAN SERVICES POLICY, RESEARCH, AND CONSULTING—WITH REAL-WORLD PERSPECTIVE.

Study of Health Outcomes in Children with Autism and their Families

Task D: ASD Risk Factor Analysis

Final Report

Prepared for: National Institute of Mental Health

Submitted by: The Lewin Group, Inc.

September 26 2012

Study of Health Outcomes in Children with Autism and Their Families

Task D: ASD Risk Factor Analysis

Final Report

Prepared for:

National Institute of Mental Health

Submitted by:

The Lewin Group, Inc.

September 26, 2012

Executive Summary

A. Introduction and Study Objectives

The National Institute of Mental Health (NIMH) contracted with The Lewin Group to conduct a two-year study from September 2010 to September 2012 entitled “The Study of Health Outcomes in Children with Autism and their Families.” This study seeks to address a significant gap in the empirical knowledge base about the trajectories of health conditions, health outcomes and utilization of health care services among children with autism spectrum disorders (ASD), their siblings, and their parents. The ability to study a very large and heterogeneous group of children with ASD using claims data and the ability to link to information about family members is unprecedented and holds promise to advance clinical and health services knowledge about ASD substantially.

The overall purpose of Task D was to explore the utility of claims data in investigating potential risk factors for ASD. Given that the etiology of ASD is not clearly known, a wide set of risk factors in parents and children has been proposed by the research community. Our selected possible risk factors span the preconception, prenatal, early postnatal, and early childhood periods; moreover, many of these factors can be completely or nearly completely captured by claims data. The four objectives of Task D were to:

- Evaluate the frequency with which an apparent mother and an apparent father can be identified within claims data and the representativeness of the children subsamples with linked parents compared to the base study sample;
- Identify the size of available subsamples meeting varying criteria for continuous enrollment over etiologic windows of potential interest;
- Explore the feasibility of using claims-based data on scheduled obstetric procedures to more accurately identify conception and trimester cut points than simpler algorithms relying on just date of birth and presence/absence of ICD-9 codes indicating preterm delivery;
- Explore the feasibility of identifying and measuring selected potential ASD risk factors in claims data, including:
 - Early postnatal and postnatal window risk factors (measurable from child claims):
 - Neonatal Intensive Care Unit (NICU)/ICU admission
 - Non-singleton pregnancy
 - Maternal fetal hypoxia
 - Prematurity/pre-term delivery
 - MMR vaccination
 - Prenatal window risk factors (measurable from maternal claims):
 - Maternal asthma
 - Maternal depression
 - Maternal autoimmune conditions
 - Maternal prenatal infection

- Maternal obesity
 - Maternal gestational diabetes
 - Maternal preeclampsia/eclampsia
 - Maternal exposure to anesthesia
 - Prenatal ultrasound
 - Maternal medication use during prenatal and early postnatal time periods (e.g., SSRIs, B2Ars, contraceptives, anticonvulsants, antibiotics, pitocin).
- Preconception window risk factors (measureable from maternal and/or paternal claims):
 - Paternal autoimmune conditions
 - Maternal obesity (prior to conception)
 - Maternal or paternal receipt of infertility treatment

B. Study Design and Analytic Strategy

This retrospective claims data study used medical data, pharmacy data, and enrollment information from the OptumInsight's research database containing claims from the large health plan affiliated with OptumInsight. All study subjects were identified among commercial enrollees who have medical, pharmacy, and behavioral health coverage.

Based on the results of the Task A Chart Study, children with at least 2 ASD claims were defined as having ASD and were included in the Task D study base sample. In the Chart Study, the positive predictive value increased from 74.2% to 87.4% when children with only 1 ASD claim were excluded from the case definition, increasing our confidence that the children with ASD in Task D are true cases. Children with only 1 ASD claim are therefore not included in our ASD or control sample in the Task D study (similar to Task B and C).

To address the research question of the frequency with which parents can be identified within claims data, we required a child's date of birth to be within +/- seven days of the mother's delivery date. Since paternal risk factors are most useful during the preconception window, we only considered father-offspring pairs where fathers were enrolled at conception. We estimated the date of conception to be the delivery date minus 280 days; we also used the delivery date to identify trimester cut-points. To assess the availability of subsamples over etiologic windows of potential interest, we identified subsamples of children with enrollment from birth to various time periods after birth, subsamples of mothers with various enrollment periods before and after their child's birth, and a subsample of fathers with enrollment during the preconception period. To explore whether claims data on scheduled obstetric procedures can more accurately identify conception and trimester cut points, we determined the appropriate trimester for various procedures and identified the proportion of mothers whose data showed a code for the procedure in the appropriate trimester. Lastly, we examined whether claims data is useful in studying risk factors for ASD by measuring the prevalence of risk factors in mothers, fathers, and children in relevant etiologic time windows and comparing them to peer-reviewed and gray literature prevalence estimates.

C. Results

We found the following results, organized by study objective:

OBJECTIVE 1: Identification of apparent mothers and fathers and representativeness of children samples with identified parents

- The final proportion of children with mothers identified in claims data is 4.1% and 4.9% for the ASD and comparison groups, respectively. This result underscores that the identification of children in claims data who can be linked to a mother with usable claims experience in etiologic windows of interest is quite limited in terms of percentages of total population. This pattern held true for fathers.
- Though the proportion of children who can be linked to mothers or fathers is low, the similarity between the ASD and comparison proportions indicates that our approach to sampling is not an overt source of selection bias.
- The Task D children subsamples linked to parents were representative of our Task A Likely ASD base sample children enrolled at birth.

OBJECTIVE 2: Identify the size of available subsamples meeting varying criteria for continuous enrollment over etiologic windows of potential interest.

- Of the 2,176 mothers of children with ASD identified via the linkage criteria in Task D, a little over three-fourths had continuous enrollment from birth minus 14 weeks to birth while just over 55% had continuous enrollment from conception (birth minus 40 weeks) to birth. Only 26% of mothers had continuous enrollment from birth minus 92 weeks (the period comprising the entire prenatal period plus a one year preconception period).
- About 70% of the mothers identified were linked to a child with ASD with continuous enrollment through one full postnatal year (where early postnatal child risk factors could be measured) and 64% of mothers were linked to a child with ASD with continuous enrollment through 24 months of age. For fathers of children with ASD identified within seven days of the date of conception, approximately half had continuous enrollment from birth minus 92 weeks to birth minus 40 weeks (1 year prior to conception to conception).
- Quite similar, but slightly higher, proportions were seen for mothers of comparison children.

OBJECTIVE 3: Explore the feasibility of using claims-based information on scheduled obstetric procedures to more accurately identify conception and trimester cut points

- Overall, the timing of the procedures corresponded well with the trimester in which they are typically scheduled validating our algorithm using days from date of birth to define conception and trimester cutpoints. For example, the inhibin screen and the alpha-fetoprotein test which only occur in the 2nd trimester, took place during the second trimester for over 98% and 99% of the women who had the procedure, respectively.

OBJECTIVE 4: Identification and Measurement of Selected Potential Risk Factors for ASD

- Some of our prevalence estimates for potential ASD risk factors were consistent with published estimates: these included preterm birth, chronic maternal health conditions

potentially initiating prior to pregnancy (e.g., asthma and depression), medication use for those conditions, anesthesia use, maternal infertility treatment, and MMR immunization.

- Some of our prevalence estimates were inconsistent with published estimates; these included: obesity, Pitocin, NICU and ICU admissions, gestational diabetes, preeclampsia, ultrasounds, and antibiotic use. In some instances we suspect that coding issues might be driving these differences (e.g., our range of codes may have been too broad in some instances).
- We examined some risk factors that, *a priori*, we expected would not be measured accurately in claims (e.g., obesity – which is not consistently coded in claims; and pitocin – which is known to be bundled into other procedure codes) – and our analyses confirmed that these are likely not well-captured through claims.
- We explored the potential of identifying groups of children who were completely free of vaccinations, a cohort that would be useful to explore continued concerns about links between ASD and immunization. While we found that nearly five percent of the children in both our ASD and comparison children samples did not have evidence of any immunization, this proportion is many times higher than reported as unvaccinated in the literature, suggesting that this most likely does not represent a truly unvaccinated cohort and that many children included in this group actually received vaccines that either did not generate a claim or that were paid for by other sources.

D. Implications and Recommendations

In summary, our ability to evaluate nine years of claims from a large database, allowed us to draw a large and heterogeneous group of children with ASD and comparison children despite various enrollment and risk factor criteria, which allows our sample sizes to be sufficient for reliable estimates of risk factors in both ASD and comparison groups. Specifically, our results lead to the following implications:

- Though a requirement of linkage to mothers has a major impact on the size of study samples (only 4% of ASD and comparison group children were linked to mothers at birth), the proportions of children retained at each step of the sequence used to link to an apparent mother were similar in the ASD and comparison groups. This suggests that requirement of linkage to mothers is not an overt source of selection bias.
- We found that despite decreases in the number of subjects as continuous enrollment requirements were extended, these smaller sample subsets generated similar estimates to the larger linked samples. This implies that researchers may not need to be overly concerned about the impact of enrollment criteria on exposure estimation.
- Our results suggest that the timing of procedures in claims data could be used to validate or improve the precision of etiologic windows for risk factors. However, a recent Canadian study found that adding date information from screening procedures did not substantially improve gestational age estimation. Nonetheless, a validation against clinical gestational age in our private claims database would still be a useful next step to confirm that the findings from the British Columbia health system generalize to the US.

- Our findings indicate that a number of risk factors for ASD are reliably captured in claims data, suggesting that claims may be a robust and comparatively inexpensive source of information for inquiry into the factors related to ASD.

While the numbers of informative children with ASD with linkage to parents and sufficient continuous enrollment to allow meaningful investigation are only a small proportion (<10%) of the very large number of children with ASD that can be located in a cross-sectional query of a large claims database, our examination of nine years of claims data still generated two to three thousand ASD cases that would be of potential use in etiologic research. We found that while it is clear that private insurance claims data will not be an adequate information source for a number of potential ASD risk factors; we saw that there was also a wide range of potential risk factors where claims-based research could have the capacity to add to the developing epidemiologic knowledge base. In general, claims would appear to be a viable data source for investigation of maternal medical conditions, that require active medical management (e.g., asthma, depression) and their treatments (in particular, pharmacologic therapies). The size of the claims database creates opportunities to control for confounding by indication as there appear to be sufficient numbers of both treated and untreated women with such conditions identifiable. Similarly, serious early postnatal complications (NICU admission, pre-term delivery) also appear amenable to investigation through claims. Risk factors explored that might be worth additional exploration include infertility (where refinement to include only plans with certain benefits policies might improve the accuracy of claims-based assessment) and parental medical conditions with somewhat less-intensive medical management. In addition, future studies using claims to explore potential ASD risk factors might seriously consider the incorporation of formal validation sub-studies on both exposure and diagnosis (gathering data on exposure and diagnosis from other data sources on a fraction of the sample), since these can often be implemented on a reasonable timetable and at a fraction of the cost of studies of comparable size that would require primary data collection on every subject.

Table of Contents

EXECUTIVE SUMMARY	i
A. Introduction and Study Objectives.....	i
B. Study Design and Analytic Strategy	ii
C. Results.....	iii
D. Implications and Recommendations.....	iv
I. INTRODUCTION AND BACKGROUND.....	8
A. Overview of Study.....	8
B. ASD Diagnosis and Treatment.....	9
C. Potential Risk Factors for ASD	10
II. STUDY OBJECTIVES	11
III. STUDY DESIGN.....	13
A. Data Sources	13
1. <i>Claims Data Sources</i>	13
2. <i>Socioeconomic Data</i>	14
B. Study Reviews.....	15
1. <i>Institutional Review Board (IRB) Review</i>	15
2. <i>OptumInsight Disclosure Limitation Program</i>	15
C. Base Study Sample	15
1. <i>Base Study Subject Eligibility Criteria</i>	16
2. <i>Time Windows for Base Sample Identification and Observation</i>	18
3. <i>Refinement of ASD Base Study Sample within Task D</i>	19
4. <i>Base Study Sample Size</i>	20
D. Variable Definitions.....	22
1. <i>Patient Characteristics</i>	22
2. <i>Patient Comorbidity Scores</i>	24
3. <i>Potential Risk Factors</i>	24

E. Methodology	29
1. Evaluate the frequency with which the apparent mother and an apparent father can be identified within claims data and the representativeness of the subsample with linked parents compared to the base study sample.....	29
2. Identify the size of available subsamples extracted from the base sample meeting varying criteria for continuous enrollment over etiologic windows of potential interest.....	30
3. Explore the feasibility of using claims-based information on scheduled obstetric procedures to more accurately identify conception and trimester cut points.....	32
4. Explore the feasibility of identifying and measuring selected potential ASD risk factors measurable in claims data.	32
IV. RESULTS	34
OBJECTIVE 1: Identification of apparent mothers and fathers and representativeness of children samples with identified parents.....	34
OBJECTIVE 2: Identify the size of available subsamples meeting varying criteria for continuous enrollment over etiologic windows of potential interest.	42
OBJECTIVE 3: Explore the feasibility of using claims-based information on scheduled obstetric procedures to more accurately identify conception and trimester cut points	43
OBJECTIVE 4: Identification and Measurement of Selected Potential Risk Factors for ASD	45
1. Literature Search Results	45
2. Potential Risk Factors' Prevalences in the Early postnatal and Postnatal Period (measured in children's' claims)	46
3. Potential Risk Factors Prevalences in the Prenatal Period (measured in maternal claims) ...	49
4. Potential Risk Factors Prevalences in the Preconception Period (measured in maternal and paternal claims)	64
V. DISCUSSION	67
A. Generalizability (External Validity)	67
B. Potential for Selection Bias and Study Size Considerations.....	68
C. Information Bias.....	68
D. Conclusion	71
VI. REFERENCES	73

I. Introduction and Background

A. Overview of Study

The National Institute of Mental Health (NIMH) contracted with The Lewin Group to conduct a two-year study from September 2010 to September 2012 entitled “The Study of Health Outcomes in Children with Autism and their Families.” The Lewin Group’s study team is a collaboration of organizations reflecting expertise in the epidemiology of autism spectrum disorders (ASD), health services research, and the clinical care of children and families. An External Advisory Committee comprised of experts in ASD research as well as stakeholders from parent advocacy groups and treatment providers was also convened to provide consultation and guidance to the project team. This study sought to address a significant gap in the empirical knowledge base about health conditions and health care service utilization among children with ASD, their siblings, and their parents. The project employed large administrative health care claims databases to fulfill four distinct aims:

- Task A: To identify a large and diverse number of children with ASD and a general population comparison group, along with their families, and describe factors related to each of these populations including age and gender distribution, geographic distribution, and socioeconomic characteristics.
- Task B: To describe and compare the health conditions of children with ASD and their families to similar families without a child with ASD.
- Task C: To describe and compare the use of health services by children with ASD and their families to similar families without a child with ASD.
- Task D: To explore the utility of claims data in investigating potential risk factors for ASD.

Task A, conducted between September 2010 and March 2012, was comprised of two subtasks: 1) a baseline claims analyses to identify and describe children with ASD, their siblings and parents, and their respective comparison groups, from the large administrative claims dataset; and 2) a medical chart review to validate the claims-based identification of children with ASD in the study population, or the “chart study.” The purpose of the chart study was to evaluate the ability to identify children with ASD within research claims databases by comparing claims-based ASD case identification to ASD status as documented in clinical (medical) charts.

The focus of this report is to present the methodology, approach, and results of the Task D potential ASD risk factor analysis. The methodology and results of the Task A baseline claims analyses and Task A chart study informed our approach for Task D and are detailed in companion reports that were delivered to NIMH October 17, 2011 and March 2, 2012.

While much research is underway to examine the prevalence and consequences of ASD, to identify the risk factors and potential causes of ASD, and to explore potential treatments, fewer efforts have been directed toward understanding the overall health status of a large heterogeneous group of children with ASD and their family members.¹ To date, few studies have

¹ See the National Institute of Mental Health web page on autism: <http://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-pervasive-developmental-disorders/index.shtml#>, the link there to clinical trials regarding autism,

used large administrative claims databases to examine health outcomes in children with ASD, especially over an extended period of time.ⁱⁱ In addition, as most studies have been clinical studies with small sample sizes that are not representative of the US population of children or children with ASD generally, a larger, more representative study drawn from existing electronic datasets can help advance the research for children with ASD and their families without the additional burden to individuals, families, clinicians or researchers of prospective data collection. Finally, longitudinal data for family members of children with ASD will inform research on how ASD impacts families in addition to its effects on the individual with ASD over time. In particular, Task D helps to inform us if we can also rely on large administrative claims databases to examine potential risk factors for ASD.

B. ASD Diagnosis and Treatment

ASD includes Autism, Asperger's Syndrome, and Pervasive Developmental Disorder not otherwise specified (PDD-NOS). Although Rett Syndrome and Childhood Disintegrative Disorder are also considered Pervasive Developmental Disorders, and thus belong on the autism spectrum, they are not a focus of this study.

ASD is a group of developmental disorders that have significant and life-long impacts on affected individuals and their families. The key features of ASD are sustained impairments in communication and social interaction, restricted interests, and repetitive behaviors. Common ASD-associated conditions and co-morbidities include anxiety, depression, epilepsy or other seizure activity, learning disabilities, obsessive-compulsive disorders and attention deficit disorder.¹

The diagnosis of ASD has been increasing in recent years, and the most recent statistics from the Centers for Disease Control and Prevention now estimate that 1 in 88 children have an autism spectrum disorder.² Whether this change can be fully explained by improved awareness and by the greater availability of services or, instead, is related to an as-yet unknown environmental exposure is still to be determined. As ASD is heterogeneous in its characteristics and presentation, the meaning of the diagnosis itself is unclear, often raising more questions than it answers regarding risk factors, heritability, health trajectories, promising treatments, and outcomes.

Since ASD also manifests along a spectrum of severity, its prognosis is also highly variable, and ranges from very poor quality of life with only minimal ability to function independently to relatively normal social and vocational functioning or even superlative skills in a focused area. While the causes of ASD are not known, both genetics and environment are believed to be etiologic factors.

Currently, the disorder does not have a cure but treatment for ASD, especially when implemented early, can help children advance social and language skills, address behavioral and learning problems and improve functioning and quality of life.³ Common therapies include educational and behavioral interventions (e.g., applied behavioral analysis, speech therapy, and occupational therapy) and medications that ameliorate associated symptoms and conditions. Such medications include antidepressants, anti-anxiety medications, stimulants, anticonvulsants (for seizures), and

and also the research summary by the Interagency Autism Coordinating Committee at <http://iacc.hhs.gov/summary-advances/2010/>.

ⁱⁱ The Request for Proposal for this study, HHS-NIH-MH-2010-018 at Attachment 3 page 2 of 12 references the "significant gaps" in this area.

antipsychotic medications (for impulsivity or other behavioral symptoms). Alternative therapeutic approaches (e.g., dietary interventions) are also used.

C. Potential Risk Factors for ASD

In spite of the rising prevalence of ASD in the last two decades, the etiology of these disorders is still not fully understood. A need for a continued, in-depth investigation into the causes of ASD, including understudied environmental risk factors, has been increasingly highlighted.⁴ One data source with underexplored potential for useful and cost efficient exploration of potential risk factors for ASD is administrative health care claims data. Claims data may allow for assembling large sample sizes within which to study a variety of etiologic questions with statistical power and broad representativeness. In fact, secondary claims datasets including Medicare and private health insurance claims have previously been used successfully to identify risk factors for other biologically-based health conditions, such as venous thromboembolism.^{5,6} Such claims data, including Medicaid and private insurance claims data, have also been fruitful in other types of ASD research, although they were primarily efforts focused on health services and cost.^{7,8,9,10}

The goal of Task D was to explore the possible value of claims data to support research about ASD etiology. ASD is a behaviorally-defined neurodevelopmental disorder with symptoms typically emerging prior to age three. As such, the etiologic windows of interest in ASD include early life, early postnatal, prenatal, and potentially, the preconception periods. Data from neuropathologic,¹¹ gene-expression,¹² and twin studies,¹³ strongly support the prenatal period as one of particular interest — though emerging research on epigenetic mechanisms in autism has also called attention to the preconception period. Furthermore, as the developmental progression of autism seems to fluctuate^{14,15} it may be important to consider the etiologic influence of exposures throughout early childhood as well. If realized, the success of strategies to use secondary data sets like claims data to detect potential risk factors could be built upon further to identify specific groups of risk factors which could ultimately inform primary prevention strategies of ASD.

There is little previous claims-based research on this topic. The potential for information bias stemming from inaccurate measurement of principal variables is a potential hindrance to this approach. In Task A, we presented data on, and discussed issues related to, the potential misclassification of ASD outcomes. Also of concern in claims data is the ability to assemble accurate data on risk factors of interest. First, this requires the opportunity to observe claims during the etiologic windows of interest as discussed above. Next, there needs to be confidence that information captured on claims measures exposures of interest with acceptable accuracy. In Task D, we focused on exploring the size of subpopulations that can be extracted from our source claims data set with sufficient opportunity to observe claims during etiologic windows of interest, the extent to which selecting these subpopulations may introduce selection bias, and the ability of the claims data during these intervals to measure risk factors of interest.

II. Study Objectives

The overall purpose of Task D was to use the base sample from Task A: Baseline Claims Analyses to explore the utility of claims data in investigating potential risk factors for ASD. Given that the etiology of ASD is not clearly known, a wide set of risk factors in parents and children has been proposed by the research community, including factors as diverse as prenatal infections, parental age, health conditions such as epilepsy, psychiatric conditions such as anxiety, prescription medication use, and many others. The particular factors we selected reflect a thorough literature review and careful consideration of how best to use claims data to measure such factors. Our selected possible risk factors span the preconception, prenatal, early postnatal and early childhood periods; moreover, many of these factors can be completely or nearly completely captured by claims data. We organized our analysis by examining potential risk factors in the categories of children, mothers, and fathers.

In order to explore the utilization of claims data to investigate potential risk factors for ASD, Task D is comprised of the following supporting objectives:

1. Evaluate the frequency with which an apparent mother and an apparent father can be identified within claims data and the representativeness of the children subsamples with linked parents compared to the base study sample.
2. Identify the size of available subsamples meeting varying criteria for continuous enrollment over etiologic windows of potential interest.
3. Explore the feasibility of using claims-based data on scheduled obstetric procedures to more accurately identify conception and trimester cut points than simpler algorithms relying on just date of birth and presence/absence of ICD-9 codes indicating preterm delivery.
4. Explore the feasibility of identifying and measuring selected potential ASD risk factors in claims data, including:
 - Early postnatal and postnatal window risk factors (measurable from child claims):
 - Neonatal Intensive Care Unit (NICU) admission
 - Non-singleton pregnancy
 - Hypoxia
 - Prematurity/pre-term delivery
 - MMR vaccination
 - Prenatal window risk factors (measurable from maternal claims):
 - Maternal asthma
 - Maternal depression
 - Maternal autoimmune conditions
 - Maternal prenatal infection
 - Maternal obesity
 - Maternal gestational diabetes
 - Maternal preeclampsia/eclampsia

- Maternal exposure to anesthesia
- Prenatal ultrasound
- Maternal medication use during prenatal and early postnatal time periods (e.g., SSRIs, B2Ars, contraceptives, anticonvulsants, antibiotics, pitocin).
- Preconception window risk factors (measureable from maternal and/or paternal claims):
 - Paternal autoimmune conditions
 - Maternal obesity (prior to conception)
 - Maternal or paternal receipt of infertility treatment

In the course of exploring the feasibility of identifying and measuring potential ASD risk factors in claims-based data, as discussed below, we will present estimates of these potential risk factors in both ASD and comparison groups. However, it is important that the contrast between these estimates not be interpreted as evidence supporting or refuting an association between those potential risk factors and ASD. We caution against such an interpretation because: 1) algorithms to identify these factors in our claims data set must be considered preliminary; 2) issues related to whether claims are an adequate source need to be more fully explored in many instances; and 3) these analyses made no attempts at causal modeling (i.e., there was no formal consideration of confounding or modifying variables, selection bias, or other sources of information bias like surveillance effects).

The remainder of this report describes the data, methods, results and implications related to these objectives.

III. Study Design

This retrospective claims data study used medical data, pharmacy data, and enrollment information from the OptumInsight's research database containing claims from the large health plan affiliated with OptumInsight. OptumInsight claims data were linked to a consumer database for select socioeconomic information. All study subjects were identified among commercial enrollees who have medical, pharmacy, and behavioral health coverage.

This section outlines the details of our study design. This includes: a) an overview of the database that was the source of the claims-based analyses and the source of sample selection for the Task D studies; b) the study reviews that were required for approval of the study to be in compliance with privacy and ethical policies; c) a description of the study sample, including subject eligibility criteria, sampling strategy, observation periods, and analytical subsamples of interest; d) detailed descriptions of the analytical variables constructed for the study; and e) our methodology.

A. Data Sources

The data sources for the Task D study included both claims data and a linked database containing socioeconomic data for study subjects.

1. Claims Data Sources

OptumInsight has access to a proprietary research database ("OptumInsight Research Database") containing medical (including behavioral health) and pharmacy claims with linked enrollment information covering the period from 1993 to 2010. For 2009, data relating to approximately 13.3 million individuals with both medical and pharmacy benefit coverage are available. The underlying population is geographically diverse across the US and reasonably representative of the privately insured US population.

■ Medical Claims

Medical claims or encounter data are collected from all available health care sites (inpatient hospital, outpatient hospital, emergency room, outpatient office, surgery center, etc.) for all types of covered services, including specialty, preventive and office-based treatments. Medical claims and coding conform to insurance industry standards. Claims for ambulatory services submitted by individual providers (e.g., physicians) use the HCFA-1500 or CMS-1500 format.¹⁶ Claims for facility services submitted by institutions (e.g., hospitals) use the UB-82, or UB-92, or UB-04 format.^{17,18} Medical claims include: diagnosis codes recorded with the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis codes; procedures recorded with ICD-9-CM procedure codes, Current Procedural Terminology (CPT), or Health care Common Procedure Coding System (HCPCS) codes; site of service codes; provider specialty codes; revenue codes (for facilities); paid amounts; and other information. Typically, facility claims do not include complete information about drugs administered within a hospital. Approximately 6 months following the delivery of services is required for complete medical data due to lags in claims submissions and final claims processing. In this report, the term "medical claims" is used to refer to claims for both physical health care and behavioral health care submitted and processed for reimbursement. Health care not processed as a medical claim (e.g., care provided as part of a wellness program or as an Employee Assistance Program - EAP) is not included.

- **Pharmacy Claims**

Claims for pharmacy services are typically submitted electronically by the pharmacy at the time prescriptions are filled. The pharmacy claims history is a profile of all outpatient prescription pharmacy services provided and covered by the health plan. Pharmacy claims data include drug name, dosage form, drug strength, fill date, days of supply, financial information, and de-identified subject and prescriber codes, allowing for longitudinal tracking of medication refill patterns and changes in medications. Pharmacy claims are typically added to the research database within 6 weeks of medication dispensing.

The OptumInsight Research Database is a unique data source for autism research, affording rich, longitudinal data on disease and comorbidity and health care utilization and costs for large samples of study subjects. Nonetheless, claims data have inherent limitations given that they are generated for payment, not research, purposes. For example, a pharmacy claim is for a filled prescription that may or may not be consumed by a patient as prescribed. Over-the-counter medications or medications provided as samples by a physician are not included in the data and therefore could not be measured. Information on diagnosis may also be inaccurate. For example, a diagnosis submitted on a claim may be an interim or transient diagnosis, while the patient is undergoing tests until a definitive diagnosis is established. Thus, in order to enhance accuracy in claims analysis, researchers frequently apply inclusion and exclusion criteria as appropriate - for example, requiring multiple appearances of a diagnosis code over time -- before considering a particular condition to be present. Similarly, diagnoses that do not impact payment or that could negatively impact payment may be under-reported. Finally, minor conditions that did not result in medical treatment at a healthcare setting and diagnoses made outside the health care setting are not captured.¹⁹ For example, diagnoses, evaluations and treatments made within the educational system are not included.

2. *Socioeconomic Data*

Many aspects of health care utilization and cost, including treatment selection, therapy patterns, and health conditions, may be associated with factors not directly measured in administrative claims data. For example, a vast literature has demonstrated differences in a variety of health-related conditions for patients of differing educational attainment, income, net worth, race/ethnicity, and family composition.^{20, 21} To allow for more powerful insight into the prevalence and burden of illness, OptumInsight has linked a unique source of patient-level data to the OptumInsight administrative claims data that allows for analysis of socioeconomic characteristics. The socioeconomic data are derived through a match done by the health plan with a marketing database maintained for a large segment of the US population. Specifically, these data elements include race, ethnicity, homeowner status, occupation type (e.g., blue collar, white collar, self-employed), household income category, and household net worth category. The data populating these socioeconomic elements are generated by a combination of self-report, modeling, census data, and a variety of other individual-level and population-level data sources. Approximately 30% of the race/ethnicity data are collected directly from public records (e.g. driver's license records), while the remaining data are imputed based on sophisticated algorithms using enhanced geocoding (e.g. address and census block data enhanced by onomastic rules). Household income and net worth are populated either by self-report or through predictive modeling. Sources for the self-reported economic measures include national surveys and consumer product registrations. Predicted household income and net worth are generated by

modeling a variety of factors including age, occupation, home ownership, and median income from the Block Group Census data. While these data have application to health economics and outcomes research, certain limitations are associated with these data, including potential inaccuracies in the assignment of socioeconomic status, missing data, and pre-defined categorizations (e.g., income level). Rates of missing data vary, depending on the specific study population and the specific data elements used. The socioeconomic variables used in this study were household income, race/ethnicity, and household size (number of adults and children within the household). Generally, these variables are populated for 60-70%, 65-75%, and 30-55% of the claims population, respectively.

The socioeconomic database is refreshed on a quarterly basis. Data used for this study were based on the most recent refresh available to OptumInsight, which varied from September 2007 through June 2011 for individual subjects. Depending on whether a subject's information changed between refreshes, the effective dates for the socioeconomic information used in this study may have been earlier than the latest refresh date and varied by subject.

B. Study Reviews

1. Institutional Review Board (IRB) Review

OptumInsight submitted the study protocol and the request for review of claim of exemption to the New England Institutional Review Board (NEIRB). In December 2011, NEIRB exempted the study from IRB review. The study was eligible for exemption under Category E (research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available OR if the information is recorded by the Investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects).

2. OptumInsight Disclosure Limitation Program

OptumInsight has implemented a Disclosure Limitation Process as part of its overall privacy initiative, in order to comply with applicable privacy laws and best business practices in protecting sensitive data in OptumInsight custody. Specifically, OptumInsight's Disclosure Limitation Program allows OptumInsight to comply with the Privacy Rule adopted by the U.S. Department of Health and Human Services under the Health Insurance Portability and Accountability Act (HIPAA). In situations where the Privacy Rule does not allow use of protected health information (PHI), the Privacy Rule does allow de-identification of the PHI. Once de-identified, PHI is no longer subject to the Privacy Rule, and can be used or disclosed without limitation (as long as it is not re-identified). OptumInsight has worked with recognized industry experts on de-identification methodology to comply with HIPAA Privacy requirements and developed a "Statistical Alternative Methodology" for de-identification of data. In December 2011, disclosure analysis of the study data was completed under OptumInsight's Disclosure Limitation Program, and it was determined that the data has been de-identified as required under applicable law and that there is a minimal risk of re-identification.

C. Base Study Sample

The base samples for this study were the subjects identified within the OptumInsight Research Database for Task A: Baseline Claims Analyses. Specifically, the OptumInsight samples of

children with ASD, the comparison group of children without ASD, and the parent samples identified were used. Task A also used data from the Impact National Benchmark Database™. However, given that the parent members were only identifiable within the OptumInsight data and that the socioeconomic data was only linkable to the OptumInsight data, study analyses under Task D focused on subjects from the OptumInsight data only.

1. Base Study Subject Eligibility Criteria

This study included commercial health plan members in the OptumInsight Research Database. To be included in the sample, individuals met the following inclusion criteria.

■ Children with ASD

Inclusion criteria:

- Commercial health plan enrolled individual with medical, pharmacy, and behavioral health coverage with at least 6 months of continuous enrollment between 01 January 2001 and 31 December 2009; the first day of the individual's first period of enrollment with all three types of coverage was set as the index dateⁱⁱⁱ
- Aged ≤ 20 years as of the index date
- At least 1 claim with an ASD diagnosis code, including Autistic Disorder, other specified PDD (including Asperger's Disorder) or unspecified PDD (ICD-9-CM 299.0x, 299.8x, 299.9x), in any position (i.e., primary or secondary position)^{iv} during enrollment between 01 January 2001 and 31 December 2009

Exclusion criteria:

- At least one claim with a diagnosis of Rett syndrome (ICD-9-CM 330.8x) in any position or childhood disintegrative disorder (CDD, ICD-9-CM 299.1x) in any position during enrollment between 01 January 2001 and 31 December 2009.^v

■ Comparison Group: Children without ASD

A general comparison group including individuals aged ≤ 20 years who did not have evidence of ASD, Rett syndrome or CDD (see diagnosis codes above) was selected.^{vi}

ⁱⁱⁱ Continuous enrollment was based on simultaneous medical, pharmacy, and behavioral health coverage. Gaps in enrollment of ≤ 32 days were bridged and included in calculation of continuous enrollment duration. Note: if a subject had more than one enrollment period with all three types of coverage, the index date was set as the first day of their first day of enrollment with all three types of coverage during the study period.

^{iv} Up to 4 diagnosis codes are recorded on physician claims and up to 9 diagnosis codes are recorded on facility claims. Primary position refers to the first diagnosis code listed; secondary position refers to any diagnosis after the first diagnosis.

^v While Rett syndrome and CDD are also considered types of pervasive development disorders similar to ASD, subjects with evidence of these disorders were excluded because these two disorders have different etiologies, disease progression and prognoses than Autistic Disorder, other specified PDD and unspecified PDD.

^{vi} An unmatched, as opposed to a matched, comparison group was selected as we felt that the large size of this unmatched comparison group would allow us to effectively employ statistical adjustment as needed for a variety of outcomes when important confounders might vary. Matching is a potentially more efficient, not a more valid, means of controlling for confounding than post hoc adjustment. The efficiency difference between matching and adjustment diminishes as available sample size increases and is greatest when there are strong confounders in play. The lack of a priori data on strong confounders for our Task B analyses coupled with the large size of the comparison group supported our decision to draw an unmatched comparison group.

The inclusion criteria for the **comparison group** were:

- Commercial health plan enrolled individual with medical, pharmacy, and behavioral health coverage with at least 6 months of continuous enrollment between 01 January 2001 and 31 December 2009; the first day of the individual's first period of enrollment with all three types of coverage was set as the index date^{vii}
- Aged ≤ 20 years as of the index date
- No evidence of ASD during enrollment between 01 January 2001 and 31 December 2009
- No evidence of Rett syndrome or CDD during enrollment between 01 January 2001 and 31 December 2009
- Not a family member of a subject with ASD

Once these individuals were identified, a random sample was selected for inclusion in the study comparison group. A sampling ratio of approximately 3 comparison subjects to 1 subject with ASD was used.

- **Family Members**

To identify subjects for the parent and sibling samples, family health plan members of both children with and without ASD were identified within the OptumInsight database using a unique system-generated family identifier variable. OptumInsight determined whether each sampled subject with ASD or comparison group member had at least one family identification (ID) value. If a subject with or without ASD had more than one family ID, OptumInsight used all family IDs associated with the subject to identify family members.

It is important to point out that the eligibility criteria for the samples of children with and without ASD were such that these samples themselves could include family members (e.g., two children with ASD within the same family could be in the sample of children with ASD). For the family member analysis, the study included family plan members assumed to be a parent, stepparent or adult domestic partner of a parent as well as family members assumed to be a sibling, step-sibling or other like child relevant to a subject with or without ASD. The family member samples did *not* include family plan members already included in the sample of children with ASD or already included in the comparison group.

In order to identify potential parents and siblings of children with ASD and of children without ASD, the *difference* between the subject's age at index date and that of his/her family members as of the subject's index date was used.^{viii} The final algorithm used to assign relationships is

^{vii} Continuous enrollment was based on simultaneous medical, pharmacy, and behavioral health coverage. Gaps in enrollment of ≤ 32 days were bridged and included in calculation of continuous enrollment duration. Note: if a patient had more than one enrollment period with all three types of coverage, the index date was set as the first day of their first day of enrollment with all three types of coverage during the study period.

^{viii} Other information within the claims data was also considered in the selection of parent and sibling samples. Relationship/dependent information (relative to the health plan subscriber) was available for many individuals with and without ASD and their family members. In a few cases, this information was detailed ("sibling," "niece/nephew," "grandchild," "stepchild"). However, the information was ultimately not used in determining the parent and sibling samples because the overwhelming majority of individuals with and without ASD were simply noted to be "child," and for the majority of family members, the available relationship/dependent information was simply another "child," "subscriber/employee," "spouse" or "domestic partner." Because detailed relationship

summarized in Table 1. Family plan members whose relative age did not meet the criteria for parent and siblings were excluded from the analysis.

Table 1. Algorithm for Identifying Parents and Siblings

Age Difference	Family Member Sample Assignment
Family member is 1-17 years younger than child with or without ASD	Sibling
Family member is 0-17 years older	Sibling
Family member is 18-49 years older	Parent
Family member is 50 or more years older	Not applicable (assumed grandparent)
Family member is 18 or more years younger	Not applicable (assumed offspring)

The final inclusion criteria for family plan members in the base study were:

- Member of the same family health plan as one of the sampled children with or without ASD
- Not a member of the sample of children with ASD or the comparison group of children without ASD
- Met the age criteria for parent or sibling relative to a sampled subject with or without ASD (see Table 1 above)
- Commercial health enrollee with medical, pharmacy, and behavioral health coverage with at least 6 months of continuous enrollment between 01 January 2001 and 31 December 2009; the first day of the family member's first period of enrollment with all three types of coverage was set as the index date^{ix}

It is possible that a sampled family member could have met the sibling criteria for one study subject and the parent criteria for another. In these cases, the family member was assigned to both family member samples.

Given Task D is focused on potential risk factors among parents and the child samples only, sibling samples were not included in the Task D base study sample. Therefore, from this point forward, the Base Study Sample description is focused solely on children with ASD, comparison children, and the associated parent samples.

2. Time Windows for Base Sample Identification and Observation

As indicated above, children with and without ASD were identified between January 2001 and December 2009. To capture the individuals' complete claims experience during the study period,

status could not be ascertained relative to the case/comparison group member, the final algorithm for the family member samples used the difference in age between the family member and case/comparison group member to determine whether a family member was assumed to be a parent or sibling relevant to the child with ASD or child without ASD. It is important to note that a family member who was classified as a sibling or parent could have been a spouse instead, that a family member classified as a parent could have been a sibling, that a family member classified as a grandparent could have been a parent, etc.

^{ix} Continuous enrollment was based on simultaneous medical, pharmacy, and behavioral health coverage. Gaps in enrollment of ≤ 32 days were bridged and included in calculation of continuous enrollment duration. Note: if a patient had more than one enrollment period with all three types of coverage, the index date was set as the first day of their first day of enrollment with all three types of coverage during the study period.

the start of the individual's first day of enrollment with simultaneous medical, pharmacy and behavioral health coverage during this time window was set as the index date. Subjects were required to have 1 period of at least 6 months of continuous enrollment during the identification window but may have had more enrollment time with all three types of coverage during the study period. Subjects with at least the minimum 6 months of continuous enrollment were studied during the time between January 2001 and December 2009. If subjects had more than 6 months of continuous enrollment *or* more than one enrollment period with simultaneous medical, pharmacy, and behavioral health coverage during the study time frame, they were studied during that additional enrollment time as well.

Parents who met the inclusion criteria were also required to have 1 period of at least 6 months of continuous enrollment between January 2001 and December 2009 and also may have had more enrollment time with simultaneous medical, pharmacy and behavioral health coverage. As with the children with and without ASD, each parent's total study observation time was the sum of all enrollment time during the study period during which the parent had all three types of coverage. It is important to note that the observation time for a sampled parent could be the same as or different than that of the subject(s) with whom that parent is affiliated. As a result, it is possible that observation time for a sampled parent may include time *before* the subject became a parent of the sampled child with or without ASD.

Task D analyses incorporated a number of other continuous enrollment criteria which are described further in Section E.2 below.

3. Refinement of ASD Base Study Sample within Task D

In Task A, eligible ASD subjects were classified into two groups: "Likely ASD" and "Possible ASD." The Likely ASD group included subjects with 2 or more medical claims with an ASD diagnosis code in any position or 1 claim with an ASD diagnosis code in a position and 1 pharmacy claim for risperidone. The Possible ASD group was defined as those children with just 1 claim with an ASD diagnosis code in any position. In the Task A: Chart Study, we conducted a medical chart review to assess the claims-based diagnoses against "gold standard" criteria. Based on the results of the Chart Study, we made two significant revisions to the ASD sample in Task D. First, we revised the Likely ASD criteria to include only children with 2 or more claims with an ASD diagnosis code in any position. The Chart Study found that a higher proportion of false positives had a prescription for risperidone than the true positives (14.3% vs. 4.4%), suggesting that risperidone may have been prescribed for conditions other than ASD. For that reason, we dropped the criteria of 1 claim with an ASD diagnosis code and 1 prescription for risperidone from the Likely ASD group definition.

Second, we dropped the Possible ASD group from the ASD sample, focusing Task D analyses on the revised definition of the Likely ASD group. In the Chart Study, the positive predictive value increased from 74.2% to 87.4% when the Possible ASD group was excluded from the case definition. Therefore, we have greater confidence that children in the Likely ASD group represent true ASD cases, and this was our Base Study Sample for the ASD group in Task D. The sampling process and study subject characteristics are presented in the next section 'Base Study Sample Size'.

4. Base Study Sample Size

Figure 1 below summarizes the identification of children with and without ASD related to the composition of the base samples used in this Task. These are the same base samples used in the other Tasks and the text description here is similar to that presented in other Task reports. Ultimately, the base sample selection process resulted in 46,236 children with ASD, although only the 33,565 in the Likely ASD category were considered for this Task, and 138,876 children without ASD identified within the OptumInsight database.

a. Children with and without ASD

To select eligible subjects for the study, first all commercial health plan enrollees with at least some type of health plan coverage between January 2001 and December 2009 were searched. Over 62 million enrollees in the OptumInsight database were identified. From these, a little over 30 million enrollees with at least 6 months of continuous enrollment with simultaneous medical, pharmacy and behavior health coverage at some point during the identification window were identified.^x Enrollees' age as of the first day of enrollment (with all three types of coverage) during the study period was calculated (based on year of birth).

Among the 30 million enrollees meeting the above criteria, individuals whose age was 20 years or younger were retained. Individuals with evidence of Rett or CDD were then excluded.^{xi} The resulting 9.5 million children comprised the sampling frame from which children with and without ASD were identified for the study. Ultimately, the sample selection process as implemented in Task A: Baseline Claims Analysis resulted in 46,236 children with ASD and 138,876 children without ASD (selected using an approximate sampling ratio of 3:1) identified within the OptumInsight database.

b. Parents

Approximately 99% of the children with and without ASD had evidence of being in a family health plan, and for all but approximately 2% of these subjects, at least one family plan member was identified within the database. The number of unique family plan members identified among all children with and without ASD was over 614,000. Specifically, 147,083 family plan members were identified for children with ASD (an average of 3.18 per subject), and 467,764 were identified for the comparison group (an average of 3.37 per subject).

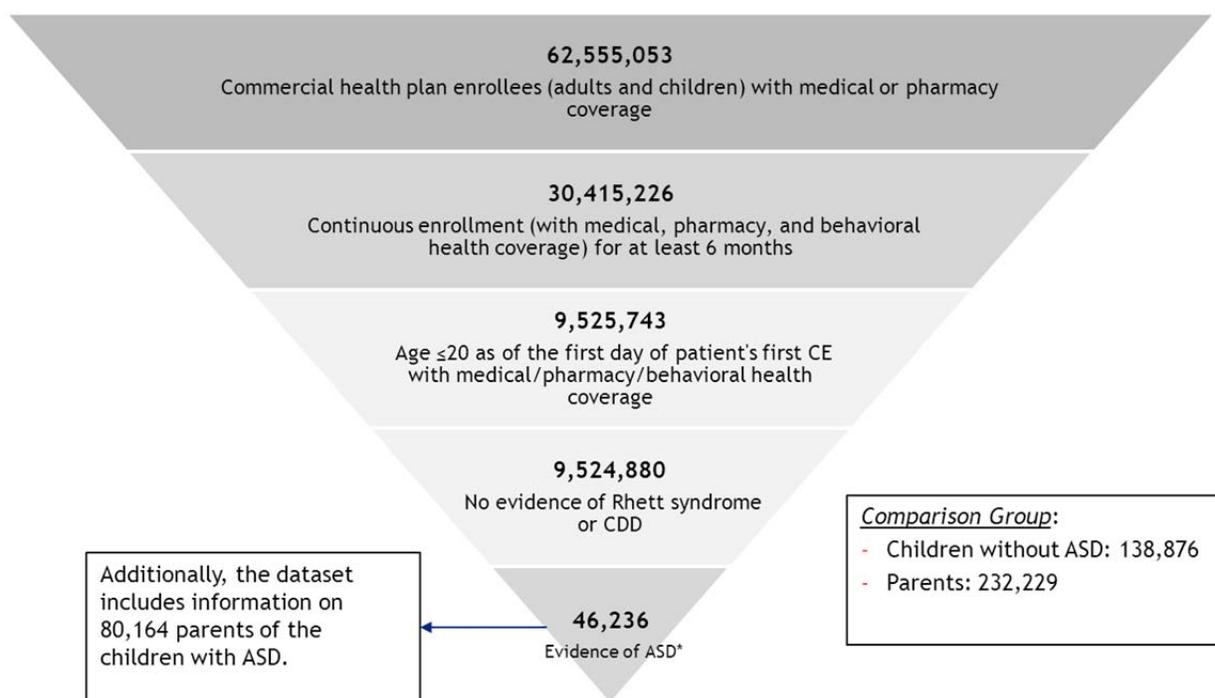
To select family plan members eligible for the study, family plan members with at least 6 months of continuous enrollment with simultaneous medical, pharmacy and behavior health coverage during the identification window of 2001 through 2009 were flagged. Family plan members who

^x While all subjects sampled for the study were required to have at least 6 months continuous enrollment, sample members were not required to have medical claims during their study enrollment time, with the exception of children with ASD (whose ASD diagnosis necessitated at least 1 medical claim). It is important to note that a subset of comparison subjects (12.4%) and a subset of members of the parent samples (10.1% of comparison parents, 4.0% of ASD parents) had no medical claims during their study observation time. Basic demographic information was available for these subjects, but, by definition, these subjects lack evidence of any of the analyzed outcomes as well as have no utilization and health care costs during the study period. Therefore, while the children with ASD sampled inherently were "health care users," the other samples, including both the comparison group and parent samples, included some "non-users."

^{xi} Of the 1,432 patients with at least one claim for Rett or CDD, approximately 60% had a claim for Rett, and approximately 40% had a claim for CDD. Very few (<1.0%) had claims for both.

met this requirement (n=568,198) represented 92% of all family members identified. From these, a tiny subset of family plan members who were linked (through system-generated family plan identification numbers) back to both children with and without ASD (n=78) were omitted.^{xiii} Finally, the age criteria outlined in Table 1 were applied to identify assumed “parents” of children with and without ASD. A total of 312,393 family plan members were designated as parents (80,164 for children with ASD and 232,229 for the comparison group).

Figure 1: Sampling Process as in Task A: Claims Based Analysis



*presence of one or more claims with an ICD-9 for Asperger's, Autism, or PDD-NOS

c. Refinement of ASD-Related Base Study Samples in Task D

In Task A: Baseline Claims Analysis, eligible ASD subjects were classified into two groups: “Likely ASD” and “Possible ASD.” The Likely ASD group included subjects with 2 or more medical claims with an ASD diagnosis code in any position or 1 claim with an ASD diagnosis code in any position and 1 claim for risperidone. The Possible ASD group included those children with just 1 claim with an ASD diagnosis code in any position.

As described above, the sample for Task D used a revised definition of the Likely ASD sample (limiting this group to only those with two ASD claims), and the Possible ASD group was excluded from not only the ASD samples but also the comparison groups in Task D. **Table 2** shows the impact of these changes on the sample of children with ASD as well as affiliated parents.

^{xiii} While comparison group members could not be a family member of an individual with ASD, 78 family members identified had family IDs that linked back to a member of both samples and were thus excluded from the study.

Table 2. Likely vs. Possible ASD Children and Affiliated Parents

	Total ASD		Parents of ASD Group	
	n	%	n	%
Total Number of Subjects in Sample	46,236	100.00	80,164	100.00
Likely ASD Subject	33,565	72.59	58,757	73.30
Possible ASD Subject	12,671	27.41	21,407	26.70
Final Sample Used in Analysis	33,565	72.59	58,757	73.30

Note: Likely ASD children include children with 2 or more claims with ASD diagnosis in any position. Possible ASD children include children with only one claim with ASD diagnosis in any position.

For Task D, the base ASD-related samples used were 33,565 children with ASD and 58,757 parents of children with ASD. Further continuous enrollment requirements as well as additional parent linkage criteria were applied to both the ASD-related and comparison Task D base samples to create the various subsamples of interest. These criteria and associated subsamples are further described in Section E.1 and E.2.

D. Variable Definitions

The variables described below include patient characteristics, patient comorbidity measures, and potential risk factors that are used in our analyses.

1. Patient Characteristics

- **Index date.** A subject's first date of enrollment.
- **Age at index year.** Using subjects' date of birth, subjects' age in years as of the year of the index date – i.e., the start of study enrollment. The definition of this variable was slightly revised from that of the Task A where age at index year was determined based on the subjects' year of birth as opposed to actual date of birth. Other age-related variables as described below were also modified so that they are based on subjects' actual date of birth in Task D.
- **Age group at index year.** Subjects' age group as of the index date. Subjects with and without ASD were categorized <2, 2-10, 11-17, and 18-20 years. Parents were categorized as <18, 18-21, 22-29, 30-49, 50-64, and 65+ years.
- **Age at first ASD claim:** Created for ASD group only. The subject's age in years as of the first medical ASD claim (in any position) during the study period. Age in years at first medical claim with ASD diagnostic code was calculated as year of birth subtracted from the year of the earliest ASD claim within the study period.
- **Gender.** Gender from enrollment data.
- **Geographic location.** The United States region in which the study subject was enrolled in a health plan as of the index date. States were categorized into geographic regions in accordance with the U.S. Census Bureau's region designations. The regions are presented below in Table 3.

Table 3. Geographic Regions

Census Region	Census Division	State
Northeast	New England	CT MA ME NH RI VT
	Mid Atlantic	NJ NY PA
Midwest	East North Central	IL IN MI OH WI
	West North Central	IA KS MN MO ND NE SD
South	South Atlantic	DC DE FL GA MD NC SC VA WV
	East South Central	AL KY MS TN
	West South Central	AR LA OK TX
West	Mountain	AZ CO ID MT NM NV UT WY
	Pacific	AK CA HI OR WA

- **Total enrollment time during study.** The sum of the number of days of enrollment during the index continuous enrollment period and additional continuous enrollment periods. For subjects with multiple enrollment periods, one or more gaps in enrollment were present during this time. The length of these gaps was not included in the calculation of total enrollment time (unless the gap was less than 33 days and was thus considered part of the continuous enrollment period).
- **Household income.** Modeled household income from the linked socioeconomic data. Available categories included: Under \$15,000 , \$15,000 - \$19,999, \$20,000 - \$29,999, \$30,000 - \$39,999, \$40,000 - \$49,999, \$50,000 - \$59,999, \$60,000 - \$74,999, \$75,000 - \$99,999, \$100,000 - \$124,999, \$125,000 - \$149,999, \$150,000 - \$249,999, and \$250,000+. For our analyses, these groups were further collapsed into five categories: <\$50,000, \$50,000 - \$74,999, \$75,000 - \$99,999, \$100,000 - \$124,999, and \$125,000+. This variable depended on the successful linkage with and the availability of information within the socioeconomic database. Data were therefore missing for some study subjects. Subjects with missing data were categorized as “unknown.”
- **Race/ethnicity.** Available categories included: White, African-American/Black, Native Hawaiian or Pacific Islander, American Indian or Alaskan Native, Asian, Hispanic or other. Because of smaller sample sizes, Native Hawaiian or Pacific Islander and American Indian or Alaskan Native were combined with the other category to form a combined “other” category. This variable depended on the successful linkage with and the availability of information within the socioeconomic database. Data were therefore missing for some study subjects. Subjects with missing data were categorized as “unknown.”
- **All-cause health care office visits.** A count of a subject’s office visits (e.g., physician offices, health clinics) was calculated. Office visits were calculated as 1 per provider per day. To adjust for varying observation time, annualized counts are presented.

2. Patient Comorbidity Scores

The comorbidity measures were measured based on a child's or parent's total number of days with medical/pharmacy/behavioral health enrollment during the study period.

- **Quan-Charlson comorbidity score.** A comorbidity score calculated based on the presence of diagnosis codes on medical claims (see Quan et al, 2005²²). Scores ranged from 0 (no comorbidity) to 29 (high comorbidity). Given that this measure was developed and validated with adults, this score was calculated only for the parent samples.
- **Overall comorbidity score for children.** There is currently no comorbidity measure for claims analysis that is universally recognized or used for children. To capture overall comorbidity for the child samples in our study, we calculated a comorbidity score modeled after Feudtner et al. 2000,²³ a comorbidity score based on the presence of diagnosis codes on medical claims for the child samples (children with and without ASD and siblings). For each subject, a dichotomous flag (0/1) was created for each of 9 categories of chronic conditions: 1) neuromuscular, 2) cardiovascular, 3) respiratory, 4) renal, 5) gastrointestinal, 6) hematologic or immunologic, 7) metabolic, 8) other congenital or genetic defect, and 9) malignant neoplasms. For each category, a subject was coded 1 if he or she had at least one claim for a diagnosis in any position for a condition within the category. These flags were then summed, which resulted in a possible score ranging from 0 to 9. While the results of this score were in line with expectations, it is important to acknowledge that the measure has not been formally validated for claims analysis. The relevant ICD-9-CM codes for each of the nine categories are provided in Appendix B.

3. Potential Risk Factors

Given that the etiology of ASD is not clearly known, a wide set of risk factors has been proposed by the research community, including factors as diverse as prenatal infections, health conditions such as epilepsy, psychiatric conditions such as anxiety, prescription medication use, and many others.²⁴ The particular factors we have selected reflect a thorough literature review and careful consideration of how best to use claims data to measure such factors. Our selected potential risk factors span the preconception, prenatal, early postnatal and early childhood periods. In general, we selected factors that we believed could be well-captured by claims data; although we did include a limited set of potential risk factors that we suspected would not be captured well as a check on our assumptions (e.g., chronic conditions that generate intermittent claims activity or are not well documented or reimbursed such as autoimmune disease and obesity).

The approach used to measure each risk factor is described below. Some potential risk factors were measured in more than one way. For example, when examining risk factors that required an ICD-9-CM code on a medical claim, some risk factors were measured with a minimum of 1 or 2 claims. For risk factors that could be represented by a medical claim with an ICD-9-CM code and a pharmacy claim for a medication, they were measured with and without pharmacy claims. The specific codes used for each risk factor are provided in Appendix B.

We categorized potential risk factors into the following groups: 1) early postnatal and postnatal window risk factors (measurable from child claims); 2) prenatal window risk factors (measurable from maternal claims); and 3) preconception window risk factors (measurable from maternal

and/or paternal claims). Potential risk factors were measured in both the ASD and comparison samples.

- **Early postnatal and postnatal window risk factors (measurable from child claims).** For purposes of this study, early postnatal was defined as birth to (birth plus one week). Postnatal was defined based on three different time periods: birth to (birth plus 12 months); birth to (birth plus 24 months); and birth to (birth plus 60 months).
 - **NICU/ICU admission.** Whether the child had evidence of a NICU admission in the early postnatal period was determined (yes/no) using CPT and HCPCS codes (See Appendix B). Whether there was an ICU admission during the first week after birth was also determined. A flag was created for either a NICU or ICU admission as well as a flag for each one separately. See Appendix B for CPT codes.
 - **Non-singleton pregnancy.** Whether the child had evidence of being a non-singleton pregnancy was determined (yes/no) using ICD-9-CM codes (a single medical claim with an ICD-9-CM code in any position). See Appendix B for conditions in the “non-singleton” designation and corresponding ICD-9-CM codes.
 - **Hypoxia.** Whether the child had evidence of hypoxia in the early postnatal period (yes/no). ICD-9-CM codes (a single medical claim with an ICD-9-CM code in any position) for hypoxia were examined. See Appendix B for conditions included and corresponding ICD-9-CM codes.
 - **Prematurity/pre-term delivery.** Whether the child had evidence of prematurity/pre-term delivery was determined. Both ICD-9-CM and CPT codes for prematurity and ICD-9-CM and CPT codes for other conditions and procedures common among premature infants were used. See Appendix B for conditions included and corresponding ICD-9-CM and CPT codes.
 - **MMR vaccine at age 0-24 months.** Whether or not a subject had evidence of complete MMR vaccination during this age period was flagged (yes/no). Complete MMR vaccination was measured by five combinations of CPT codes for MMR vaccination: 1) separate claims for measles, mumps, and rubella vaccinations; 2) a claim for measles and rubella combination vaccination and a separate claim for mumps vaccination; 3) a claim for mumps and rubella combination vaccination and a separate claim for measles vaccination; 4) a claim for a MMR combination vaccination; and 5) a claim for an MMR plus varicella combination vaccination. A flag was set for each of the five indicating complete MMR vaccination so that the manner in which a subject was vaccinated could be determined. See Appendix B for the list of CPT codes used.
 - **Non-vaccinated status.** Children with no evidence of childhood vaccinations were identified. The number and percentage of children without evidence of any vaccination in both the ASD and comparison subsamples was identified for each of the following age periods, if available: birth to 12 months, birth to 24 months, and birth to 60 months. See Appendix B for vaccinations included and corresponding CPT codes.

- **Prenatal window risk factors (measurable from maternal claims).** For purposes of this study, prenatal was defined as ((birth minus 40 weeks) to (birth minus 1 day)). The prevalence of some risk factors were also measured during specific trimesters: first (conception to (conception plus 90 days)), second ((conception plus 91 days) to (conception plus 181 days)), and third ((conception plus 182 days) to birth) and some were measured in early postnatal period (birth to one week after birth).
 - **Maternal asthma.** Whether the mother had evidence of asthma from conception (birth minus 40 weeks) to birth was determined using both ICD-9-CM codes and pharmacy codes for asthma medications (See Appendix B). Separate flags were created based on three different asthma definitions: one or more medical claims with an ICD-9-CM code for asthma in any position; two or more medical claims with an ICD-9-CM code for asthma in any position 30+ days apart; OR one medical claim with an ICD-9-CM code for asthma in any position and a prescription for an asthma medication. As asthma is a chronic condition and could be present before the child was born, before the pregnancy, or after the child was born, its presence was also examined from the mother's index date to child's birth, and child's birth to end of enrollment for the mother. When stratifying the use of B2Ars and non-B2Ar asthma medication by an indication of asthma (discussed further below), the definition of 1+ diagnosis was used to define having the indication for asthma. See Appendix B for conditions included and corresponding ICD-9-CM codes and prescription drugs.
 - **Maternal depression.** Whether a mother had evidence of depression from conception (birth minus 40 weeks) to birth was determined using ICD-9-CM codes. Separate flags were created based on two different definitions: one or more medical claims with an ICD-9-CM code for depression in any position; two or more medical claims with an ICD-9-CM code for depression in any position 30+ days apart. As depression is a chronic condition and could be present before the child was born, before the pregnancy, or after the child was born, its presence was also examined from mothers' index date to child's birth, and child's birth to end of enrollment for the mother. When stratifying the use of SSRIs and non-SSRI depression medication by an indication of depression (discussed further below), the definition of 1+ diagnosis was used to define of having the indication for depression. See Appendix B for conditions included and corresponding ICD-9-CM codes.
 - **Maternal autoimmune conditions.** Whether a mother had evidence of an autoimmune condition prior to conception [from (birth minus 92 weeks) to conception (birth minus 40 weeks)] and/or during the pregnancy [from (birth minus 40 weeks) to birth] was determined using ICD-9-CM codes. Separate flags were created for a single medical claim with an ICD-9-CM code for an autoimmune condition in any position and for 2 or more medical claims for an autoimmune condition in any position 30+ days apart. As autoimmune conditions are a chronic condition and could be present before the child was born, before the pregnancy, or after the child was born, its presence was also examined from mothers' index date to child's birth, and child's birth to end of enrollment for the mother. See Appendix B for conditions included and corresponding ICD-9-CM codes.

- **Maternal prenatal infection.** Whether a mother had evidence of a prenatal infection from conception (birth minus 40 weeks) to birth was determined using ICD-9-CM codes (a single medical claim with an ICD-9-CM code in any position). The trimester in which the infection occurred was also determined. See Appendix B for conditions included and corresponding ICD-9-CM codes.
- **Maternal prenatal obesity.** Whether a mother had evidence of obesity from conception (birth minus 40 weeks) to birth was determined using ICD-9-CM codes (a single medical claim with an ICD-9-CM code in any position). See Appendix B for conditions included and corresponding ICD-9-CM codes.
- **Maternal gestational diabetes.** Whether a mother had evidence of gestational diabetes from conception (birth minus 40 weeks) to birth was determined using ICD-9-CM codes (a single medical claim with an ICD-9-CM code in any position). For mothers without evidence of diabetes in the preconception period (specifically no claims with a diagnosis code for diabetes or pharmacy claims for diabetic medications), diagnosis codes for diabetes or gestational diabetes were used as well as 1 or more pharmacy claims for a diabetic medication. See Appendix B for conditions included and corresponding ICD-9-CM codes and prescription drugs.
- **Maternal preeclampsia/eclampsia.** Whether a mother had evidence of preeclampsia/eclampsia from conception (birth minus 40 weeks) to birth was determined using ICD-9-CM codes (a single medical claim with an ICD-9-CM code in any position) and medications. See Appendix B for conditions included and corresponding ICD-9-CM codes.
- **Maternal anesthesia.** Whether a mother had evidence of anesthesia use in the prenatal ((birth minus 40 weeks) to birth) or early postnatal (birth to birth plus one week) periods was determined using ICD-9-CM (a single medical claim with an ICD-9-CM code in any position) or CPT procedure codes. See Appendix B for conditions included and corresponding ICD-9-CM and CPT codes.
- **Prenatal ultrasound.** Whether a mother had evidence of an ultrasound from conception (birth minus 40 weeks) to birth was determined. Both CPT and HCPCS codes were examined. The number of ultrasounds and the trimester in which the service occurred were also determined. See Appendix B for CPT codes.
- **Maternal prenatal medication use.** Whether a mother had evidence of prenatal medication use was determined. Medications were examined by class- SSRIs (selective serotonin reuptake inhibitors), B2Ars (Beta-2 adrenergic receptors), contraceptives (separate flags for oral vs. non-oral), anticonvulsants, and antibiotics. Use of medication was identified with a single prescription for the class of drugs identified. For some medications, we also determined the trimester(s) in which the prenatal medication were used. Pitocin use was measured during early postnatal period (birth to birth plus one week) to capture use at delivery. See Appendix B for prescription drugs.
- **Non-SSRI depression medication use.** Whether a mother had evidence of prenatal non-SSRI depression medication use was examined. While this is not a risk factor in this study, it allowed a comparison of the proportion of mothers

who used non-SSRI depression medications to the proportion of mothers with an SSRI medication. Use of medication was identified with a single prescription for the class of drugs identified. We also determined the trimester(s) in which the medication was used. See Appendix B for prescription drugs.

- **Non-B2Ar asthma medication use.** Whether a mother had evidence of prenatal asthma medication use (excluding non-B2Ars) was examined. While this is not a risk factor in this study, it allowed a comparison of the proportion of mothers who used Non-B2Ar asthma medications to the proportion of mothers with a B2Ar medication. Use of medication was identified with a single prescription for the class of drugs identified. We also determined the trimester(s) in which the medication was used. See Appendix B for prescription drugs.
- **Preconception window risk factors (measureable from maternal and/or paternal claims).** For purposes of this study, preconception was defined as (date of conception minus 52 weeks or birth minus 92 weeks) to date of conception.
 - **Maternal preconception obesity.** Whether the mother had evidence of obesity prior to conception [from (birth minus 92 weeks)] was determined using ICD-9-CM codes (a single medical claim with an ICD-9-CM code in any position). See Appendix B for conditions included and corresponding ICD-9-CM codes.
 - **Paternal autoimmune conditions.** Whether the father had evidence of an autoimmune condition from 1 year prior to conception (birth minus 92 weeks) to conception (birth minus 40 weeks) was determined using ICD-9-CM codes. Separate flags were created for a single medical claim with an ICD-9-CM code for an autoimmune condition in any position and for 2 or more medical claims for an autoimmune condition in any position 30+ days apart. As autoimmune conditions are a chronic condition and could be present before the child was conceived or after the child was born, its presence was also examined from fathers' index date to child's conception, and child's conception to end of enrollment for the father. See Appendix B for conditions included and corresponding ICD-9-CM codes.
 - **Maternal infertility treatment.** Whether the mother had evidence of infertility treatment from 1 year prior to conception (birth minus 92 weeks) to conception (birth minus 40 weeks) was determined. Separate flags were created for a single medical claim with an ICD-9-CM for infertility in any position, a single medical claim with a HCPCS code for infertility, and for medications used for infertility (some medications were only counted if an ICD-9-CM for infertility was also present during the time prior to conception (birth minus 92 weeks) to conception (birth minus 40 weeks). See Appendix B for conditions included and corresponding ICD-9-CM, CPT codes, and prescription drugs.
 - **Paternal infertility treatment.** Whether the father had evidence of infertility treatment from 1 year prior to conception (birth minus 92 weeks) to conception (birth minus 40 weeks) was determined using ICD-9-CM codes. Separate flags were created for a single medical claim with an ICD-9-CM for infertility in any position, a single medical claim with a HCPCS code for infertility, and for medications used for infertility. See Appendix B for conditions included and corresponding ICD-9-CM, CPT codes, and prescription drugs.

E. Methodology

This section outlines our methodology for analyzing the data to meet the study objectives outlined in Section II. Methods are summarized by objective.

- 1. Evaluate the frequency with which the apparent mother and an apparent father can be identified within claims data and the representativeness of the subsample with linked parents compared to the base study sample.*

The first objective of this task was to evaluate the frequency with which apparent mothers and fathers could be identified in claims data. As described above under Section III.C and, in more detail, in the Final Study Report for Task A: Baseline Claims Analyses, children and mothers were initially linked via a family identification code with difference in age between the child with ASD or the comparison child and his/her family plan member used to infer initial family relationship. To refine this, we relied on an approach taken by the STORK²⁵ study. In the STORK study, maternal delivery dates and child's date of birth were examined to provide further evidence of the maternal-offspring relationship. Consistent with STORK, we required the ASD and comparison child's date of birth to be within +/- seven days^{xiii} of the mother's delivery date to be considered a match. Any infants linked to more than one mother were considered a non-match and were removed from the subsequent analyses reported here. This approach, while providing an additional measure of confidence in the maternal-offspring match, also serves to limit the identified mother-child pairs to those for whom information from both mother and child was available in claims around the time of birth (a 'midpoint' of sorts for the etiologic windows of interest).

Linkages with fathers are of most use in risk factor investigations if claims data are available on fathers during the preconception window (the time period where paternal claims data would be most relevant to ASD risk). Consequently we considered only father-offspring pairs where fathers were enrolled at conception.

Date of conception was considered (delivery date minus 280 days). The first trimester and second trimester cut-points were (delivery date minus 189 days) and (delivery date minus 98 days), respectively. In cases where there was evidence of a non-singleton pregnancy or pre-term delivery in either the mother or child's claims, the date of conception and the first and second trimester cut-points were adjusted as outlined in Table 4 below. Using these dates we explored three alternative windows around our estimated conception date to require verification of paternal enrollment (i.e., a potential father had to be enrolled for a certain number of days prior to and after the estimated date of conception. These were +/- 7, 14 and 21 days, respectively.

^{xiii} A range of 7 days is used because of the potential delay in processing of claims. A more strict range of days may be considered.

Table 4. Initial Conception and Trimester Cut-points by Prematurity Category

Description	ICD-9-CM Code	Initial Conception Date (from delivery date)	Trimester Cutpoint 1	Trimester Cutpoint 2
Standard pregnancy	N/A	- 280 days	- 189 days	- 98 days
Non-singleton pregnancy	651.0-651.6, 660.50	- 252 days	- 161 days	- 70 days
Prematurity				
Unspecified	765.20	- 252 days	- 161 days	- 70 days
<24 weeks	765.21	- 160 days	- 69 days	N/A
24 weeks	765.22	-168 days	- 77 days	N/A
25-26 weeks	765.23	- 178 days	- 87 days	N/A
27-28 weeks	765.24	- 192 days	- 101 days	- 10 days
29-30 weeks	765.25	- 206 days	-115 days	- 24 days
31-32 weeks	765.26	- 220 days	- 129 days	- 38 days
33-34 weeks	765.27	- 234 days	- 143 days	- 52 days
35-36 weeks	765.28	- 248 days	- 157 days	- 66 days

In addition to considering the size of the children base study subsamples that link to mothers and fathers, we also explored the representativeness of these groups by examining the distribution of demographic characteristics between the identified children subsamples and the base sample (“Likely ASD”) population. In addition, we make comparisons to the subsample of the base population who were enrolled at birth since the algorithms used to refine linkage of children to parents required subject enrollment at birth.

2. *Identify the size of available subsamples extracted from the base sample meeting varying criteria for continuous enrollment over etiologic windows of potential interest.*

In order to explore how subsample availability would affect studies over the full range of potential risk factors of interest, we identified and enumerated subsamples from the base study population that would potentially be selected for various studies focusing on different preconception, prenatal, early postnatal and early childhood risk windows. The subsamples we considered, for both our ASD and comparison samples, were as follows:

■ **Child subsamples:**

- Continuous enrollment^{xiv} from birth to ≥ 24 months after birth. This is a group that would be selected for a study of child exposures that might be etiologically relevant over the first two years of life. The size of the available subsample was determined with and without linkages to an apparent mother and/or an apparent father using the Task D methodology for identifying parents described in the prior section.

^{xiv} Enrollment is defined based on a patient’s date joining the health plan and the date they left the health plan. A patient is considered continuously enrolled if there is no gap in enrollment of more than 32 days.

- Continuous enrollment from birth to ≥ 12 months after birth. This group would be selected for a potential study of child exposures that might be etiologically relevant over the first year of life. The size of the available subsample was determined with and without linkages to an apparent mother and/or an apparent father using the Task D methodology for identifying parents described in the prior section.
 - Continuous enrollment from birth to ≥ 60 months after birth. This group would be selected to estimate the proportion of children with no evidence of vaccination from claims. The ≥ 60 months after birth provided an expanded time period in which to better identify children with no evidence of vaccination during the time period during which many vaccinations and their boosters are recommended to be received. Any evidence of vaccination at any time from birth to 60 months (including the period of ≥ 24 months after birth) was identified. The size of the available subsample was determined without linkages to an apparent mother and/or an apparent father using the Task D methodology for identifying parents described in the prior section.
- **Mother subsamples:**
- All mother subsamples described here are drawn from the group of mothers that linked to children based on the linkage requirement described in the prior section (which followed the STORK approach that required a match between child's date of birth and date of delivery which, in turn, means that both mother and child needed to be enrolled around the time of the child's birth).
- Continuous enrollment from the third trimester (birth minus 14 weeks) to birth^{xv}. This group would be selected in a study interested primarily in late pregnancy risk factors.
 - Continuous enrollment from (birth minus 40 weeks) to birth. This group would be selected in a study that was interested in exposures that could occur, and might be etiologically significant, at any point during pregnancy.
 - Continuous enrollment from (birth minus 92 weeks) to birth. This group would be selected in a study that was interested in exposures that could occur, and might be etiologically significant, a year prior to pregnancy or at any point during pregnancy.
 - Continuous enrollment from (birth minus 40 weeks) to (birth plus 12 months). This group would be selected in a study that was interested in exposures that could occur, and might be etiologically significant, at any point during pregnancy and also follow the mother one year after the birth of the child.

^{xv} This subsample was reviewed to determine the continuous enrollment length among the subjects. This allowed us to see if mothers tend to appear in claims around a specific time during the pregnancy.

- **Father subsamples:**

All father subsamples described here are drawn from the group of fathers that linked to children based on the linkage requirement described in the prior section (enrollment at the time of estimated date of conception).

- Continuous enrollment in father from (birth minus 92 weeks) to (birth minus 40 weeks). This group would be selected in a study that was interested in potential risk factors in the father during the preconception period.

3. *Explore the feasibility of using claims-based information on scheduled obstetric procedures to more accurately identify conception and trimester cut points.*

In addition, one of the analytic goals of this task was to explore the feasibility of refining approaches to determining conception and trimester cut-points. To do this we first determined the proportion of our mother samples that have a code for each obstetric procedure listed in Appendix C. Next, we determined the trimester in which the procedure took place based on our current trimester cut-points. Based on the results, we could then determine the feasibility of using these procedures codes in an algorithm to accurately define alternative conception and trimester cut-points in future analyses.

4. *Explore the feasibility of identifying and measuring selected potential ASD risk factors measurable in claims data.*

The final objective of this study was to determine the feasibility of using claims data to measure a select set of potential ASD risk factors. Potential risk factors were considered in the most relevant subject (mother, father and/or child) and in the relevant etiologic time windows. When alternate definitions (e.g. requiring one diagnosis vs. two diagnosis codes minimum) of potential risk factors were explored, prevalence estimates are shown for both definitions. In addition to providing estimates on the prevalence of potential ASD risk factors, we also took note of challenges met and assumptions made during this process. Included are descriptions of decisions made around potential risk factors excluded from consideration because of concerns over our ability to operationalize variable definitions in claims (i.e. whether the limitation was lack of coverage, bundling with other procedures, etc.) and explanations of assumptions made in coding risk factor indicators (e.g. numbers of claims with a particular code required, whether gaps of fixed length between codes were used, etc.).

This task was not intended to include a comparison of risk factor data from claims to any external gold standard data (e.g., medical record review) – contrary to what was done in Task A Chart Study with respect to ASD diagnosis. Consequently, in order to provide some information on the potential adequacy of claims-based risk factor measurement, claims-based risk factor prevalence estimates were compared to available data in the published medical literature. A search of both the peer-reviewed and gray literature was conducted to locate published potential risk factor prevalence estimates in relevant populations (ASD-affected (“ASD”), ASD unaffected (“non-ASD”), and the general population of children, mothers, and fathers (“national”). This literature search was not meant to be exhaustive; rather, the objective of our search was to obtain a range of values for each of the specific risk factors. However, if only one estimate was available from the literature for a given risk factor it was included in our results. Given the literature search was not

exhaustive, caution needs to be used in interpreting comparisons to our data, especially when only one estimate from the literature is cited.

To conduct the literature review we built upon the initial literature review conducted during the early stages of this task which had been conducted to assist with selecting candidate potential risk factors. This initial review, conducted in Fall 2011, identified 164 articles. We re-examined these articles and found 16 articles that could contribute reference prevalence estimates for one or more subsamples of interest. A subsequent highly focused search of the peer-reviewed literature via PubMed was then conducted, using medical subject headings and key text associated with the potential ASD risk factors identified in the initial review. While more recent articles and US-based estimates were preferred, older and international studies were used to provide estimates when no recent or US-based studies were found. In addition, while studies using claims data were of particular interest, all estimates were included regardless of the data source. Additionally, a search of the gray literature, primarily reports from government-sponsored surveillance programs, identified other sources that reported US general population estimates. Finally, for those potential risk factors that only had one available estimate, we reviewed the references cited in the article providing that estimate in order to locate additional data, if it existed. Altogether we identified 66 sources that reported relevant estimates of risk factor prevalence for the ASD, non-ASD, or general population. A summary of these sources is presented in Section IV.D Identification and Measurement of Potential Risk Factors.

IV. Results

The results below are presented by study objective. Please note that for some tables, a valid N is presented. This represents the total number of subjects upon which the specific proportion presented is based. In some cases, the valid N is different than the Column total N due to the approach to identifying the subsamples when the child is very premature. For example, a mother of a premature child is included in the prenatal period Valid N, but would not be included in the third trimester Valid N as the child was born during the second trimester. These instances are rare, but they do result in differences in sample size depending on the time period of interest. In cases where a valid N is not shown, the specific percentage is based upon the total N for the column.

OBJECTIVE 1: Identification of apparent mothers and fathers and representativeness of children samples with identified parents

Table 5 presents the sample selection for the mothers and fathers of the ASD and comparison groups following the above-described approach that emphasized the identification of parents with enrollment at the time of the child's birth. In the sequence of steps used to identify mothers, the proportions of potential mothers retained at each step were similar in the ASD and comparison groups with the final proportion of children with mothers identified quite similar at 4.1% and 4.9%, respectively. This similar result underscores that the identification of children in claims data who can be linked to a mother with usable claims experience in that window is quite limited in terms of percentages of total population. However, the process has similar yield for ASD cases and comparison group mothers, suggesting that the approach to identifying mothers is not an overt source of selection bias. When identifying fathers, requiring a longer enrollment period around the estimated date of conception made little difference in the proportion of fathers identified (consequently a requirement of +/- seven days was used in all subsequent analyses). Overall the finding was similar to that for mothers in that only a small percentage of fathers with enrollment around conception were identified overall but proportions were similar across ASD and comparison groups.

Table 5. ASD and Comparison Group Mothers and Fathers Sample Selection using the STORK Methodology for Mothers and Enrolled at Conception for Fathers

	ASD		Comparison	
	n	%	n	%
Female family members	53,378	100.00	216,827	100.00
Mothers identified via FAMID and age differences (Task A)	31,329	58.69	119,143	54.95
Mothers with children born between 2001-2009	13,485	25.26	48,171	22.22
Children enrolled at birth*	3,400	6.37	17,959	8.28
Mothers with evidence of enrollment +/- 7 days of child birth date	2,807	5.26	16,066	7.41
Mothers identified via STORK methods**	2,176	4.08	10,703	4.94
Male family members	50,507	100.00	215,960	100.00
Fathers identified via FAMID and age differences (Task A)	30,191	59.78	115,829	53.63
Fathers with children born between 2001-2009	8,109	26.86	28,202	24.35

	ASD		Comparison	
	n	%	n	%
Enrolled at conception (+/- 7 days)***	1,513	5.01	7,204	6.22
Enrolled at conception (+/- 14 days)***	1,486	4.92	7,110	6.14
Enrolled at conception (+/- 21 days)***	1,462	4.84	7,004	6.05

*Enrolled within 32 days of birth; **Mothers identified via STORK method will be mother sample used in all tables going forward (not required to be linked to child with continuous enrollment requirement), no CE requirements where implemented in the selection of this population. The STORK method identifies mothers for 12,814 unique children.*** This table presents number of mother/child or father/child combinations found in the dataset. The full female population is 270,205 (Case: 216,827/ASD: 53,378), this represents 265,336 unique females. The full male population is 266,467 (Case: 215,960/ASD: 50,507), this represents 261,749 unique males.

In order to further explore the possibility that the selection of samples based on criteria related to the ability to link to parents resulted in selection bias, the distribution of demographic and socioeconomic characteristics and comorbidity measures in the ASD and comparison children subsamples linked to parents using the Task D methodology was compared to the relevant Task A “Likely ASD” base samples. As mentioned, for the Task D children with ASD subsample, we show the distribution relative to the Task A “Likely ASD” base sample and to those children in the Task A “Likely ASD” base sample who were enrolled at birth.

Table 6 shows these comparisons for ASD children. As expected, children linked to parents had markedly younger age at index than did Task A base children in general but, again as expected, similar age at index to the Task A base children enrolled at birth. Gender, race, income, and geographic distributions were generally similar across all groups. ASD children linked to parents had a longer (nearly one year) average period of continuous enrollment than the Task A base Likely ASD children but similar enrollment to the subsample of the Task A base children enrolled at birth. ASD children linked to parents had more annual office visits (mean of 98 vs. 57) and comorbidity (e.g., 9.50% vs. 4.95% with scores ≥ 3) than did the Task A base but these differences were not apparent when the Task A base was limited to those enrolled at birth suggesting that the differences compared to the base population are related to longer length of enrollment and the inclusion of enrollment early in life. Findings were similar for the Comparison group (Table 7). Office visits and comorbidity levels were, as expected, lower in the Comparison group than in the ASD children but the contrasts between those linked to parents and both the Task A “Likely ASD” base and Task A “Likely ASD” base enrolled at birth were similar.

Table 6. Demographic Characteristics of ASD Group with Linked Parents

	ASD Children with Identified Mother (N=2,168)			ASD Children with Identified Father (N=1,499)			ASD Children with Identified Mother or Father (N=2,564)			ASD Children with both Parents Identified (N=1,103)			Task A Likely ASD Enrolled at Birth (N=2,227)			Task A Likely ASD (N=33,565)		
	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI
Age at first ASD diagnosis in claims data	3.42	3.35	3.48	3.30	3.23	3.38	3.44	3.38	3.50	3.22	3.13	3.30	3.34	3.27	3.40	8.84	8.78	8.89
Age at index date	0.35	0.31	0.39	0.33	0.28	0.38	0.42	0.38	0.46	0.16	0.12	0.20	0.00	-	-	6.73	6.68	6.78
		n	%		n	%		n	%		n	%		n	%		n	%
Age at group at index year																		
<2		1,940	89.48		1,361	90.79		2,248	87.68		1,053	95.47		2,227	100.00		5,609	16.71
2-8		228	10.52		138	9.21		316	12.32		50	4.53		0	0.00		19,987	59.55
9-17		0	0.00		0	0.00		0	0.00		0	0.00		0	0.00		7,277	21.68
18-20		0	0.00		0	0.00		0	0.00		0	0.00		0	0.00		692	2.06
Gender																		
Male		1,777	81.96		1,224	81.65		2,101	81.94		900	81.60		1,824	81.90		27,479	81.87
Female		391	18.04		275	18.35		463	18.06		203	18.40		403	18.10		6,086	18.13
Race/Ethnicity																		
White		1,021	47.09		718	47.90		1,210	47.19		529	47.96		1,009	45.31		17,796	53.02
African-American/Black		39	1.80		27	1.80		50	1.95		16	1.45		44	1.98		691	2.06
Native Hawaiian or Pacific Islander		0	0.00		0	0.00		0	0.00		0	0.00		0	0.00		3	0.01
American Indian or Alaskan Native		1	0.05		2	0.13		2	0.08		1	0.09		1	0.04		57	0.17

	ASD Children with Identified Mother (N=2,168)		ASD Children with Identified Father (N=1,499)		ASD Children with Identified Mother or Father (N=2,564)		ASD Children with both Parents Identified (N=1,103)		Task A Likely ASD Enrolled at Birth (N=2,227)		Task A Likely ASD (N=33,565)	
Asian	44	2.03	24	1.60	48	1.87	20	1.81	42	1.89	466	1.39
Hispanic	116	5.35	79	5.27	142	5.54	53	4.81	108	4.85	1,366	4.07
Other	33	1.52	15	1.00	35	1.37	13	1.18	27	1.21	279	0.83
Unknown/missing	914	42.16	634	42.29	1,077	42.00	471	42.70	996	44.72	12,907	38.45
Household income												
Under \$50,000	177	8.16	116	7.74	212	8.27	81	7.34	166	7.45	3,090	9.21
\$50,000 - \$74,999	283	13.05	198	13.21	334	13.03	147	13.33	268	12.03	5,149	15.34
\$75,000 - \$99,999	333	15.36	227	15.14	393	15.33	167	15.14	337	15.13	4,838	14.41
\$100,000 - \$124,999	244	11.25	173	11.54	287	11.19	130	11.79	254	11.41	3,596	10.71
\$125,000+	182	8.39	133	8.87	215	8.39	100	9.07	167	7.50	2,915	8.68
Unknown/missing	949	43.77	652	43.50	1,123	43.80	478	43.34	1,035	46.48	13,977	41.64
Geographic Location												
Northeast	377	17.39	272	18.15	439	17.12	210	19.04	407	18.28	5,271	15.70
Midwest	633	29.20	426	28.42	754	29.41	305	27.65	641	28.78	11,561	34.44
South	809	37.32	587	39.16	981	38.26	415	37.62	857	38.48	12,090	36.02
West	349	16.10	214	14.28	390	15.21	173	15.68	322	14.46	4,643	13.83

	ASD Children with Identified Mother (N=2,168)			ASD Children with Identified Father (N=1,499)			ASD Children with Identified Mother or Father (N=2,564)			ASD Children with both Parents Identified (N=1,103)			Task A Likely ASD Enrolled at Birth (N=2,227)			Task A Likely ASD (N=33,565)		
	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI
Mean number of days of continuous enrollment	1,617.88	1,589.69	1,646.08	1,561.14	1,529.20	1,593.07	1,587.60	1,561.56	1,613.63	1,611.16	1,575.10	1,647.22	1,727.44	1,700.43	1,754.44	1,322.81	1,314.24	1,331.38
	n	%		n	%		n	%		n	%		n	%		n	%	
6 mons	27	1.25	1.73	26	1.73		40	1.56		13	1.18		7	0.31		1,928	5.74	
12 mons	155	7.15	7.27	109	7.27		199	7.76		65	5.89		97	4.36		6,563	19.55	
24 mons	327	15.08	15.54	233	15.54		401	15.64		159	14.42		303	13.61		6,426	19.14	
36 mons	416	19.19	21.08	316	21.08		503	19.62		229	20.76		435	19.53		5,533	16.48	
≥48 mons	1,243	57.33	54.37	815	54.37		1,421	55.42		637	57.75		1,385	62.19		13,115	39.07	
	lower 95% CI	upper 95% CI		lower 95% CI	upper 95% CI		lower 95% CI	upper 95% CI		lower 95% CI	upper 95% CI		lower 95% CI	upper 95% CI		lower 95% CI	upper 95% CI	
Number of office visits	97.75	92.47	103.03	95.80	89.62	101.97	95.31	90.55	100.08	100.76	93.30	108.21	100.12	95.06	105.19	56.90	55.97	57.84
Comorbidity score	0.89	0.84	0.93	0.87	0.82	0.93	0.88	0.84	0.92	0.89	0.83	0.96	0.93	0.89	0.98	0.66	0.65	0.67
	n	%		n	%		n	%		n	%		n	%		n	%	
Comorbidity score (categorical)																		
0	1,039	47.92	47.90	718	47.90	48.36	1,240	48.36	46.87	517	46.87	45.08	1,004	45.08	18,756	55.88		
1	641	29.57	30.35	455	30.35	29.41	754	29.41	31.01	342	31.01	31.12	693	31.12	9,685	28.85		
2	282	13.01	12.34	185	12.34	12.83	329	12.83	12.51	138	12.51	13.79	307	13.79	3,461	10.31		
≥ 3	206	9.50	9.41	141	9.41	9.40	241	9.40	9.61	106	9.61	10.01	223	10.01	1,663	4.95		

Children include any child with an identified mother via the stork method or an identified father via the +/- 7 days around conception. No CE requirements were imposed on the children

Table 7. Demographic Characteristics of Comparison Group with Linked Parents

	Comparison Children with Identified Mother (N=10,646)			Comparison Children with Identified Father (N=7,098)			Comparison Children with Identified Mother or Father (N=12,376)			Comparison Children with both Parents Identified (N=5,368)			Task A Comparison Enrolled at Birth (N=12,255)			Task A Comparison (N=138,876)		
	mean	lower 95% CI	upper 95% CI	Mean	lower 95% CI	upper 95% CI	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	
Age at index date	0.15	0.14	0.16	0.17	0.15	8.66	8.63	8.69	0.21	0.08	0.09	0.00	-	-	8.66	8.63	8.69	
	n	n	%	n	%	%	n	%	n	%	n	%	n	%	n	%	%	
Age at group at index year																		
<2		10,219	95.99		6,795	18.39	25,534	94.97		5,260	97.99		12,255	100.00		25,534	18.39	
2-8		427	4.01		303	40.54	56,305	5.03		108	2.01		0	0.00		56,305	40.54	
9-17		0	0.00		0	31.38	43,584	0.00		0	0.00		0	0.00		43,584	31.38	
18-20		0	0.00		0	9.69	13,453	0.00		0	0.00		0	0.00		13,453	9.69	
Gender																		
Male		5,374	50.48		3,593	50.64	70,321	50.61		2,704	50.37		6,192	50.53		70,321	50.64	
Female		5,272	49.52		3,505	49.36	68,555	49.39		2,664	49.63		6,063	49.47		68,555	49.36	
Race/Ethnicity																		
White		3,767	35.38		2,653	40.53	56,286	35.71		2,001	37.28		4,110	33.54		56,286	40.53	
African-American/Black		242	2.27		161	3.52	4,883	2.42		103	1.92		293	2.39		4,883	3.52	
Native Hawaiian or Pacific Islander		2	0.02		2	0.03	44	0.03		0	0.00		3	0.02		44	0.03	
American Indian or Alaskan Native		14	0.13		7	0.15	203	0.13		5	0.09		12	0.10		203	0.15	
Asian		157	1.47		120	1.37	1,899	1.44		99	1.84		179	1.46		1,899	1.37	
Hispanic		481	4.52		295	5.35	7,434	4.61		205	3.82		498	4.06		7,434	5.35	
Other		105	0.99		68	0.54	754	0.93		58	1.08		97	0.79		754	0.54	
Unknown/missing		5,878	55.21		3,792	48.51	67,373	54.73		2,897	53.97		7,063	57.63		67,373	48.51	

	Comparison Children with Identified Mother (N=10,646)		Comparison Children with Identified Father (N=7,098)		Comparison Children with Identified Mother or Father (N=12,376)				Comparison Children with both Parents Identified (N=5,368)		Task A Comparison Enrolled at Birth (N=12,255)		Task A Comparison (N=138,876)		
Household income															
Under \$50,000	837	7.86	606	15,193	10.94	8.26	421	7.84	951	7.76	15,193	10.94	15,193	10.94	
\$50,000 - \$74,999	1,289	12.11	885	18,226	13.12	12.21	663	12.35	1,385	11.30	18,226	13.12	18,226	13.12	
\$75,000 - \$99,999	1,123	10.55	843	13,789	9.93	10.57	658	12.26	1,209	9.87	13,789	9.93	13,789	9.93	
\$100,000 - \$124,999	732	6.88	531	9,030	6.50	6.84	416	7.75	792	6.46	9,030	6.50	9,030	6.50	
\$125,000+	436	4.10	333	6,854	4.94	4.10	261	4.86	468	3.82	6,854	4.94	6,854	4.94	
Unknown/missing	6,229	58.51	3,900	75,784	54.57	58.02	2,949	54.94	7,450	60.79	75,784	54.57	75,784	54.57	
Geographic Location															
Northeast	1,125	10.57	763	14,537	10.47	10.42	598	11.14	1,337	10.91	14,537	10.47	14,537	10.47	
Midwest	3,370	31.66	2,315	42,064	30.29	31.53	1,783	33.22	3,835	31.29	42,064	30.29	42,064	30.29	
South	4,428	41.59	2,926	61,497	44.28	42.21	2,130	39.68	5,167	42.16	61,497	44.28	61,497	44.28	
West	1,723	16.18	1,094	20,778	14.96	15.84	857	15.96	1,916	15.63	20,778	14.96	20,778	14.96	
	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI
Mean number of days of continuous enrollment	860.03	848.17	871.90	927.57	931.19	863.97	867.16	883.32	852.04	874.51	923.95	927.57	923.95	931.19	
	n	%	n	n	%	%	n	%	n	%	n	%	n	%	
6 months	2,484	23.33	1,562	23,672	17.05	23.21	1,173	21.85	2,930	23.91	23,672	17.05	23,672	17.05	
12 months	3,269	30.71	2,229	43,361	31.22	31.04	1,657	30.87	3,717	30.33	43,361	31.22	43,361	31.22	
24 months	1,941	18.23	1,338	26,808	19.30	18.46	995	18.54	2,256	18.41	26,808	19.30	26,808	19.30	
36 months	1,195	11.22	807	17,307	12.46	10.94	648	12.07	1,327	10.83	17,307	12.46	17,307	12.46	
≥48 months	1,757	16.50	1,162	27,728	19.97	16.35	895	16.67	2,025	16.52	27,728	19.97	27,728	19.97	

	Comparison Children with Identified Mother (N=10,646)			Comparison Children with Identified Father (N=7,098)			Comparison Children with Identified Mother or Father (N=12,376)			Comparison Children with both Parents Identified (N=5,368)			Task A Comparison Enrolled at Birth (N=12,255)			Task A Comparison (N=138,876)		
	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI	mean	lower 95% CI	upper 95% CI
Number of office visits	19.61	19.28	19.94	19.88	19.45	10.94	10.85	11.03	19.69	20.52	20.04	20.99	20.23	19.91	20.55	10.94	10.85	11.03
Comorbidity score	0.26	0.25	0.27	0.26	0.24	0.20	0.20	0.21	0.26	0.26	0.25	0.28	0.26	0.25	0.28	0.20	0.20	0.21
		n	%		n		n	%		%	n	%		n	%	n	n	%
Comorbidity score (categorical)																		
0		8,456	79.43		5,621		114,717	82.60	79.44		4,246	79.10		9,664	78.86		114,717	82.60
1		1,799	16.90		1,214		20,666	14.88	16.99		910	16.95		2,119	17.29		20,666	14.88
2		280	2.63		201		2,885	2.08	2.61		158	2.94		341	2.78		2,885	2.08
≥ 3		111	1.04		62		608	0.44	0.96		54	1.01		131	1.07		608	0.44

Children include any child with an identified mother via the stork method or an identified father via the +/- 7 days around conception. No CE requirements were imposed on the children

OBJECTIVE 2: Identify the size of available subsamples meeting varying criteria for continuous enrollment over etiologic windows of potential interest.

Table 8 presents the available samples for mothers and fathers of children with ASD or comparison children based on various continuous enrollment criteria that represent periods where potential risk factors for ASD would be examined. Of the 2,176 mothers of children with ASD identified via the STORK criteria in Task D, a little over three-fourths had continuous enrollment from birth minus 14 weeks to birth while just over 55% had continuous enrollment from conception (birth minus 40 weeks) to birth. Only 26% of mothers had continuous enrollment from birth minus 92 weeks (the period comprising the entire prenatal plus a one year preconception period). Approximately 48% of mothers linked to a child with continuous enrollment over the entire prenatal and one-year postnatal period (a group where maternal prenatal and early postnatal risk factors could be studied). Finally, about 70% of the mothers identified via the Stork methodology were linked to a child with ASD with continuous enrollment through one full postnatal year (where early postnatal child risk factors could be measured) and 64% of mothers were linked to a child with ASD with continuous enrollment through 24 months of age. For questions pertaining to risk factors in children only, without the need to measure maternal risk factors, there were an additional 419 and 376 children (data not shown) that could be added to these two respective groups, increasing the effective sample size available by 27% for both the one and two year postnatal periods. Quite similar, but slightly higher, proportions were seen for mothers of comparison children. For fathers of children with ASD identified within seven days of the date of conception, approximately half had continuous enrollment from birth minus 92 weeks to birth minus 40 weeks (1 year prior to conception to conception).

Compared to the impact of imposing criteria to find children with parents enrolled around the time of the subject's birth, the impact of criteria around enrollment corresponding to etiologic windows of interest had a relatively smaller impact on decreasing effective sample size.

Table 8. ASD and Comparison Group Mother and Father Samples with Continuous Enrollment Requirements

	ASD		Comparison	
	N	%	n	%
Mother identified via STORK methodology	2,176	100.00	10,703	100.00
Mothers with CE (birth - 14 weeks) to birth*	1,670	76.75	8,799	82.21
Mothers with CE (birth - 40 weeks) to birth*	1,206	55.42	6,329	59.13
Mothers with CE (birth - 92 weeks) to birth*	573	26.33	3,270	30.55
Mothers with CE (birth - 40 weeks) to (birth + 12 months)*	1,035	47.56	4,448	41.56
Linked to child with CE \geq 12 months	1,532	70.40	6,538	61.09
Linked to child with CE > 24 months	1,391	63.92	3,748	35.02
Fathers identified with CE +/- 7 days of conception	1,513	100.00	7,204	100.00
Fathers with CE (birth - 92 weeks) to (birth - 40)	752	49.70	3,867	53.68
Linked to child with CE > 12 months	1,049	69.33	4,369	60.65
Linked to child with CE > 24 months	962	63.58	2,520	34.98

*subsamples of mothers are not mutually exclusive; for instance, a mother who has CE from birth minus 40 weeks would also be included in the CE from birth minus 14 weeks.

OBJECTIVE 3: Explore the feasibility of using claims-based information on scheduled obstetric procedures to more accurately identify conception and trimester cut points

The aim of the following analysis was to explore the feasibility of using claims-based information on normal obstetric procedures that are often scheduled to occur at set times in pregnancy to more accurately identify conception and trimester cut-points. This analysis was based on the subsample of mothers of children with ASD and mothers of the comparison group with CE (birth minus 40 weeks) to birth. In Table 9, the frequency of various procedures (See Appendix C for CPT codes) are shown distributed across mothers' pregnancy as a whole as well as across the first, second and third trimesters using our standard approach of defining date of conception (Table 4) and then establishing trimester cutpoints based on fixed periods (see Section III.D). First, the prevalence of these procedures among women in the sample ranged from 8% to 72% of women. Overall, the timing of the procedures corresponded well with the trimester in which they are typically scheduled. For example, the inhibin screen and the alpha-fetoprotein test which only occur in the 2nd trimester, took place during the second trimester for over 98% and 99% of the women who had the procedure, respectively. Similarly, the oral glucose tolerance test (OGTT), usually performed routinely in the third trimester, appears in the third trimester about 89% of the time, although it is sometimes tested earlier in women with a history of diabetes or gestational diabetes. Other procedures are slightly less well-matched to particular trimesters, but are tests, like the OGTT that may be done for various indications or concerns. For example, a diagnostic ultrasound is routinely recommended during the second trimester, but may also be conducted if there are suspicions about an anatomic abnormality because of family history or other reasons. We did observe some "spillover" of a few procedures into adjacent trimesters, although we did not determine the length of the discrepancy between procedure date and trimester cutpoint (therefore even a subject with a discrepancy as small as one day is included here). These results suggest that the timing of procedures in claims data could potentially be used to validate or improve the precision of etiologic windows for risk factors although there are very few procedures, like the inhibin screen, that are reliably administered during narrow enough time frames to potentially improve the ability to determine exact gestational age for risk factor research.

Table 9. Proportion of Mothers with Trimester Procedures, by Trimester*

PROCEDURE	TRIMESTER							
	Any trimester (N=7,535)		1 st trimester (N=7,535)		2 nd trimester (N=7,533)**		3 rd trimester (N=7,533)**	
	n	%	n	% ^{***}	n	% ^{***}	n	% ^{***}
Procedures that can take place during any trimester								
HCG test	4,691	62.26	2,240	47.76	3,623	77.25	18	0.38
Obstetric panel/complete blood count	5,414	71.85	4,540	83.89	854	15.78	55	1.02
Complete blood count	4,478	59.43	1,159	25.89	1,114	24.88	3,433	76.68
Procedures that take place during first or early second trimester								
Ultrasound (including maternal-fetal evaluation, single gestation)	1,830	24.29	1,420	77.64	558	30.51	17	0.93
Ultrasound (including test of nuchal translucency or for multiple gestation)	577	7.66	339	58.85	256	44.44	5	0.87
Pregnancy-associated plasma protein-A (11-13.9 weeks)	701	9.30	314	44.86	387	55.29	0	0.00
Second trimester procedures								
Inhibin Screen	1,896	25.16	23	1.21	1,871	98.73	7	0.37
Ultrasound (Diagnostic; 18 - 20 weeks preferred)	2,648	35.14	19	0.72	2,455	92.75	391	14.77
Alpha-fetoprotein (AFP)	3,715	49.30	36	0.97	3,678	99.00	13	0.35
Second or third trimester procedures								
Ultrasound (≥14 weeks)	5,259	69.79	693	13.18	4,538	86.32	1,776	33.78
Third trimester procedures								
Oral glucose tolerance test (OGTT; 26-28 weeks)	1,234	16.38	25	2.03	151	12.24	1,101	89.22

*Based on subsample of mothers of children with ASD and comparison group with CE (birth minus 40 weeks) to birth.

*** For those babies that were extremely premature, we did not provide a trimester 2 cut point (as described in our methods). The research team decided these particular mothers may not have a complete second or third trimester, but do still have a full prenatal period (and first trimester) to evaluate. Therefore, there are a small number of mothers for which risk factors are only measured during the entire prenatal period or the first trimester only, resulting in the sample size being somewhat smaller for measurement during the second and third trimester.

***Represents the proportion of mothers who had the procedure during the trimester of interest. Mothers could have the procedure in multiple trimesters and therefore the proportions across a given row may be greater than 100%.

OBJECTIVE 4: Identification and Measurement of Selected Potential Risk Factors for ASD

The objective of this section was to determine the feasibility of identifying selected potential risk factors for ASD within claims data. This was done by identifying the proportion of subjects in the subsamples within the ASD and comparison groups and contrasting them with estimates from the literature. We first present a summary of the literature to which we compared our estimated prevalences of the risk factors. Then, we present the prevalence of the risk factors in our data compared to the estimates found in the literature.

1. Literature Search Results

Our literature search located 56 studies with published prevalence estimates of the potential risk factors we included in our study among children, mothers, and fathers. As stated above, this literature search was not meant to be exhaustive; rather, the objective of our search was to obtain a range of values for the specific risk factors that could serve as a comparison to our results. Table 10 summarizes the number of estimates found for each population by risk factor. The actual value of literature estimate(s) is presented in later tables (Table 11 and onward) to allow ease of comparison with our prevalence estimates. While some risk factors had a rich literature, others lacked risk factor estimates in both the peer-reviewed and gray literature, while for others we were only able to locate one estimate. In addition, several ASD risk factors had multiple international studies that were cited for estimates including: Pitocin, SSRI medication, prenatal infection, gestational diabetes, anesthesia, and maternal fertility treatments. No information was found for the ASD and non-ASD population regarding paternal fertility treatments. Limited data was available for the ASD and non-ASD population for the medication risk factors, with the exception of Pitocin where “induced labor” was used as a measure of Pitocin-induced labor. Limited data was also available for prenatal ultrasound, NICU admission, and depression. At least one general population estimate was found for the majority of risk factors. General population estimates were not found for maternal and paternal autoimmune conditions and maternal fetal hypoxia. While this literature search provides some prevalences to which we can compare our data, caution needs to be used, especially in the cases where only one to two estimates were located. The range of prevalence estimates from these studies is reported in the risk factor tables in the next section. A brief summary of each article and the corresponding citation is organized by risk factor and provided in Appendix D and the complete bibliography is located in Appendix E.

Table 10. Number of Risk Factor Estimates Located in Literature Review

POTENTIAL RISK FACTORS	ASD ESTIMATES	NON –ASD ESTIMATES	GENERAL POPULATION ESTIMATES
Children risk factors			
NICU Admission	1	1	2
Non-singleton Birth	5	4	2
Maternal Fetal Hypoxia	4	4	0
Prematurity/Pre-term Delivery	6	5	2
Postnatal MMR Vaccine	3	2	5
No Vaccinations	0	0	2
Mother risk factors			
Asthma	2	2	1
Depression	1	1	3
Autoimmune Condition	5	5	1
Infections	5	3	1
Obesity	3	2	1
Gestational Diabetes	7	5	3
Preeclampsia/Eclampsia	4	3	2
Fertility Treatment	2	2	1
Prenatal Anesthesia	3	3	1
Prenatal Ultrasound	1	1	1
SSRI Medications	1	1	1
B2Ar Medications	1	1	1
Contraceptives	1	0	0
Antibiotics	1	0	2
Anticonvulsants	1	1	3
Pitocin	4	4	1
Father risk factors			
Fertility Treatment	0	0	1
Autoimmune Condition	3	3	1

2. Potential Risk Factors' Prevalences in the Early postnatal and Postnatal Period (measured in children's' claims)

Tables 11 and 12 present the prevalence estimates of postnatal risk factors in the ASD and comparison groups for the subsamples with 12 month and 24 month continuous enrollment, respectively. As expected, prevalences do not vary between the 12 vs 24 month continuous enrollment groups for two reasons 1) these two samples are not mutually exclusive as children with 24 months of continuous enrollment are also included in the 12 month continuous enrollment subsample; and 2) for these risk factors, the most relevant claims data will come from very early in life. For the ASD group, NICU/ICU stay and non-singleton pregnancy estimates appeared lower

than the published literature — a finding seen also for the comparison group, suggesting some underreporting in the children’s’ claims which may not be differential with respect to ASD diagnosis. Hypoxia and preterm delivery prevalences were within the published range (though hypoxia was at the lower end of that range) for both the ASD and comparison groups.

Table 21. Children with ASD with Early postnatal Risk Factors and Estimates from the Literature

	CE birth to ≥ 12 months (N=1,836)		CE birth to ≥ 24 months (N=1,664)		Estimate from Literature	
	n	%	n	%	ASD	National
					% (range)	% (range)
NICU or ICU Admission	204	11.11	182	10.94		
NICU only	172	9.37	152	9.13	13.0%	6.0%,7.0%
ICU only	79	4.30	73	4.39	-	-
Non-singleton pregnancy	2	0.11	2	0.12	2.0-9.7%	3.1-3.4%
Hypoxia	51	2.78	47	2.82	2.9-17.6%	-
Prematurity/pre-term delivery	259	14.11	236	14.18	6.2-16.7%	11.6-12.2%

*This table is subset to children with enrollment during the early postnatal period, because 12 and 24 month enrollment allowed children to enroll during the first 32 days after birth, therefore some children have been dropped from assessment of postnatal risk factors.

Table 12. Comparison Children with Early postnatal Risk Factors and Estimates from the Literature

	CE birth to ≥ 12 months (N=7,964)		CE birth to ≥ 24 months (N=4,546)		Estimate from Literature	
	n	%	n	%	Non-ASD	National
					% (range)	% (range)
NICU or ICU Admission	527	6.62	263	5.79		
NICU only	402	5.05	210	4.62	7.0%	6.0%,7.0%
ICU only	299	3.75	138	3.04	-	-
Non-singleton pregnancy	4	0.05	3	0.07	2.2-2.9%	3.3-3.4%
Hypoxia	156	1.96	88	1.94	0.8-16.2%	-
Prematurity/pre-term delivery	696	8.74	375	8.25	2.0-9.3%	11.6-12.2%

*This table is subset to children with enrollment during the early postnatal period, because 12 and 24 month enrollment allowed children to enroll during the first 32 days after birth, therefore some children have been dropped from assessment of postnatal risk factors.

Tables 13 and 14 present prevalences of MMR vaccination in the ASD and comparison groups with continuous enrollment from birth to 24 months. The proportion of children with ASD with MMR from 0-24 months was within the range found in the literature (Table 13). The same was found for the comparison group (Table 14). It should be noted that there is one low estimate, 34%, reported in the literature for the national rate of MMR vaccination. This is the only estimate based on claims data that we found in the literature. Our estimates, in contrast, were within the range of the other non-claims-based estimates which strengthens the likely validity of our data for appropriately capturing MMR vaccinations in these populations.

Table 13. Children with ASD with Postnatal MMR Vaccination and Estimates from the Literature

	CE birth to ≥ 24 months (N=1,767)		Estimate from Literature	
	N	%	ASD	National
			% (range)	% (range)
MMR at 0-24 months	1,463	82.80	78.1-84.1%	34.3*-91.2%
90705 (measles), 90704 (mumps) and 90706 (rubella)	6	0.34		
90704 (mumps) and 90708 (measles and rubella)	0	0.00		
90705 (measles) and 90709 (Rubella and mumps)	0	0.00		
90707 (MMR)	1,330	75.27		
90710 (MMR and varicella)	127	7.19		

*34.3% is an estimate for one study based on claims data. Excluding that estimate, the range is 64%-91.2%

Table 14. Comparison Children with Postnatal MMR Vaccination and Estimates from the Literature

	CE birth to ≥ 24 months (N=4,876)		Estimate from Literature	
	N	%	Non-ASD	National
			% (range)	% (range)
MMR at 0-24 months	4,072	83.51	82.0%	34.3*-91.2%
90705 (measles), 90704 (mumps) and 90706 (rubella)	7	0.14		
90704 (mumps) and 90708 (measles and rubella)	0	0.00		
90705 (measles) and 90709 (Rubella and mumps)	0	0.00		
90707 (MMR)	3,651	74.88		
90710 (MMR and varicella)	421	8.63		

*34.3% is an estimate for one study based on claims data. Excluding that estimate, the range is 64%-91.2%

With respect to vaccination variables, we also sought to estimate the proportion of children in the ASD and comparison groups with no evidence of any vaccinations from claims. As mentioned above, for this particular variable we also considered a third subsample with continuous enrollment from birth up until age 5. The proportion of children in the ASD group with no evidence of vaccination was 6.8% for children with continuous enrollment 12 months or greater, 5.6% in children with continuous enrollment of 24 months or greater, and 4.1% in children with continuous enrollment for 60 months or greater (Table 15). These proportions are considerably higher than the estimates of 0.3% and 0.7% reported in the literature.¹⁶ This likely indicates that either claims for vaccinations are not reported consistently or that vaccinations may be included in more general claims (such as preventive care visit) and not specifically distinct (i.e., rolled up). This is further evidence that vaccination rates are likely considerably underestimated in claims

¹⁶ The two studies mentioned include the CDC National and State Vaccination Coverage among Children Aged 19-35 Months-United States, 2010 and study by Smith et al published in Pediatrics in 2004. Full citations are available in Appendix E.

data. In the comparison group, the proportions were of roughly similar magnitude as those for the children with ASD for all the continuous enrollment periods.

Table 15. Children with ASD with No Vaccinations and Estimates from the Literature

No Vaccinations	CE birth to ≥ 12 months (N=1,951)		CE birth to ≥ 24 months (N=1,767)		CE birth to ≥ 60 months (N=784)		Estimate from Literature	
	n	%	n	%	n	%	ASD	National
							% (range)	% (range)
Birth to 12 months	132	6.77	121	6.85	73	9.31	–	–
Birth to 24 months	–	–	99	5.60	62	7.91	–	0.3,0.7%
Birth to 60 months	–	–	–	–	32	4.08	–	–

Table 16. Comparison Children with No Vaccinations and Estimates from the Literature

No Vaccinations	CE birth to ≥ 12 months (N=8,512)		CE birth to ≥ 24 months (N=4,876)		CE birth to ≥ 60 months (N=987)		Estimate from Literature	
	n	%	n	%	n	%	Non-ASD	National
							% (range)	% (range)
Birth to 12 months	601	7.06	356	7.30	96	9.73	–	–
Birth to 24 months	–	–	281	5.76	73	7.40	–	0.3,0.7%
Birth to 60 months	–	–	–	–	44	4.46	–	–

3. Potential Risk Factors Prevalences in the Prenatal Period (measured in maternal claims)

As described above, we identified four subsamples of mothers of children in the ASD and comparison group based on various continuous enrollment criteria that represent periods of interest when potential risk factors for ASD would be examined: third trimester ((birth minus 14 weeks) to birth); prenatal period ((birth minus 40 weeks) to birth); preconception and prenatal period ((birth minus 92 weeks) to birth); and prenatal and one year following birth ((birth minus 40 weeks) to (birth plus 12 months)). Our initial analysis measured the potential ASD risk factor prevalences for each of the four subsamples. We found that the prevalences were fairly constant across all four continuous enrollment subsamples. This was as expected for two reasons. First, these subsamples are not mutually exclusive – mothers enrolled (birth minus 40 weeks) to birth would also be enrolled (birth minus 14 weeks) to birth and therefore represented in both subsamples. Second, most of the risk factors were measured during the same time period (e.g., first trimester) regardless of the mother subsample and therefore the period of observation was the same across subsamples. Therefore, with the exception of the first table which presents chronic conditions during the subject’s entire enrollment period, we present the prevalences of potential risk factors for only one of these subsamples – mothers enrolled the entire prenatal time period ((birth minus 40 weeks) to birth).

a. Chronic health conditions

Tables 17 and 18 present the proportions of mothers of children in the ASD and comparison group with chronic conditions considered as potential ASD risk factors. We consider chronic conditions as those potentially having etiologic influence in the prenatal period but potentially detectable by querying claims over longer periods of time because they are of long duration (in other words, a claim indicating diagnosis outside the etiologic window of interest potentially suggests that the condition was also present in the prenatal period). Prevalences are shown using different criteria for numbers of claims (e.g., one or two claims), different intervals for identifying these claims (e.g., by trimester, prenatal, preconception), and in the subsamples meeting different continuous enrollment criteria (e.g., mothers enrolled 14 weeks prior to birth, mothers enrolled during entire prenatal and preconception periods).

Regardless of which interval was used to query maternal asthma claims, the proportions of mothers in the ASD group (Table 17) with evidence of maternal asthma were similar and fell within the range seen in the literature. Requiring 2+ ICD-9 codes did, as expected, reduce prevalence estimates somewhat, but not dramatically. The proportion of mothers of children in the ASD group with evidence of depression was more dependent on the length of the period over which claims were queried than was true for asthma, with longer periods leading to higher prevalence estimates. In particular, inclusion of the postnatal observation time led to much higher prevalences. This suggests that post-partum depression episodes (which may be of less etiologic significance) are influencing estimates substantially and, consequently, intervals including postnatal experience should probably not be incorporated into a claims-based definition of depression. Estimates that included only prenatal periods for querying codes are not wildly out of range with those found in the literature, although these published estimates are based on heterogeneous observation periods. As with depression, autoimmune condition prevalences were also influenced by observation period. Pregnancy can be a trigger for certain autoimmune conditions but the association between pregnancy and autoimmune diagnoses is not as strong as for depression. The estimates of autoimmune prevalence are markedly higher than published estimates, reaching prevalences as high as 70%, which may reflect a generous list of included ICD-9-CM codes (included in Appendix B) we called “autoimmune” compared to definitions in the literature. In addition, some of the articles cited were strictly based on hospital admissions with an autoimmune diagnosis. The estimates for the comparison group of mothers followed generally the same pattern as of mothers of children with ASD.

When querying codes during the entire enrollment period, mother’s index date to birth, and birth to mother’s end of enrollment, the prevalence estimates were higher for the subsamples of mothers with longer continuous enrollment requirements. For example, when measuring the prevalence of asthma from index date to birth, the proportion of mothers with asthma increased from 6.7% for mothers continuously enrolled (birth minus 14 weeks) to birth, to 8% for mothers continuously enrolled (birth minus 40 weeks) to birth, to 9.4% for mothers continuously enrolled (birth minus 92 weeks) to birth. This is as expected given the subsamples with longer continuous enrollment requirements have more time in which to incur a claim for asthma or another chronic condition. On the other hand, when querying codes during the prenatal period only, the prevalence estimates are quite similar across subsamples regardless of continuous enrollment requirements, ranging from 4.7% to 5.1%. As mentioned earlier in this section, this is as expected given the mother subsamples are not mutually exclusive. Similar results were found across all potential risk factors measurable in mothers’ claims. Therefore, the remaining tables on acute

health conditions, maternal medical procedures, and medication use only present the prevalence estimates for mothers enrolled the entire prenatal time period (birth minus 40 weeks) to birth.

Table 17. Mothers of Children with ASD with Chronic Condition Risk Factors and Estimates from the Literature

	Mothers CE (birth-14 weeks) to birth (N=1,670)			Mothers CE (birth-40 weeks) to birth (N=1,206)		Mothers CE (birth-92 weeks) to birth (N=573)		Mothers CE (birth-40 weeks) to (birth plus 12 months) (N=1,035)		Estimate in Literature	
										ASD	National
	valid N	n	%	n	%	n	%	n	%	% (range)	% (range)
Maternal asthma											
1+ diagnoses										4.2-15.5%	3.7-8.4%
Index date to birth	1,670	112	6.71	97	8.04	54	9.42	83	8.02		
During pregnancy	1,214*	57	4.70	56	4.64	29	5.06	47	4.54		
Birth to end of CE	1,670	226	13.53	165	13.68	73	12.74	151	14.59		
Entire CE (index date to end of CE)	1,670	255	15.27	192	15.92	92	16.06	172	16.62		
1 dx & prescription OR 2+ diagnoses											
Index date to birth	1,670	92	5.51	83	6.88	48	8.38	71	6.86		
During pregnancy	1,214	37	3.05	36	2.99	18	3.14	30	2.90		
Birth to end of CE	1,670	204	12.22	148	12.27	62	10.82	134	12.95		
Entire CE (index date to end of CE)	1,670	236	14.13	177	14.68	84	14.66	159	15.36		
2+ diagnoses											
Index date to birth	1,670	60	3.59	55	4.56	30	5.24	47	4.54		
During pregnancy	1,214	23	1.89	22	1.82	9	1.57	18	1.74		
Birth to end of CE	1,670	133	7.96	96	7.96	42	7.33	90	8.70		
Entire CE (index date to end of CE)	1,670	158	9.46	120	9.95	57	9.95	109	10.53		
Maternal depression											
1+ diagnoses										1.8%	8.5-15.4%
Index date to birth	1,670	171	10.24	146	12.11	74	12.91	120	11.59		
During pregnancy	1,214	86	7.08	85	7.05	31	5.41	66	6.38		
Birth to end of CE	1,670	561	33.59	412	34.16	173	30.19	364	35.17		

	Mothers CE (birth-14 weeks) to birth (N=1,670)			Mothers CE (birth-40 weeks) to birth (N=1,206)		Mothers CE (birth-92 weeks) to birth (N=573)		Mothers CE (birth-40 weeks) to (birth plus 12 months) (N=1,035)		Estimate in Literature	
										ASD	National
	valid N	n	%	n	%	n	%	n	%	% (range)	% (range)
Entire CE (index date to end of CE)	1,670	610	36.53	451	37.40	195	34.03	393	37.97		
2+ diagnoses											
Index date to birth	1,670	109	6.53	98	8.13	53	9.25	81	7.83		
During pregnancy	1,214	48	3.95	47	3.90	20	3.49	36	3.48		
Birth to end of CE	1,670	402	24.07	292	24.21	114	19.90	261	25.22		
Entire CE (index date to end of CE)	1,670	446	26.71	328	27.20	136	23.73	286	27.63		
Maternal autoimmune conditions											
1+ diagnoses										6.2-25.7%	5.0-8.0%
Index date to birth	1,670	530	31.74	464	38.47	264	46.07	410	39.61		
During pregnancy	1,214	270	22.24	270	22.39	130	22.69	236	22.80		
Birth to end of CE	1,670	962	57.60	701	58.13	337	58.81	621	60.00		
Entire CE (index date to end of CE)	1,670	1,091	65.33	816	67.66	412	71.90	720	69.57		
2+ diagnoses											
Index date to birth	1,670	252	15.09	227	18.82	145	25.31	206	19.90		
During pregnancy	1,214	97	7.99	97	8.04	50	8.73	87	8.41		
Birth to end of CE	1,670	629	37.66	456	37.81	219	38.22	420	40.58		
Entire CE (index date to end of CE)	1,670	764	45.75	581	48.18	296	51.66	523	50.53		

*The Valid N is smaller than the subsample of 1,670 because in order to measure the prevalence of the potential risk factor during conception to birth, the mother had to have continuous enrollment from conception to birth. Only 1,214 mothers of the 1,670 enrollment during their last trimester were enrolled during the entire prenatal period based on our algorithm to define conception date.

Table 18. Mothers of Comparison Children with Chronic Condition Risk Factors and Estimates from the Literature

	Mothers CE (birth-14 weeks) to birth (N=8,799)			Mothers CE (birth-40 weeks) to birth (N=6,329)		Mothers CE (birth-92 weeks) to birth (N=3,270)		Mothers CE (birth-40 weeks) to (birth plus 12 months) (N=4,448)		Estimate in Literature	
				n	%	n	%	n	%	n	%
	valid N	n	%	n	%	n	%	n	%	% (range)	% (range)
Maternal asthma											
1+ diagnoses										4.8-10.5%	3.7-8.4%
Index date to birth	8,799	478	5.43	394	6.23	270	8.26	270	6.07		
During pregnancy	6,374	217	3.40	216	3.41	132	4.04	144	3.24		
Birth to end of CE	8,799	617	7.01	443	7.00	228	6.97	367	8.25		
Entire CE (index date to end of CE)	8,799	846	9.61	639	10.10	370	11.31	487	10.95		
1 dx & prescription OR 2+ diagnoses											
Index date to birth	8,799	373	4.24	319	5.04	215	6.57	212	4.77		
During pregnancy	6,374	163	2.56	162	2.56	92	2.81	108	2.43		
Birth to end of CE	8,799	473	5.38	345	5.45	174	5.32	293	6.59		
Entire CE (index date to end of CE)	8,799	704	8.00	544	8.60	307	9.39	412	9.26		
2+ diagnoses											
Index date to birth	8,799	212	2.41	184	2.91	137	4.19	129	2.90		
During pregnancy	6,374	63	0.99	63	1.00	40	1.22	47	1.06		
Birth to end of CE	8,799	281	3.19	201	3.18	104	3.18	171	3.84		
Entire CE (index date to end of CE)	8,799	436	4.96	334	5.28	204	6.24	256	5.76		
Maternal depression											
1+ diagnoses										0.01%	8.4-15.4%
Index date to birth	8,799	719	8.17	637	10.06	429	13.12	443	9.96		
During pregnancy	6,374	264	4.14	263	4.16	137	4.19	181	4.07		
Birth to end of CE	8,799	1,354	15.39	972	15.36	510	15.60	796	17.90		
Entire CE (index date to end of CE)	8,799	1,728	19.64	1,312	20.73	746	22.81	1,007	22.64		

	Mothers CE (birth-14 weeks) to birth (N=8,799)			Mothers CE (birth-40 weeks) to birth (N=6,329)		Mothers CE (birth-92 weeks) to birth (N=3,270)		Mothers CE (birth-40 weeks) to (birth plus 12 months) (N=4,448)		Estimate in Literature	
				n	%	n	%	n	%	n	%
	valid N	n	%	n	%	n	%	n	%	% (range)	% (range)
2+ diagnoses											
Index date to birth	8,799	430	4.89	382	6.04	271	8.29	263	5.91		
During pregnancy	6,374	132	2.07	131	2.07	66	2.02	89	2.00		
Birth to end of CE	8,799	821	9.33	592	9.35	315	9.63	497	11.17		
Entire CE (index date to end of CE)	8,799	1,116	12.68	862	13.62	509	15.57	668	15.02		
Maternal autoimmune conditions											
1+ diagnoses										2.0-9.9%	5-8%
Index date to birth	8,799	2,445	27.79	2,125	33.58	1,373	41.99	1,487	33.43		
During pregnancy	6,374	1,073	16.83	1,068	16.87	578	17.68	737	16.57		
Birth to end of CE	8,799	3,036	34.50	2,214	34.98	1,169	35.75	1,810	40.69		
Entire CE (index date to end of CE)	8,799	4,184	47.55	3,239	51.18	1,850	56.57	2,432	54.68		
2+ diagnoses											
Index date to birth	8,799	1,136	12.91	1,031	16.29	752	23.00	712	16.01		
During pregnancy	6,374	387	6.07	386	6.10	217	6.64	253	5.69		
Birth to end of CE	8,799	1,609	18.29	1,171	18.50	642	19.63	1,006	22.62		
Entire CE (index date to end of CE)	8,799	2,446	27.80	1,930	30.49	1,193	36.48	1,501	33.75		

b. Acute health conditions

Tables 19 and 20 present the estimated prevalences of acute or episodic conditions likely to have etiologic influence during the prenatal period in the ASD and comparison groups. Prevalence estimates are shown for the entire prenatal period and, when relevant, for each trimester in each of the four subsamples based on continuous enrollment criteria. In the ASD group (Table 19) the prevalence of maternal prenatal infection indicators was substantially higher than published estimates (which tend to be studies with varying outcome definitions based on interviews and registries, and were mostly non-US based). Our working definition of prenatal infection may have been too broad; however, it is clear that infection-related codes are recorded frequently on claims during pregnancy and further exploration different coding algorithms for this variable in claims may be necessary.

As expected, the maternal code for obesity was seen at less than expected levels in pregnancy in both the ASD and comparison group mothers. Claims data are likely not well-suited for investigation of obesity in general and/or excessive weight gain during pregnancy.

The prevalence of gestational diabetes in the ASD group mothers was substantially higher than published ranges from the literature based on mothers of children with ASD and the general population suggesting that our claims-based algorithm requiring one ICD-9-CM code for gestational diabetes could have been capturing some rule-out testing or screening for gestational diabetes. We then expanded our algorithm to include two definitions: 1) any subject that had a regular diabetes code during pregnancy as long as the subject did not have a diagnosis of diabetes or a prescription for a diabetic medication during the preconception period or 2) a gestational diabetes code during the prenatal period. While the sample size for this revised algorithm was smaller (n=576) due to the requirement of having continuous enrollment during the entire preconception period, the result remained much higher than the literature. Further exploration of claims-based coding for this risk factor would be indicated for future studies of this risk factor using claims data. Similarly, the prevalence of preeclampsia in ASD group mothers exceeded that reported for other ASD groups in the literature suggesting that, again, the claims-based algorithm applied here may be too generous in identifying true preeclampsia.

Findings for the comparison group mothers were very similar. For gestational diabetes and preeclampsia, the upper ends of the range of prevalences from the published literature were lower than that seen for the comparison group mothers but the prevalence estimates were similar, underscoring the likely over-capture in claims of rule-out or screening applications around these particular diagnoses.

Table 19. Mothers of Children with ASD with Acute Health Condition Risk Factors and Estimates from the Literature

	Mothers CE (birth-40 weeks) to birth (N=1,206)			Estimate from Literature	
				ASD	National
	valid N	N	%	% (range)	% (range)
Maternal prenatal infection	1,206	618	51.24	2.3-14.7%	22.0%
First trimester	1,206	304	25.21		
Second trimester	1,204*	245	20.35		
Third trimester	1,204	306	25.42		
Maternal obesity (prenatal)	1,206	34	2.82	21.5%	
Maternal gestational diabetes (prenatal)				1.0-10.5%	3.0-5.2%
Gestational ICD9 only	1,206	174	14.43		
Gestational and Diabetes ICD9**	576	79	13.72		
Maternal preeclampsia/eclampsia (prenatal)	1,206	129	10.70	2.4-11.0%	3.7%

* For those babies that were extremely premature, we did not provide a trimester 2 cut point (as described in our methods). The research team decided these particular mothers may not have a complete second or third trimester, but do still have a full prenatal period (and first trimester) to evaluate. Therefore, there are a small number of mothers for which risk factors are only measured during the entire prenatal period or the first trimester only, resulting in the sample size being somewhat smaller for measurement during the second and third trimester.

Table 20. Mothers of Comparison Children with Acute Health Condition Risk Factors and Estimates from the Literature

	Mothers CE (birth-40 weeks) to birth (N=6,329)			Estimate from Literature	
				Non-ASD	National
	valid N	n	%	% (range)	% (range)
Maternal prenatal infection	6,329	3,011	47.57	2.0-14.7%	22.0%
First trimester	6,329	1,488	23.51		
Second trimester	6,329	1,228	19.40		
Third trimester	6,329	1,435	22.67		
Maternal obesity (prenatal)	6,329	114	1.80		
Maternal gestational diabetes (prenatal)				0.5-5.5%	3.0-5.2%
Gestational ICD9 only	6,329	887	14.01		
Gestational and Diabetes ICD9	3,285	445	13.55		
Maternal preeclampsia/eclampsia (prenatal)	6,329	539	8.52	1.3-6.3%	0.2-3.7%

c. Maternal medical procedures

Tables 21 and 22 present potential prenatal maternal medical procedure risk factors in the ASD and comparison groups. Two potential risk factors are considered – anesthesia administration (data presented as prevalence of any occurrence) and ultrasound administration (data presented as both prevalence of occurrence and average procedure counts). Estimates are shown for the entire prenatal period and, when relevant, for each trimester. Anesthesia administration is also measured during the early postnatal period (birth to birth plus one week) to account for anesthesia received during delivery. For all potential risk factors in both of these tables, similar to prevalences presented above, estimates did not vary across the continuous enrollment criteria subsamples (as stated previously, data not shown for all subsamples), suggesting that enrollment criteria would not introduce selection bias with respect to these potential risk factors.

In the ASD group (Table 21), anesthesia exposure during the early postnatal period (birth to birth plus one week) was approximately 74% which is within the range of estimates published in the literature (57-89%). The comparison mothers had results that were similarly comparable to the literature (Table 22). Proportions of mothers of children with ASD with any ultrasound exposure were higher than published estimates, perhaps reflecting the fact that insured populations are more likely to receive this procedure than other samples in the literature that included both insured and uninsured mothers. Ultrasound frequency was also much higher in our claims data than that reported in the literature; over two times higher for mothers of children with ASD (4.6) when compared to the highest national rate (2.7) and a similarly higher rate for the comparison mothers (3.9).

Table 21. Mothers of Children with ASD with Medical Utilization Risk Factors and Estimates from the Literature

	Mothers CE (birth-40 weeks) to birth (N=1,206)			Estimate from Literature	
	valid N	n	%	ASD % (range)	National % (range)
Maternal evidence of anesthesia prenatal	1,206	131	10.86		
Maternal evidence of anesthesia (early postnatal)	1,204	893	74.17	56.6-88.6%	61.0%
Prenatal ultrasound (Y/N)	1,206	1,123	93.12	86.7%	N/A
First trimester	1,206	744	61.69		
Second trimester	1,204	1,070	88.87		
Third trimester	1,204	712	59.14		
	mean	lower 95% CI	upper 95% CI		
Prenatal ultrasound (counts)	4.64	4.42	4.86	1.8	1.5,2.7
First trimester	2.11	1.98	2.25		
Second trimester	1.84	1.76	1.92		
Third trimester	2.33	2.19	2.47		

Table 22. Mothers of Comparison Children with Medical Utilization Risk Factors and Estimates from the Literature

	Mothers CE (birth-40 weeks) to birth (N=6,329)			Estimate from Literature	
	valid N	n	%	Non-ASD % (range)	National % (range)
Maternal evidence of anesthesia prenatal	6,329	644	10.18		
Maternal evidence of anesthesia (early postnatal)	6,319	4,562	72.19	40.0-93.5%	61.0%
Prenatal ultrasound (Y/N)	6,329	6,014	95.02	87.5%	N/A
First trimester	6,329	3,891	61.48		
Second trimester	6,329	5,738	90.66		
Third trimester	6,329	3,475	54.91		
	mean	lower 95% CI	upper 95% CI		
Prenatal ultrasound (counts)*	3.93	3.85	4.01	1.7	1.5-2.7
First trimester	1.79	1.74	1.84		
Second trimester	1.64	1.61	1.67		
Third trimester	2.09	2.03	2.15		

Tables 23 and 24 present the estimated prevalences of maternal prenatal medication exposure considered as potential ASD risk factors in the ASD and comparison groups. Prevalence estimates are shown for the entire prenatal period and, when relevant, for each trimester. Prevalence estimates are also shown for the early postnatal period (birth to birth plus one week) for Pitocin to capture use at delivery. For all potential risk factors in both of these tables, prevalences did not vary meaningfully across the continuous enrollment criteria subsamples (as stated previously, data not shown for all subsamples), suggesting that enrollment criteria would not introduce selection bias with respect to these potential risk factors.

Oral contraceptive exposure prevalence was low, as expected among a group of pregnant women. Of particular interest is oral contraceptive exposure early in the first trimester when mothers continue with birth control because they are unaware that they have conceived. We were able to locate only one estimate from the literature that was based on only 51 mothers which reported that 12% of mothers of children with ASD were using *contraception* (oral and other forms) at time of conception. However, given the infrequency of this exposure and the uncertainty around date of conception (given that this can be crudely estimated in this data source), it appears unlikely that claims-based investigation of this exposure would be fruitful.

Similarly, women who regularly receive antiepileptic medications may cease taking these drugs during pregnancy because of the risk of adverse effects. However, for a small percentage that have severe or intractable epilepsy, for whom the risks of stopping medications outweigh the risk of adverse effects on the fetus, or for women who continue medications because they are not yet aware they are pregnant, antiepileptic medication may still be an important risk factor and would most likely appear in claims. Thus our estimate (1.16%) is aligned with those in the literature for women with children with ASD (1.1%) and somewhat higher than those reported among a more general population of women (0.1-0.4%). The comparison group also compared similarly to the literature.

Our data for antibiotic use was consistent with the national estimates in the literature. Particularly, our estimate compares favorably to the 40.8% estimate which represents the proportion of females who delivered an infant in 1996-2000 with anti-infective oral or injectable use during pregnancy. The only estimate for antibiotic exposure in mothers of children with ASD was remarkably lower, but was based only 144 mothers who had *only* antibiotic use and therefore is not a comparable estimate.

Pitocin use during the early postnatal period (birth to birth plus one week) was remarkably low with only one of our 1,204 mother subsample being flagged as having received Pitocin. In contrast, the literature cites anywhere from 17% to 37% for induced labor or Pitocin use. As stated above, we are unable to capture inpatient drug administration in claims data. This suggests claims data is not a reliable source for capturing Pitocin use among mothers.

Table 23. Mothers of Children with ASD with Medication Risk Factors and Estimates from the Literature

Class of Medication	Mothers CE (birth-40 weeks) to birth (N=1,206)			Estimate in Literature	
	valid N	n	%	ASD	National
				% (range)	% (range)
SSRI	1,206	93	7.71	4.5%	1.4-10.1%
First trimester	1,206	78	6.47		
Second trimester	1,204	55	4.57		
Third trimester	1,204	59	4.90		
Beta2 Agonists	1,206	125	10.36	19.0%	2.8%
First trimester	1,206	26	2.16		
Second trimester	1,204	44	3.65		
Third trimester	1,204	91	7.56		
Contraceptives				12.0%*	
Oral (first trimester)	1,206	27	2.24		
Other (first trimester)	1,206	3	0.25		
Anticonvulsants	1,206	14	1.16	1.0%	0.1-0.6%
First trimester	1,206	14	1.16		
Second trimester	1,204	7	0.58		
Third trimester	1,204	7	0.58		
Antibiotics	1,206	600	49.75	3.5%	40.8,62%
First trimester	1,206	283	23.47		
Second trimester	1,204	259	21.51		
Third trimester	1,204	308	25.58		
Pitocin (early postnatal period)	1,204	1	0.08	17.0-37.3%	23.2%

*12.0% is based on any use of contraception at or before conception, not just oral contraception.

Table 24. Mothers of Comparison Children with Medication Risk Factors and Estimates from the Literature

Class of Medication	Mothers CE (birth-40 weeks)to birth (N=6,329)			Estimate in Literature	
	valid N	n	%	Non-ASD	National
				% (range)	% (range)
SSRI	6,329	349	5.51	1.7%	1.4%-10.1%
First trimester	6,329	267	4.22		
Second trimester	6,329	202	3.19		
Third trimester	6,329	214	3.38		
Beta2 Agonists	6,329	519	8.20	15.0%	2.8%-7.5%
First trimester	6,329	148	2.34		
Second trimester	6,329	192	3.03		
Third trimester	6,329	354	5.59		
Contraceptives					
Oral (first trimester)	6,329	144	2.28		
Other (first trimester)	6,329	20	0.32		
Anticonvulsants	6,329	46	0.73	1.0%	0.1-0.6%
First trimester	6,329	38	0.60		
Second trimester	6,329	19	0.30		
Third trimester	6,329	20	0.32		
Antibiotics	6,329	2,998	47.37		40.8%,62%
First trimester	6,329	1,380	21.80		
Second trimester	6,329	1,328	20.98		
Third trimester	6,329	1,491	23.56		
Pitocin (early postnatal period)	6,319	8	0.13	17.0-44.7%	23.2%

In the ASD group (Table 23 above), SSRI exposure is in the range of published estimates and, of note, was very similar to a recent estimate published based on an insured Kaiser HMO population (See Appendix D and E for more information on the Kaiser study). B2Ar exposure prevalence is below that reported in the literature and was also lower than a recent estimate from the Kaiser system. The discrepancy with Kaiser data was greatest in the third trimester. This may be due to our inability to capture the inpatient administration of drugs (e.g., antitocolytics) that the mothers receive when hospitalized for delivery or for acute conditions during pregnancy (such as asthma).

Tables 25 through 28 present cross-tabulations of data on medication exposure prevalence and the prevalence of corresponding indicating conditions. In investigation of pharmacologic risk factors the problem of confounding by indication, where the risk is conferred by the indicating condition (or related factors) as opposed to the drug exposure itself, is a major methodological concern. Large claims databases have promise to allow empirical exploration of confounding by indication because sufficiently sized subsamples of individuals who are exposed only to either the medication or the indication can be identified. Explorations performed in our study support this notion. For example, in the ASD group, there were 53 mothers exposed prenatally to SSRIs who did not have a depression diagnosis – a number roughly equal to those with exposure that carried a depression diagnosis. Data in the comparison groups, and for B2ARs, also suggest that identification of subsamples that provide information on the independent effects of exposure and indication is feasible in claims-based investigations. However, caution needs to be maintained when considering the meaning of the appearance of a diagnostic code. For mental health conditions in particular, a diagnosis may not appear because there is concern about labeling and subsequent stigma, not because the condition does not exist. Thus the absence of an asthma diagnosis is probably more reliable than the absence of a diagnosis for depression or anxiety.

Table 25. Mothers of Children with ASD with depression medication Risk Factor

	Depression Diagnosis (N=85; 7%)		No Depression Diagnosis (N=1,121; 93%)	
	n	%*	n	%*
Prenatal period				
SSRI	40	3.32	53	4.39
Non-SSRI medication	16	1.33	20	1.66
Both SSRI and non-SSRI medication	10	0.83	3	0.25
No depression medication	39	3.23	1,051	87.15
First trimester				
SSRI	35	2.90	43	3.57
Non-SSRI medication	14	1.16	17	1.41
Both SSRI and non-SSRI medication	8	0.66	3	0.25
No depression medication	44	3.65	1,064	88.23
Second trimester				
SSRI	26	2.16	29	2.40
Non-SSRI medication	9	0.75	8	0.66
Both SSRI and non-SSRI medication	5	0.41	0	0
No depression medication	55	4.56	1,082	89.72
Third trimester				
SSRI	28	2.32	31	2.57
Non-SSRI medication	11	0.91	10	0.83
Both SSRI and non-SSRI medication	5	0.41	0	0
No depression medication	51	4.23	1,078	89.39

*percentages are based on total table N=1,206.

Table 26. Mothers of Comparison Children with depression medication Risk Factor*

	Depression Diagnosis (N=263; 4%)		No Depression Diagnosis (N=6,066; 96%)	
	n	%*	n	%*
Prenatal period				
SSRI	93	1.47	256	4.04
Non-SSRI medication	42	0.66	79	1.25
Both SSRI and non-SSRI medication	19	0.30	19	0.30
No depression medication	147	2.32	5,750	90.85
First trimester				
SSRI	73	1.15	194	3.07
Non-SSRI medication	38	0.60	69	1.09
Both SSRI and non-SSRI medication	14	0.22	14	0.22
No depression medication	166	2.62	5,817	91.91
Second trimester				
SSRI	61	0.96	141	2.23
Non-SSRI medication	17	0.27	32	0.51
Both SSRI and non-SSRI medication	5	0.08	7	0.11
No depression medication	190	3.00	5,900	93.22
Third trimester				
SSRI	62	0.98	152	2.40
Non-SSRI medication	19	0.30	35	0.55
Both SSRI and non-SSRI medication	5	0.08	7	0.11
No depression medication	187	2.95	5,886	93.00

*percentages are based on total table N=6,329.

Table 27. Mothers of Children with ASD with asthma medication Risk Factor

	Asthma Diagnosis (N=56; 5%)		No Asthma Diagnosis (N=1,150; 95%)	
	n	%*	n	%*
Prenatal period				
B2AR	32	2.65	93	7.71
Non-B2AR medication	16	1.33	50	4.15
Both B2AR and non-B2AR medications	12	1.00	12	1.00
No asthma medication	20	1.66	1,019	84.49
First trimester				
B2AR	11	0.91	15	1.24
Non-B2AR medication	9	0.75	28	2.32
Both B2AR and non-B2AR medications	3	0.25	2	0.17
No asthma medication	39	3.23	1,109	91.96
Second trimester				
B2AR	19	1.58	25	2.07
Non-B2AR medication	7	0.58	7	0.58
Both B2AR and non-B2AR medications	3	0.25	1	0.08
No asthma medication	33	2.74	1,117	92.62
Third trimester				
B2AR	20	1.66	71	5.89
Non-B2AR medication	8	0.66	19	1.58
Both B2AR and non-B2AR medications	5	0.41	5	0.41
No asthma medication	33	2.74	1,063	88.14

*percentages are based on total table N=1,206.

Table 28. Mothers of Comparison Children with asthma medication Risk Factor

	Asthma Diagnosis (N=216; 3%)		No Asthma Diagnosis (N=6,113; 97%)	
	n	%*	n	%*
Prenatal period				
B2AR	145	2.29	374	5.91
Non-B2AR medication	94	1.49	216	3.41
Both B2AR and non-B2AR medications	77	1.22	41	0.65
No asthma medication	54	0.85	5,564	87.91
First trimester				
B2AR	72	1.14	76	1.20
Non-B2AR medication	48	0.76	138	2.18
Both B2AR and non-B2AR medications	32	0.51	14	0.22
No asthma medication	128	2.02	5,913	93.43
Second trimester				
B2AR	90	1.42	102	1.61
Non-B2AR medication	58	0.92	67	1.06
Both B2AR and non-B2AR medications	39	0.62	9	0.14
No asthma medication	107	1.69	5,953	94.06
Third trimester				
B2AR	85	1.34	269	4.25
Non-B2AR medication	58	0.92	59	0.93
Both B2AR and non-B2AR medications	33	0.52	10	0.16
No asthma medication	106	1.67	5,795	91.56

*percentages are based on total table N=6,329.

4. Potential Risk Factors Prevalences in the Preconception Period (measured in maternal and paternal claims)

Tables 29 through 32 present the estimated prevalences of potential preconception risk factors measured in maternal and paternal claims in the ASD and comparison groups. Again, for all potential risk factors in both of these tables, prevalences did not vary meaningfully across the continuous enrollment criteria subsamples (data not shown), suggesting that these criteria would not introduce selection bias with respect to these potential risk factors.

In the ASD group (Table 29) the prevalence of infertility treatment in mothers was within the range of published estimates when all coding source data (ICD-9, CPT, and medication) were used, although higher than the one national average (1.4%) when comparing prevalence using CPT codes alone. However, it should be noted that the national estimate shown is the proportion of all pregnancies in the year 2000 in five states that resulted from assisted reproductive technology (ART) while our definition includes other types of infertility treatment as well. The prevalence of infertility treatment in fathers of children with ASD when all coding source data (ICD-9, CPT, and medication) were used was 0.5%; lower than the one estimate located in the

literature (1.4%) based strictly on ICD-9 codes. Estimates in the comparison population (Table 30) showed comparable patterns.

Similarly to the prenatal period, the codes for obesity during preconception was reported at low levels (1% and 2%) for both the ASD and comparison group mothers and confirm that claims are likely not well-suited for a baseline measurement of preconception obesity and/or excessive weight gain (Tables 29 & 30).

Finally, this data also includes information on the prevalence of autoimmune conditions in fathers assessed from their index date to end of eligibility (Tables 31 & 32). This is considered a preconception risk factor in that the role of paternal autoimmune condition in ASD etiology would likely be through some shared genetic susceptibility mechanism (physiologic consequences of autoimmune disease in fathers, unlike mothers, would not be etiologically significant). Similar to the result seen for prenatal autoimmune conditions in mothers, prevalence estimates here were above expectation, with the exception of the definition using two or more claims with the autoimmune condition diagnoses codes from index date to conception, suggesting that some refinement or increased specificity of coding criteria may also be needed here.

Table 29. Mothers and Fathers of Children with ASD with infertility treatment Risk Factor

	Mothers CE (birth-40 weeks) to birth (N=1,206)			Estimate from Literature		Fathers (N=1,513)			Estimate from Literature	
				ASD	National				ASD	National
	valid N	n	%	% (range)	% (range)	valid N	n	%	% (range)	% (range)
Infertility treatment from (birth - 92 weeks) to (birth - 40 weeks) (ICD9, CPT, or Rx)	576	51	8.85	2.3,14.1%		752	4	0.53		
ICD9	576	2	0.35			752	0	0.00		1.4%
CPT	576	34	5.90		1.4%	752	2	0.27		
Medication	576	37	6.42			752	3	0.40		
Maternal Obesity Preconception	576	6	1.04	7.6%, 9.0%	27.4%					

Table 30. Mothers and Fathers of Comparison Children with infertility treatment Risk Factor

	Mothers CE (birth-40 weeks) to birth (N=6,329)			Estimate from Literature		Fathers (N=7,204)			Estimate from Literature	
				Non-ASD	National				Non-ASD	National
	valid N	n	%	% (range)	% (range)	valid N	n	%	% (range)	% (range)
Infertility treatment from (birth - 92 weeks) to (birth - 40 weeks) (ICD9, CPT, or Rx)	3,285	157	4.78	5.4%,7.9%		3,867	16	0.41		
ICD9	3,285	8	0.24			3,867	0	0.00		1.4%
CPT	3,285	88	2.68		1.4%	3,867	8	0.21		
Medication	3,285	118	3.59			3,867	8	0.21		
Maternal Obesity Preconception	3,285	72	2.19	7.0%, 10.0%	27.4%					

Table 31. Fathers of Children with ASD with autoimmune condition Risk Factor

	Fathers (N=1,513)			Estimates from the Literature	
	valid N	n	%	ASD % (range)	National % (range)
Paternal autoimmune conditions					
1+ diagnosis				5.0-8.6%	5.0-8.0%
Index date to Conception	752	123	16.36		
Conception to end of CE	1,513	597	39.46		
2+ diagnoses					
Index date to Conception	752	34	4.52		
Conception to end of CE	1,513	308	20.36		

Table 32. Fathers of Comparison Children with autoimmune condition Risk Factor

	Fathers (N=7,204)			Estimates from Literature	
	valid N	n	%	Non-ASD % (range)	National % (range)
Paternal autoimmune conditions					
1+ diagnosis				2.0-4.5%	5-8%
Index date to Conceptions	3,867	629	16.27		
Conception to end of CE	7,204	2,001	27.78		
2+ diagnoses					
Index date to Conceptions	3,867	170	4.40		
Conception to end of CE	7,204	940	13.05		

V. Discussion

This report sought to provide an initial exploration of the potential of using study populations assembled from, and information contained in, administrative private health care claims data to conduct meaningful research on potential ASD risk factors. In risk factor research, the relevant windows for potential risk factors of etiologic influence include preconception, prenatal, postnatal, and the early life of a child. Consequently, samples from claims data will not only need to have enrollment experience sufficient to allow assessment of the ASD outcome but also relevant claims experience covering these windows. This implies that a risk factor investigation where the primary period of interest is in the period before the birth of the child, will need to involve data on children themselves enrolled at birth and potentially also linked to parents who also have an enrollment period covering a period beginning at the child's birth date and extending back from there.

Given this, a risk factor researcher would most likely approach a claims database by first ascertaining a birth cohort sample meeting specific eligibility criteria. The criteria would likely include linkage of parent and child, parental enrollment for a defined period prior to birth, and child enrollment for a defined period from birth through which ASD diagnosis may be expected to be reliably identified through a claims-based algorithm. From this cohort the rate of children with evidence of ASD would be contrasted to groups with different risk factor profiles. However, in this Task, we began with the same basic samples used in previous project Tasks. Consequently, rather than a prospective birth cohort approach, here we examine data in samples defined by ASD case status and estimate risk factor prevalence prior to ASD diagnosis. The inferences drawn from the contrasts we make, nonetheless, are applicable to a cohort design that could be implemented in a claims-based study of ASD risk factors.

A. Generalizability (External Validity)

Privately insured populations may not be demographically representative of the entire US population regarding risk factors and risk factor prevalences though they do represent the majority of individuals. When comparing to the literature, some of which presumably included individuals without private insurance, our results were often comparable. We did, however, undertake empirical analysis of the generalizability, or lack thereof, of private insurance samples as part of our Task A: Baseline Claims Analysis study and found that both our comparison group of children and children with ASD were similar to the broader privately insured population in the US with regards to age and gender.

However, it is also likely that our privately insured study samples (with and without ASD) are not representative of the entire US population in that the privately insured population is generally healthier, has better access to care, has higher income, and is less racially and ethnically diverse than the US population as a whole.²⁶ Consequently, this report does not address the extent to which findings, even if valid for a private insurance sample, apply or do not apply to populations dissimilar to those receiving private insurance. See Task A: Baseline Claims Analyses Report submitted to NIMH on October 17, 2011 for more information.

B. Potential for Selection Bias and Study Size Considerations

We conducted a series of comparisons that provided data on the likelihood of potential selection effects in the manner in which samples would be constructed in claims data. For selection effects to introduce selection bias, selection into a sample has to be differential with respect to both the risk factor of interest and the outcome. When we examined the sequence of steps used to retain those children who could be linked to an apparent mother enrolled at the time of the child's birth, the proportions of children retained at each step were similar in the ASD and comparison groups. This suggests that a requirement of linkage to mothers is not an overt source of selection bias. Also reassuring in this regard was the fact that the demographic profile of those children linking to mothers was similar to the Task A "Likely ASD" base sample and the utilization experience and comorbidity levels were similar to all children who were enrolled at birth (regardless of whether they linked to a mother). We saw similar results and drew similar conclusions with respect to linkage to apparent fathers enrolled at the estimated time of conception.

The linkage requirement does, however, have a major impact on the size of study samples and, consequently, the power and precision of analyses. Only 4% of ASD and comparison group children were linked to mothers and enrolled at birth. Large segments of children were excluded because no enrolled female met FAMID and age requirements used to 'find' a mother and because even among identified mothers, women often did not have enrollment at the time of the child's birth. Nonetheless, this series of nine years of claims data gave rise to more than 2,000 ASD cases enrolled at birth and linked to mothers who were also enrolled at that time.

For some risk factors, specifically delivery and early life risk factors, enrollment at birth may be sufficient to capture data of interest but we also needed to consider the effect of requiring enrollment extending over longer periods. Only one-quarter of children with ASD (and similar proportions of the comparison group) who were linked to mothers who were enrolled all the way back over the entire prenatal and preconception periods; however, about three-quarters had mothers with enrollment extending back over the third trimester. So, the number of children with ASD available for etiologic investigation depends greatly on the extent of enrollment coverage needed to capture particular risk factors of interest. For studies of early childhood risk factors where linkage to parents is not necessary, numbers of children increased approximately 25%. As was the case with linkage to parents, restriction of samples to various continuous enrollment windows led to similar changes in sample size for the ASD case and comparison groups and resulted in samples with a similar distribution of characteristics compared to the Task A "Likely ASD" base sample with enrollment at birth.

C. Information Bias

Whereas selection bias is a byproduct of the definition and assembly of a sample, information bias results from inaccurate information on the factors under study, including risk factor measures, outcomes and relevant covariates. Task A included a criterion validity study evaluating the accuracy of claims-based ASD definitions against medical record review. Task D, on the other hand, included analyses that explored information accuracy issues with respect to a range of potential ASD risk factors. These did not include formal criterion validity studies but, as mentioned above, did include estimation of risk factor prevalence and comparison to published data on ASD, non-ASD, and general population samples. It should be noted that while the best comparisons to our prevalence estimates are prevalences available in the published literature, the

literature may not be a perfect ‘gold standard’. Additionally, the literature review was not all-encompassing and in some cases, our comparison is only based on one estimate from the literature. Before discussing the implications of these findings, two general topics related to accuracy of claims-based risk factor data are discussed.

First, because a number of risk factors were measured in parental claims, errors in establishing accurate linkage would translate into errors on risk factor status (e.g., an ASD subject linked to the wrong woman as the mother would generate potentially inaccurate data on prenatal risk factors). While, we did not undertake any empirical assessment of the accuracy of these linkages we applied the STORK methodology that was developed and used extensively to establish accurate linkages between mothers and newborns in claims data research minimizing the errors in establishing accurate linkage. However, validating this or another approach remains an area for future investigation with respect to the utility of claims data for ASD risk factor research.

Next, because the timing of risk factor exposure can be of importance in establishing biologic plausibility or when considering mechanisms underlying detected associations in epidemiologic analyses, we wanted to consider the accuracy of our ability to establish trimester cutpoints using claims data. As has been done in other claims-based research on pregnancy data, we employed a relatively simple algorithm to estimate date of conception and, from there, determine cutpoints for second and third trimesters. We did not validate these dates and cutpoints against an external gold standard (like medical records). We did explore whether codes for tests and procedures that typically occur at fixed points in pregnancy shed any light on the accuracy of our approaches to estimate date of conception and subsequent trimester cutpoints. We found that our results for procedures during pregnancy corresponded well to the trimesters in which they are expected to occur. These results suggest that the timing of procedures in claims data could then be used to validate or improve the precision of etiologic windows for risk factors although there are very few procedures, like the inhibin screen, that are reliably administered during narrow enough time frames to potentially improve the ability to determine exact gestational age for risk factor research. Of note, a recently published paper took a similar approach using claims data from the British Columbia Medical Services Plan.²⁷ They contrasted gestational age estimated through algorithms based on an assumed fixed term and preterm pregnancy lengths to algorithms incorporating date of occurrence from four screening procedures that are recommended to occur at targeted gestational ages. The four procedures were alpha-fetoprotein testing, guided amniocentesis, ultrasound at 14+weeks and gestational diabetes screening. Unlike our study, these Canadian researchers also had access to clinical birth data which they considered as a gold standard, i.e the clinical gestational age determination that was recorded at birth. Their analyses contrasted each algorithm’s estimate with the clinical gold standard, focusing on the proportion in which the algorithm-based gestational age was within one week of the clinical gestational age. They concluded that the addition of date information from screening procedures did not substantively improve gestational age estimation. This suggests that adding in the procedure date data may not improve our conception date and trimester cutpoint estimation. Nonetheless, a validation against clinical gestational age in our private claims database would still be a useful next step to confirm that the findings from the British Columbia health system generalize to the US.

With respect to the estimation of the prevalence of various potential risk factors of interest using our claims data, a number of key findings emerged. First, although we examined risk factor prevalence in samples with differing continuous enrollment requirements when relevant (i.e., the

prevalence of maternal prenatal infection in the third trimester was estimated in each of the three linked sample groups where mothers had continuous enrollment extending back 14, 40 and 92 weeks before birth) this consistently did not substantially affect prevalence. So despite decreases in the number of subjects as continuous enrollment requirements were extended, these smaller sample subsets generated similar estimates to the larger linked samples. This implies that researchers may not need to be overly concerned about the impact of enrollment criteria on exposure estimation – though the sample size impact and the flexibility provided to consider different kinds of risk factors in a range of etiologic windows may still be of importance.

The potential ASD risk factors where our claims-based prevalence estimates tended to be consistent with published estimates from the literature included preterm birth, chronic maternal health conditions potentially initiating prior to pregnancy (e.g., asthma and depression), medication use for those conditions, anesthesia use, maternal infertility treatment, and MMR immunization (note that we explored DPT immunization prevalence but do not present these results due to limitations with one of the CPT codes associated with this vaccination). With respect to medication use, our exploration also indicated that claims data have the potential to allow for empirical analysis of confounding by indication (separating effects of the medication from effects of the underlying diagnosis). The fact that maternal infertility prevalence was within the range of published estimates was somewhat surprising given that coverage of these procedures can vary considerably across different health plans (paternal infertility procedure prevalence, on the other hand, was much lower than published estimates which could reflect inconsistent coverage).

We examined some risk factors that, *a priori*, we expected would not be measured accurately in claims (e.g., obesity – which is not consistently coded in claims; and pitocin – which is known to be bundled into other procedure codes) - and our analyses confirmed that these are likely not well-captured through claims. There were, however, other potential risk factors where we hoped claims data would yield results comparable with published estimates. For example, NICU and ICU admissions were much lower in our ASD and comparison subsamples than in published data. If this is driven primarily by the fact that our privately insured sample is at lower risk for these events than other US populations, our sample would still have internal validity for estimation of associations between these factors and ASD risk but findings may be less generalizable to other populations. There were also a number of potential risk factors where prevalence estimates were higher than expected in our sample. In some instances we suspect that coding issues might be driving these differences. Our range of codes may have been too broad in some instances (e.g., autoimmune disease, prenatal infections – See Appendix B for full list of codes) and in other instances coding practice may be such that codes on our lists are used for rule-out diagnoses and work ups, rather than as an indication that the diagnosis is confirmed (e.g., gestational diabetes and preeclampsia). In other instances (e.g., ultrasound and antibiotic use), the higher prevalences we observe could be due to the composition of our study population and/or the influence of private insurance.

Finally, we explored the potential of identifying groups of children who were completely free of vaccinations, a cohort that would be useful to explore continued concerns about links between ASD and immunization. While we found that nearly five percent of the children in both our ASD and comparison children samples did not have evidence of any immunization, this proportion is many times higher than reported as unvaccinated in the literature, suggesting that this most likely does

not represent a truly unvaccinated cohort and that many children included in this group actually received vaccines that either did not generate a claim or that were paid for by other sources.

D. Conclusion

This analysis demonstrated that the mechanics needed to select samples from large US private insurance claims databases to allow risk factor research for childhood ASD do not appear to introduce selection effects that will lead to selection bias. However, and not unexpectedly, the numbers of informative children with ASD with linkage to parents and sufficient continuous enrollment to allow meaningful investigation are only a small proportion (<10%) of the very large number of children with ASD that can be located in a cross-sectional query of a large claims database. Nonetheless our examination of nine years of claims data still generated two to three thousand ASD cases that would be of potential use in etiologic research. This is a very similar “yield” to large registry-based studies done in Scandinavia (e.g., Atladottir and colleagues²⁸ queried 22 years of registry data to generate a sample including approximately 7,000 ASD cases) that have been major contributors to knowledge about ASD epidemiology in recent years. While it is clear that private insurance claims data will not be an adequate information source for a number of potential ASD risk factors, we saw that there was also a wide range of potential risk factors where claims-based research could have the capacity to add to the developing epidemiologic knowledge base.

In general, claims would appear to be a viable data source for investigation of maternal medical conditions, that require active medical management (e.g., asthma, depression) and their treatments (in particular, pharmacologic therapies). Prevalence estimates for these conditions estimated in claims tended to agree with published estimates. Further, the size of the claims database creates opportunities to control for confounding by indication as there appear to be sufficient numbers of both treated and untreated women with such conditions identifiable. Similarly, serious early postnatal complications (NICU admission, pre-term delivery) also appear amenable to investigation through claims. Risk factors explored that might be worth additional exploration include infertility (where refinement to include only plans with certain benefits policies might improve the accuracy of claims-based assessment) and parental medical conditions with somewhat less-intensive medical management. We saw a higher than expected prevalence of indicators of maternal infection in pregnancy and autoimmune disease history which suggests that some coding refinement is necessary but, at least in the case of infection, claims data that capture outpatient encounters likely represent an improvement over Scandinavian registry studies that focus on inpatient diagnoses only. Finally, chronic conditions requiring less active medical management (e.g., maternal obesity) would not appear to be good candidates for claims-based investigation and, though our data suggested the presence of a fairly large unvaccinated group, the possibility of over-estimation is real given the likelihood that vaccines are bundled with other primary care services. It should be noted again that the potential risk factors selected here were chosen to represent a broad range of different variable types occurring in different subjects (parents, child) and those examined should not be considered more, and those unexamined should not be considered less, plausible than others potentially measurable in claims that we did not consider. In addition, future studies using claims to explore potential ASD risk factors might seriously consider the incorporation of formal validation sub-studies on both exposure and diagnosis (gathering data on exposure and diagnosis from other data sources on a fraction of the sample), since these can often be implemented on a reasonable timetable and at a

fraction of the cost of studies of comparable size that would require primary data collection on every subject.

VI. References

- ¹ Newschaffer C, Croen LA, Daniels J, Giarelli E, et al. The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health*. 2007;28:21.1–21.24.
- ² Centers for Disease Control and Prevention. “Prevalence of Autism Spectrum Disorders --- Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008.” Available at: <http://www.cdc.gov/mmwr/pdf/ss/ss6103.pdf>. Accessed March 30, 2012.
- ³ Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, Veenstra-Vanderweele J. A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*. 2011;127:e1303-e1311.
- ⁴ Interagency Autism Coordinating Committee. 2011 IACC Strategic Plan for Autism Spectrum Disorder Research. 2011 Jan. Retrieved from the Department of Health and Human Services Interagency Autism Coordinating Committee website. Available at: <http://iacc.hhs.gov/strategic-plan/2011/index.shtml>.
- ⁵ Baine W. Systematic screening of secondary diagnoses in Medicare administrative data to identify candidate risk factors for the principal diagnosis. *Annals of Epidemiology*. 2003;13(6):443-449.
- ⁶ Russell M, Taylor D, Cummins G, Huse D. Use of managed care claims data in the risk assessment of venous thromboembolism in outpatients. *American Journal of Managed Care*. 2002;8(1 Suppl):S3-9.
- ⁷ Mandell DS, Morales KH, Xie M, Lawer LJ, Stahmer AC, Marcus SC. Age of diagnosis among Medicaid-enrolled children with autism, 2001-2004. *Psychiatric Services*. 2010;61(8):822-829.
- ⁸ Mandell DS, Morales KH, Marcus SC, Stahmer AC, Doshi J, Polsky DE. Psychotropic medication use among Medicaid-enrolled children with autism spectrum disorders. *Pediatrics*. 2008;121(3):E441-E448.
- ⁹ Wang L, Leslie DL. Health care expenditures for children with autism spectrum disorders in Medicaid. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2010;49(11):1165-1171.
- ¹⁰ Leslie DL, Martin A. Health care expenditures associated with autism spectrum disorders. *Archives of Pediatrics & Adolescent Medicine*. 2007;161(4):350-355.
- ¹¹ Casanova MF. Neuropathological and genetic findings in autism: the significance of a putative minicolumnopathy. *Neuroscientist*. 2006; 12(5), 435-441.
- ¹² Berg JM, & Geschwind DH. Autism genetics: searching for specificity and convergence. *Genome Biology*. 2012; 13(7), 247.
- ¹³ Hallmayer, J., et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of general psychiatry*. 2011;68(11),1095-1102.
- ¹⁴ Landa RJ, Gross AL, Stuart EA, Bauman M. Latent class analysis of early developmental trajectory in baby siblings of children with autism. *J Child Psychol Psychiatry*. 2012;53(9):986-96.
- ¹⁵ Smith LE, Maenner MJ, Seltzer MM. Developmental trajectories in adolescents and adults with autism: the case of daily living skills. *J Am Acad Child Adolesc Psychiatry*. 2012;51(6):622-31.

- 16 Centers for Medicare & Medicaid Services. "Medicare Claims Processing Manual, Chapter 26: Completing and Processing Form CMS-1500 Data Set." Available at: <https://www.cms.gov/manuals/downloads/clm104c26.pdf>. Accessed September 7, 2011.
- 17 Cecil G. Sheps Center for Health Services Research. "Implementation of the UB-04." Available at: http://www.shepscenter.unc.edu/research_programs/hosp_discharge/links/ub04_fact_sheet.pdf. Accessed September 7, 2011.
- 18 National Uniform Billing Committee. "History of the NUBC." Available at: <http://www.nubc.org/history.html>. Accessed September 7, 2011.
- 19 Shimabukuro, T, Grosse, S, Rice, C. Medical expenditures for children with an autism spectrum disorder in a privately insured population. *J. Autism Dev Dis.* 2007;546-552.
- 20 Commission to Build a Healthier America. "Education Matters for Health." Available at: <http://www.commissiononhealth.org/PDF/c270deb3-ba42-4fbd-baeb-2cd65956f00e/Issue%20Brief%20%20Sept%2009%20-%20Education%20and%20Health.pdf>. Accessed July 10, 2012.
- 21 Ettner SL. New evidence on the relationship between income and health. *J Health Econ.* 1996;15:67-85.
- 22 Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care.* 2005; 43:1130-39.
- 23 Feudtner C, Christakis DA, Connell FA. Pediatric deaths attributable to complex chronic conditions: a population-based study of Washington State, 1980-1997. *Pediatrics.* 2000;106:205-209.
- 24 Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, Levy SE, Mandell DS, Miller LA, Pinto-Martin J, Reaven J, Reynolds AM, Rice Ce, Schendel D, Windham GC. The epidemiology of autism spectrum disorders. *Annu Rev Public Health.* 2007;28:235-58.
- 25 Cole JA, Model JG, Haight BR, Cosmatos IS, Stoler JM, Walker AM. Bupropion in pregnancy and the prevalence of congenital malformation; *Pharmacoepidemiology;* 2007;16:474-84.
- 26 DeNavas-Walt C, Proctor BD, Lee, CH. "Income, Poverty and Health Insurance Coverage in the United States, 2005," United States Census Bureau. 2006. Available at <http://www.census.gov/prod/2006pubs/p60-231.pdf>. Accessed September 2, 2011.
- 27 Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf.* 2012.
- 28 Atladóttir HO, Thorsen P, Schendel DE, Østergaard L, Lemcke S, Parner ET. Association of hospitalization for infection in childhood with diagnosis of autism spectrum disorders: a Danish cohort study. *Arch Pediatr Adolesc Med.* 2010;164(5):470-7.