



# HHS Public Access

Author manuscript

*J Autism Dev Disord.* Author manuscript; available in PMC 2015 July 23.

Published in final edited form as:

*J Autism Dev Disord.* 2013 November ; 43(11): 2650–2663. doi:10.1007/s10803-013-1815-x.

## The International Collaboration for Autism Registry Epidemiology (iCARE): Multinational Registry-Based Investigations of Autism Risk Factors and Trends

**Diana E. Schendel,**

Department of Public Health and Department of Economics and Business, University of Aarhus, 8000 Aarhus C, Denmark

**Michaeline Bresnahan,**

Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA. New York State Psychiatric Institute, New York, USA

**Kim W. Carter,**

Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, WA, Australia

**Richard W. Francis,**

Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, WA, Australia

**Mika Gissler,**

National Institute for Health and Welfare, Helsinki, Finland. Nordic School of Public Health, Gothenburg, Sweden. Turku University, Turku, Finland

**Therese K. Grønberg,**

Department of Public Health, University of Aarhus, Aarhus, Denmark

**Raz Gross,**

Division of Psychiatry, Sheba Medical Center, Tel Hashomer, Israel. Department of Epidemiology and Preventive Medicine, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel

**Nina Gunnes,**

Norwegian Institute of Public Health, Oslo, Norway

**Mady Hornig,**

Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA. Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, USA

**Christina M. Hultman,**

Karolinska Institutet, Stockholm, Sweden

**Amanda Langridge,**

Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, WA, Australia

**Marlene B. Lauritsen,**

Research Unit for Child and Adolescent Psychiatry, Aalborg Psychiatric Hospital, Aarhus University Hospital, Aalborg, Denmark

**Helen Leonard,**

Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, WA, Australia

**Erik T. Parner,**

Department of Public Health, University of Aarhus, Aarhus, Denmark

**Abraham Reichenberg,**

Department of Psychosis Studies, Institute of Psychiatry, King's College London, London, UK.  
Department of Psychiatry, Mount Sinai School of Medicine, New York, USA

**Sven Sandin,**

Karolinska Institutet, Stockholm, Sweden

**Andre Sourander,**

Child Psychiatry Research Center, Department of Child Psychiatry, Turku University and Turku University Hospital, Turku, Finland

**Camilla Stoltenberg,**

Norwegian Institute of Public Health, Oslo, Norway

**Auli Suominen,**

Department of Child Psychiatry, Turku University, Turku, Finland

**Pål Surén, and**

Norwegian Institute of Public Health, Oslo, Norway

**Ezra Susser**

Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA. New York State Psychiatric Institute, New York, USA

Diana E. Schendel: [diana.schendel@folkesundhed.au.dk](mailto:diana.schendel@folkesundhed.au.dk)

## Abstract

The International Collaboration for Autism Registry Epidemiology (iCARE) is the first multinational research consortium (Australia, Denmark, Finland, Israel, Norway, Sweden, USA) to promote research in autism geographical and temporal heterogeneity, phenotype, family and life course patterns, and etiology. iCARE devised solutions to challenges in multinational collaboration concerning data access security, confidentiality and management. Data are obtained by integrating existing national or state-wide, population-based, individual-level data systems and undergo rigorous harmonization and quality control processes. Analyses are performed using database federation via a computational infrastructure with a secure, web-based, interface. iCARE provides a unique, unprecedented resource in autism research that will significantly enhance the ability to detect environmental and genetic contributions to the causes and life course of autism.

## Keywords

Autism; Epidemiology; Study methods; Risk factors; Multinational

Autism spectrum disorders (ASD) are a group of neurodevelopmental disorders affecting about 1 % of children (Centers for Disease Control and Prevention 2012) that are defined by qualitative impairments in social interaction and communication and the presence of restricted interests and repetitive behaviors (American Psychiatric Association 2000). Autism was differentiated as a specific set of child disorders in the 1940s, and early work considered mother–child interactions as potential causes (Bettelheim 1972). By the 1980s, however, studies had established that there was a strong genetic contribution to ASD (Folstein and Rutter 1977; Steffenburg et al. 1989; Ritvo et al. 1989). Since then, etiological research in ASD has focused primarily on genetic factors. The marked increase in ASD prevalence as well as lower heritability estimates for autism reported recently, however, suggest that non-genetic and environmental factors may also be prominent in ASD etiology (Fombonne 2009; Hallmayer et al. 2011). The etiological complexity in autism is shared with other multifactorial complex disorders that prompted the establishment of multinational consortia to facilitate data sharing and collaborative analyses on pooled data (e.g., Society for Cerebral Palsy in Europe 2000; Muilu et al. 2007; Sullivan 2010).

The International Collaboration for Autism Registry Epidemiology (iCARE) was established as a multinational consortium to promote and facilitate research on ASD. It is a unique resource in ASD research: seven multinational partners contribute existing national- or state-wide prospectively-collected data for analyses using data federation approaches. The resource occupies a unique niche for: (1) analyses that require very large sample sizes to detect effects that could not be achieved with single data systems and (2) direct analytical comparison of findings across populations that differ in sociocultural norms, potential sources of bias, geography, familial characteristics and temporal trends. Both large sample sizes and comparisons across populations are often necessary (though not sufficient) for causal inference from epidemiological studies.

The International Collaboration for Autism Registry Epidemiology (iCARE) was established among collaborators in Australia, Denmark, Finland, Israel, Norway, Sweden, and the USA with the goal of conducting multinational population-based autism epidemiological research which is enabled by a data federation infrastructure. Achieving this goal established a wide range of potential benefits including: (1) cost efficiency through use of existing resources; (2) flexible infrastructure accommodating current research needs and future network growth and data upgrades; (3) flexibility in study designs to suit particular analyses (e.g., cohort, case-cohort, multigenerational or sibling designs); (4) large sample sizes based on federated data that would enhance statistical precision, especially in individual strata based on ASD phenotype characteristics or risk factors; (5) ability to characterize population trends in reported diagnoses over time (e.g., by age at reporting, birth cohort or time period), as well as changes over the life course of affected individuals; and (6) enhanced comparison and interpretation of between-site results based on data harmonization and application of uniform analytic methods to multi-site data.

The purpose of this paper is to describe iCARE collaboration and resources, including site characteristics, cross-site data harmonization methods and our federated dataset approach to multi-site analysis.

## **iCARE Structure and Site Characteristics, Including ASD Data Sources**

### **Organization**

An iCARE memorandum of agreement (MOA) articulated the guiding principles and practical processes for the multinational collaborative activities. The MOA (available on request) describes the structural framework for the collaborative arrangements (terms of membership, member roles and responsibilities, governance and communication guidelines) and outlines the processes of implementing collaborative analyses, assembly and sharing of data, authorship and publication. Principles underlying successful collaboration, such as trust, open communication and respect, were core features of the MOA language and were intended to protect the time and resource investment of the group as well as rights of the individual members; these were believed to be especially important considerations in a multinational setting. iCARE administrative and scientific oversight is provided by the analytical oversight committee (AOC) on an ongoing basis and is comprised of representatives with voting authority from all sites. The AOC reviews and approves all proposals for analyses prior to their initiation. iCARE member sites and roles are described in Table 1.

### **Site and Population-Register Characteristics**

All six data contributing sites have complete population coverage (national or state-wide) and can identify a local iCARE study population as a complete birth cohort (e.g., via birth registries that record all births in the population) (Table 2). All sites also have existing, ongoing, population-wide health information systems that were established for administrative, health, socioeconomic, or demographic purposes. Typically these data collection systems are longstanding, independently and continuously maintained and contain individual-level information of interest for initial iCARE investigations such as pre- and perinatal data (e.g., parental age, birth weight, gestational age, parity, multiplicity, apgar score, mode of delivery, fetal presentation), medical diagnoses (e.g., genetic, developmental or psychiatric conditions, infections, birth defects), death, immigration status, social outcomes, and sometimes biological specimens (i.e., biobanks) or other more specialized information (e.g., prescribed medications). The unique personal identification number available to each site (and encrypted in Western Australia) permits linkage and merging of individual data within and across data systems at a site, including linkages between family members (although no personal identification numbers are retained in any iCARE dataset). Universal health care coverage that is publically financed and utilized is available to all citizens of Denmark, Finland, Israel, Norway and Sweden and contacts with the public health care system are required to be reported. In Western Australia both public and private health care provisions are available although both public and private health care providers are required to report to the public health system. This means that for iCARE purposes health information from registries across all sites is not biased by differential access to health care and is population-based. The registry data across sites, however, do differ on the

basis of the types of contacts with the health care system that are reported to the registries. These differences of relevance to autism data are described in greater detail below.

### ASD Data Characteristics

Three sites (Denmark, Finland, Sweden) retrieve autism diagnosis data from government-maintained medical registries that record diagnoses or procedures from each in- or outpatient clinic or hospital contact (Table 2). These registry data do not include contact with general practitioners who serve as primary care providers; general practitioners refer children to clinic- or hospital-based specialists (e.g., pediatricians, child psychiatrists) for diagnosis or treatment of conditions outside of routine primary care. Since it would be quite rare for a child with ASD to be diagnosed and treated only by a general practitioner, the absence of these contacts in the registry data is unlikely to impact iCARE analyses. Contacts with private health care providers are also not included in these registries, but these are also rare in the context of universal health care in these countries. For the other three sites (Israel, Norway, Western Australia) diagnostic information is derived from government-maintained service/benefits registries that record contacts with individuals receiving services/benefits for autism (Table 2). Since diagnosed individuals who do not receive services/benefits are not included in these sites' data, the omission could impact iCARE pooled analyses or across site comparisons. We describe below the proposed approaches to assess and mitigate the potential impact of this source of site data variability on ICARE results. Also, the data from Israel and Western Australia do not include data from the non-Jewish and Aboriginal populations, respectively. Due to cultural differences and barriers to access to the health care system, individuals in these minority groups may be substantially under-represented in the registry data. For example, in Western Australia a formalized multidisciplinary approach to diagnosing children with ASD has been operating since 1991 and involves multiple assessments by a pediatrician or psychiatrist, psychologist and speech therapist over a period of several hours. This approach does not accommodate well the needs and situations of Aboriginal families and makes it likely that Aboriginal children may be underdiagnosed. Although it is of great interest to study these minority groups, their under-ascertainment in registries precludes us from using current iCARE data for this purpose because our results would be biased.

Whereas the individual steps leading to the recording of an autism diagnosis in the data sources are specific to each country, the overall process is fairly uniform across the different countries and strengthened by public health care provisions wherein access to the services that lead to the reporting of a diagnosis is free of charge and universally available. Thus, across all countries, as shown in Table 3, if a child's health care or education provider or parent suspects the child to have a developmental problem then the child is referred to specialist care where s/he is evaluated, typically by a multidisciplinary team (including specialist observations, interviews with parents and/or administration of standardized tests). If an autism diagnosis (ASD or one of the diagnostic subtypes of ASD) is made based on the evaluation then the diagnosis is reported by the specialist to the relevant data system or, as in Norway, approval of the application by a parent for benefits for autism following the specialist evaluation and subsequent receipt of benefits prompts the recording of the diagnosis in the database.

Validation studies of the reported ASD diagnosis to the data source have been, or are being, performed in four sites with positive predictive value of over 90 % between the recorded and validation study diagnoses at the three sites with published or provisional results (Table 3). Table 3 also lists potential sources of variation in the reporting of ASD diagnoses to the data source within each site. All sites have had marked increases in the reporting of ASD diagnoses in recent years and common sources of variation that at least partly contribute to this change include the transitions between ICD versions (e.g., ICD-8 to ICD-9 or ICD-10) and the introduction of reporting diagnoses made at outpatient clinic or hospital contacts where previously only those diagnoses made at inpatient contacts had been reported. In the service/benefits-based registry systems there are potential sources of variation due to changes in the reporting process over time, or by age of child or specific ASD diagnoses that are tied to service/benefits (Nassar et al. 2009).

To illustrate the iCARE ASD sample size potential, Table 4 provides, by site, the numbers of individuals, prevalence and sex ratio for ASD and autistic disorder for two birth cohorts. Overall, across all 6 sites, 23,314 individuals with an ASD were ascertained from 3.6 million births in the years 1987–1996 and 13,422 individuals with an ASD were ascertained from 3.0 million births in 1997–2004 (with follow-up for ASD identification for both cohorts from birth through 2009 at 4 sites, through 2006 in Norway and through 2004 in Western Australia). For 1987–1996 births, the ASD prevalences were highest in Denmark, Sweden and Finland (from 7.4 to 9.5 per 1,000), somewhat lower in Western Australia (3.3 per 1,000), and lowest in Israel and Norway (1.4–1.7 per 1,000 births); note the Israel and Norway data are based on service/benefits registries and not full population diagnostic registries. Somewhat lower prevalences were observed in the more recent cohort but with a similar between-site pattern; the lower prevalences are most likely due to less follow-up time for identification (Parner et al. 2008) since other factors that might impact reported prevalence either didn't change during the time period of follow-up (e.g., change in diagnostic criteria or exclusion of minority data from iCARE datasets) or were likely to increase reported prevalence (e.g., addition to registries of diagnoses made in outpatient contacts, increasing public awareness or decreasing age of diagnosis).

### Data Access

Each iCARE data contributing site has obtained all necessary ethical and administrative approvals for data access, applying for data from local data systems, and managing and preparing the local data for iCARE use, including secure data sharing. Each site's application for approval outlines the means by which it would adhere to rules governing data access and use, rules based partly on citizen perspectives. It is important to note that the population-based registry data, including personal identifiers, are reported to the registries as part of routine business and without individual consent. Further, a researcher's access to the registry data based on an approved application does not require individual consent. Under these circumstances, to preserve confidentiality, no personal identifying information is permitted in any iCARE harmonized analytic datasets. At one site ethical approval only permitted sharing of aggregated data for iCARE analyses. To satisfy this restriction, the site prepares the local dataset using individual-level data per the iCARE protocol and then

calculates summary counts depending on the needs of an analysis. The summary counts constitute this site's iCARE data for multi-site analyses.

## Across-Site Data Harmonization Methods

The overall goal of the harmonization process is to build a data platform of uniform datasets, one at each site, comprised of harmonized variables and to obtain information from each site that permits the identification of potential data artifact and sources of bias.

The harmonization process involves three steps: first, information surveys are completed by each site for each variable of interest and corresponding data source(s); second, the information is summarized and used to create a data dictionary with instructions for site data managers on constructing harmonized variables; third, a quality control procedure is implemented prior to use of the site's harmonized dataset in analyses wherein the program logs of variable creation along with resulting data tables from each site are reviewed for accuracy. The local datasets used for harmonization are fixed snapshots of registry data at a particular point of time. The harmonization process can be repeated following registry data updates (new "snapshots"), the addition of new variables, or variable modifications.

### Information Surveys

The information surveys inform development of the harmonization protocol for a specific set of variables and insure that proposed analyses with any given variable can take into account the availability and comparability of the variable across sites and facilitate across-site results interpretation. The surveys request detailed information for each target variable: availability over time; registry source(s); methods of data collection; data quality; preferred data source for a variable if there are multiple sources; data source modifications over time (e.g., method of collection); reasons for missing data that may be systematic; and variable name, description, data format, and values. Sites supply details regarding any changes in this information over time, e.g., variable name or value changes.

### Data Dictionary

On the basis of the information surveys, a data dictionary is prepared specifying the name, definition and eligible range and values for each iCARE variable, including specifications for assigning missing values, and programming syntax for complex variables. To illustrate, the most recent data dictionary specifications for harmonizing ASD diagnostic data are provided in Table 5. Each site then creates a local dataset in accordance with the data dictionary specifications.

### Quality Control

For purposes of quality control (QC) each site generates a program log of the local dataset creation and produces descriptive statistics for each created variable; these are reviewed by the Data Management site to uncover programming errors or data inconsistencies. The same descriptive statistics are generated after the site has uploaded the dataset onto the local iCARE server (described below) and the Data Management site compares the pre- and post-upload statistics as a final QC step to confirm that the upload process did not affect dataset

integrity. Once all QC measures have been performed and no problems are detected, then the site's harmonized dataset is "certified" for analysis-readiness.

## Federated Dataset Approach to Multi-Site Analysis

Historically, analyzing datasets from disparate locations involved merging and transferring the separate datasets into a single master dataset. However, ethico-legal and data ownership concerns potentially hamper the pooling of datasets into a single resource, especially if the process involves export of data across national boundaries. Database federation techniques offer a viable solution to this problem by permitting transparent access to datasets located and managed in disparate locations without the need for data export for pooling or permanent archiving at a single location (Haas et al. 2002). Through database federation techniques, iCARE created a computational infrastructure with a secure, web-based, interface to facilitate analysis of the federated, harmonized, iCARE research datasets. The iCARE approach to federation is illustrated in Fig. 1 and briefly described below.

We provisioned a local iCARE database (LID) server (as either a physical computer server or virtual server running on a virtualization platform) at each data contributing site (Component A in Fig. 1). Currently there are 6 connected sites, with further expansion planned as new investigator sites become part of the collaboration. With each new version of the iCARE data dictionary, we create a database table based on the data dictionary on each LID to store the site's corresponding harmonized data. Each dataset version is stored in a separate database table to allow for maximum flexibility while maintaining data and result integrity. This allows an existing variable to be redefined or a new variable to be added in a subsequent data dictionary version, while maintaining the same data definition (and thus results) within each version.

The iCARE information technology (IT) site in Perth, Western Australia, houses the iCARE master server. The master server hosts two key components of the iCARE infrastructure, namely the Perl data federation component (Component B in Fig. 1), and the iCARE web-based analysis portal (iWAP; Component C in Fig. 1). The Perl data federation component is connected to each LID using an industry-standard secure connection method called Secure Shell (SSH) tunneling. The SSH tunneling creates a secure, encrypted communication channel over the Internet to transparently retrieve data from the harmonized datasets in each LID when an analysis is initiated. Hardware and software firewall rules are applied to the SSH tunnel to ensure these are uni-directional connections that can only be established from the master server to the LIDs. This extra security layer ensures that none of the LIDs can connect back to the master site, or any other LID. Furthermore, connections are dynamic and only active when a connection is made to one or more LIDs.

The iWAP (Component C in Fig. 1) forms the central access portal through which individual harmonized and federated datasets can be analyzed without users having to interact directly with the data. The system is secured with both client-side and server-side security certificates, in addition to strong password protection. To initiate an analysis session, an authorized iCARE investigator logs into the iWAP securely through their web browser and selects the datasets and variables they have been approved to analyse. The user also enters

the analysis syntax using one of the available statistical analysis packages (currently R, Stata or SAS). Once the user selects to run the analysis, the iWAP system initiates a federation process through the Perl data federation component. The federation component retrieves the data pertaining to the variables specified in the analysis query from each LID and virtually pools all data into computer memory. The virtually pooled data are passed directly to the specified analysis software package (also running on the master server). The user-provided statistical syntax is then executed on the data, and once the analysis task has finished running, the results are made available to the researcher in a project-specific file manager. At no stage is the virtual dataset saved to disk or stored in any permanent way, it is simply created as the temporary step between the analysis program and the analysis output. This federation technique allows the data to be transiently pooled only for the purposes of a given analysis, with data permanently stored only on the LID server at each remote site.

## Discussion

The iCARE collaboration has a number of strengths based on its use of long-standing, population-based data sources; rigorous data harmonization and quality control processes; the data federation approach; and a powerful analysis portal. These strengths, plus the multi-site approach, give it several analytical advantages. Its greatest scientific challenge is assurance of valid data.

### Analytical Advantages and Limitations

The use of longstanding population-based data systems in iCARE provides the capability to assess population-based time trends in both autism reported diagnoses and risk factors and to evaluate trends in characteristics of affected individuals over the life course (up to and after the age of diagnosis). The combination of time depth and use of unique personal identifiers at the individual sites permits the construction of family linkages (e.g., linking offspring through their parents or grandparents) with associated family history data regarding autism diagnoses and risk factors. The broad population coverage in the reported health data provided by the data systems also permits flexibility in selection of specific study populations (e.g., birth cohorts, multigenerational families, siblings) and comparison groups (e.g., individuals with certain conditions or health characteristics) extending the range of available analytical designs.

The iCARE multi-site approach creates an unprecedented opportunity for creation of large study sample sizes, enhancing the statistical power of analyses within individual strata of exposure or outcome, and enabling assessments of the independent effects of multiple risk factors. The multi-site approach also makes it possible to apply uniform analytical methods to pooled, multi-site data or individual, site-by-site, comparisons. These features may enhance the ability to detect etiological mechanisms and facilitate comparison of results and interpretation of data variability across different populations and health care systems. The multinational feature of the multisite approach also enhances our understanding of the generalizability of results within the global context.

Autism spectrum disorders (ASD) risk factors selected for initial investigations comprise a set of independent variables anticipated to benefit from the iCARE analytical advantages.

Variables include seasonality of conception or birth, inter-pregnancy interval, grandparental/parental age, grandparental/parental immigration, gestational age, birth weight, birth weight for gestational age, and maternal medication use during pregnancy. The data also enable us to carry out genetically informative analyses such as computation of the risk of ASD recurrence in siblings and estimation of the contribution of shared environmental factors to the liability to ASD among family members. iCARE partners are also examining temporal trends and geographic variation in the reported diagnoses of four conditions including ASD (ASD, hyperkinetic disorder, obsessive compulsive disorder, and Tourette's syndrome), as well as the impact of change in diagnostic criteria on the incidence of ASD reported diagnoses. As described previously, reported diagnoses and risk factor data are prospectively and continuously collected by population-wide health information systems that were established for administrative, health, socioeconomic, or demographic purposes, e.g., provision of universal health care. Reporting to the health information systems is carried out by designated persons with relevant expertise, e.g., midwives report to birth registries or attending medical specialists report to hospital discharge registries. These data collection systems contain individual-level information such as pre- and perinatal data, medical diagnoses, pharmacy data, death, immigration status, and social outcomes. The unique personal identification number permits linkage and merging of individual data within and across data systems at a site, e.g., birth and parent information can be used to link parents and children, thereby identifying families for use in studies of pregnancy history (e.g., inter-pregnancy interval), disease recurrence or heritability. In general, validation of registry data is usually undertaken for specific registry variables for a specific study. When such studies have been performed for a variable of interest to an iCARE analysis (e.g., validation of registry-reported ASD diagnoses), the study results will be referenced in iCARE manuscripts.

Both a limitation and strength of the iCARE resource includes the fact that all member sites are in economically developed countries and with relatively small, non-Caucasian minority populations. Thus, investigation of the impact of economic or ethnic diversity on trends or risk factor associations with ASD will be limited. A comparable level of economic development across sites, however, enhances the comparability of variables as they are more likely to reflect similar constructs. And minority populations from less developed countries in the iCARE countries provide an opportunity to investigate immigration effects in autism. Another limitation is that since 3 of the 6 sites rely all, or in part, on service/benefits registries for ASD case identification the data from these sites will be limited to cases that are registered for services or benefits. In the following section we discuss our approach to assessing the impact on case characteristics and associations with risk factors arising from site differences in case ascertainment and other sources of data variability.

### **Challenges in Data Assurance and Comparability**

Despite great care in construction, the iCARE harmonization and QC processes cannot resolve all issues of data validity and comparability across sites. As seen in Table 4, site differences in autism birth prevalence overall and across time make it evident that iCARE investigators must give careful consideration to the consequences of site differences in case ascertainment (e.g., registry-specific variation in ascertainment of different ASD diagnostic

or phenotypic subtypes), differences in registry reporting, and changes in diagnostic criteria across sites and over time, and their impact on case characteristics and associations with risk factors. Adoption of several approaches to assess site heterogeneity in harmonized ASD and risk factor data, and their potential impact on observed associations, will be key to addressing questions focused on data validity and comparability. To that end, pre-analytical, descriptive steps to assess data heterogeneity include exploration, variable by variable, of autism and risk factor differences over time, by site, and diagnostic system. These early descriptive steps are intended to identify variance in the data that will inform pre-analysis resolutions for handling such differences. In the analysis phase, several approaches to handling data variability may be implemented: (1) site-specific modeling to assess site heterogeneity in the association between ASD and the exposure of interest, and (2) sensitivity analyses wherein results using different site or variable inclusion criteria are compared to detect systematic differences and their impact on the results and interpretation of an analysis.

## Conclusion

iCARE is a unique resource for autism research. It offers a powerful answer to the call for multi-study data sharing to advance understanding of complex diseases through efficient leveraging of costly investments in data collection (Pisani et al. 2010; Walport and Brest 2011). The iCARE approach includes rigorous data harmonization processes and data federation that yields unprecedented autism study sample size and flexible research application capabilities. The infrastructure flexibility enhances the collaborative and analytical breadth among current iCARE sites and increases the long-term potential of the collaboration. iCARE solutions to concerns around data access security, patient confidentiality and data management address common challenges in multi-site collaboration, among which multinational collaborations pose a special case. In the end, however, the utility of the iCARE autism epidemiological approach will be reflected by the demonstrated research benefits of multi-site data sharing that is as rigorous as it is innovative.

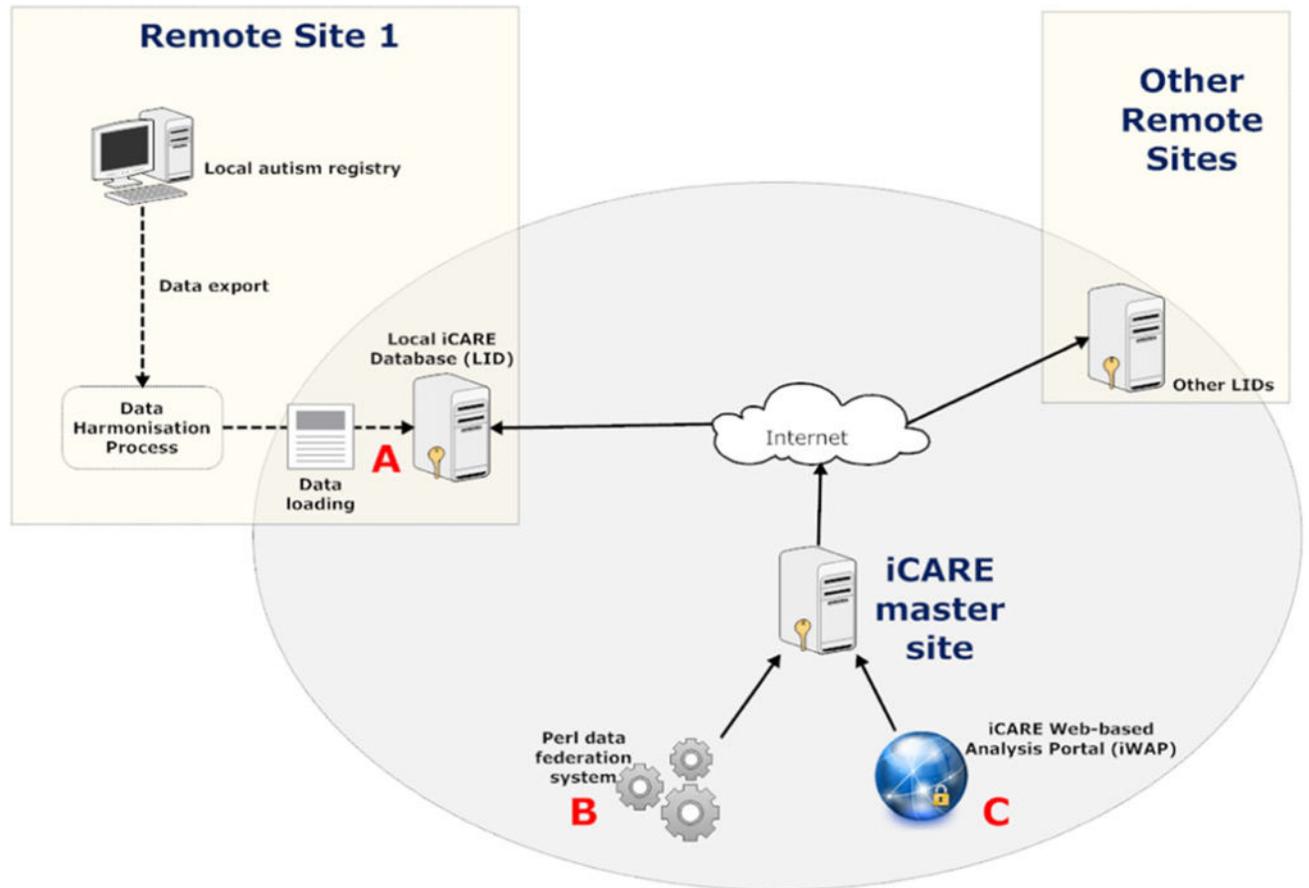
## Acknowledgments

We wish to acknowledge iCARE funding support from Autism Speaks grant numbers: 6230, 6246–6249, 6251, 6295. We also gratefully acknowledge the contributions to iCARE development and implementation of the Centers for Disease Control and Prevention, Manuel Posada MD, PhD, Anastasia Nyman Iliadou, PhD, JukkaHuttunen, PhD and TanjaSarlin, MSc.

## References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fourth edition, text revision. Washington, DC: American Psychiatric Association; 2000.
- Bettelheim, B. The empty fortress: Infantile autism and the birth of the self. New York: Free Press; 1972.
- Centers for Disease Control and Prevention. Surveillance summaries: Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, United States 2008. *MMWR Surveill Summ.* 2012; 61(SS03)
- Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. *Journal of Child Psychology and Psychiatry.* 1977; 18(4):297–321. [PubMed: 562353]

- Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatric Research*. 2009; 65:591–598. [PubMed: 19218885]
- Haas LM, Lin ET, Roth MA. Data integration through database federation. *IBM Systems Journal*. 2002; 41(4):578–596.
- Hallmayer J, Cleveland S, Torres A, et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry*. 2011 Published online July 4, 2011. 10.1001/archgenpsychiatry.2011.76
- Lampi KM, Banerjee PN, Gissler M, et al. Finnish prenatal study of autism and autism spectrum disorders. (FIPS-A): Overview and design. *Journal of Autism and Developmental Disorders*. 2010 Published online. 10.1007/s10803-010-1132-6
- Lauritsen MB, Jørgensen M, Madsen KM, Lemcke S, Toft S, Grove J, et al. Validity of childhood autism in the Danish psychiatric central register—findings from a cohort sample born 1990–1999. *Journal of Autism and Developmental Disorders*. 2010; 40:139–148. [PubMed: 19728067]
- Muilu J, Peltonen L, Litton JE. The federated database—basis for biobank-based post genome studies, integrating phenome and genome data from 600,000 twin pairs in Europe. *European Journal of Human Genetics*. 2007; 15:718–723. [PubMed: 17487219]
- Nassar N, Dixon G, Bourke J, Bower C, Glasson E, de Klerk N, et al. Autism spectrum disorders in young children: Effect of changes in diagnostic practices. *International Journal of Epidemiology*. 2009; 38(5):1245–1254. [PubMed: 19737795]
- Parner E, Schendel D, Thorsen P. Autism prevalence trends over time: The confounding of changes in age at diagnosis. *Journal of the American Academy of Child Psychiatry*. 2008; 162(12):1–8.
- Pisani E, Whitworth J, Zaba B, Abou-Zahr C. Time for fair trade in research data. *Lancet*. 2010; 375:703–705. [PubMed: 19913902]
- Ritvo ER, Jorde LB, Mason-Brothers A, Freeman BJ, Pingree C, Jones MB, et al. The UCLA-university of Utah epidemiologic survey of autism: recurrence risk estimates and genetic counselling. *The American Journal of Psychiatry*. 1989; 146:1032–1036. [PubMed: 2750975]
- Society for Cerebral Palsy in Europe. Surveillance of cerebral palsy in Europe: A collaboration of cerebral palsy surveys and registers. *Surveillance of Cerebral Palsy in Europe (SCPE)*. *Developmental Medicine and Child Neurology*. 2000; 42:816–824. [PubMed: 11132255]
- Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, et al. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of Child Psychology and Psychiatry*. 1989; 30(3):405–416. [PubMed: 2745591]
- Sullivan PF. The psychiatric GWAS consortium: Big science comes to psychiatry. *Neuron*. 2010; 68:182–186. [PubMed: 20955924]
- Walport M, Brest P. Sharing research data to improve public health. *Lancet*. 2011; 377:537–539. [PubMed: 21216456]



**Fig. 1.** Federated dataset approach to multi-site analysis in the International Collaboration for Autism Registry Epidemiology (iCARE). A local iCARE database (LID) physical/virtual server is provisioned at each site running a MySQL database containing harmonized research data (*Component A*). The iCARE master site houses the Perl data federation system (*Component B*) and the iCARE Web-based Analysis Portal (iWAP; *Component C*). The former creates secure connections to and controls retrieval of the data on each LID and the latter hosts the web interface for project management and data analysis. Arrows indicate the uni-directional connections between entities in the infrastructure

**Table 1**

## Member sites and roles in the International Collaboration for Autism Registry Epidemiology (iCARE)

<b>Role</b>	<b>Description</b>	<b>Site</b>
Data contributor	Acquire data from local data systems and maintain local harmonized dataset for collaborative analyses Obtain all necessary local approvals Manage and prepare local data according to iCARE protocol	Denmark, Finland, Israel, Norway, Sweden, Western Australia
IT operations	Establish and maintain the necessary hardware/software for centralized web-based data access	Western Australia
Data management	Prepare standardized data harmonization protocols and provide oversight of harmonized dataset preparation and quality control	USA/Columbia university
Project lead	Provide general coordination and oversight of consortium and responsible for general iCARE communication	Denmark
Founding collaborator	Participate in and support iCARE funding proposal development and contribute to iCARE activities in supportive roles as needed	Spain
Analysis	Specific site designated as lead for an analysis: Conduct the analysis Lead discussion on results and interpretation Prepare draft manuscripts Conduct analysis-specific communication	All sites are eligible
Scientific oversight	Review and approve proposed analyses (as member of Analytical Oversight Committee)	All sites participate with designated representatives
Scientific input and oversight	Provide input to analysis and manuscript preparation (as member of a specific analysis group)	All sites participate

**Table 2** Characteristics of the data contributing sites in the International Collaboration for Autism Registry Epidemiology (iCARE)

Site	Denmark	Finland	Israel	Norway	Sweden	Western Australia (WA) <sup>d</sup>
Catchment area	Nation-wide	Nation-wide	Nation-wide	Nation-wide	Nation-wide	State-wide
Average births/year	62,000	60,000	86,000	61,000	107,000	24,000
Country of birth or ethnic profile	90 % Danish	95 % born in Finland 3 % born else-where in Europe 2 % born outside Europe	98 % born in Israel 2 % born elsewhere	88 % Norwegian 12 % other	> 90 % born in Sweden < 10 % born elsewhere	93 % Caucasian 7 % other
Health care provision	Public	Public	Public	Public	Public	Public and private
Type of study population	Birth cohort	Birth cohort	Birth cohort	Birth cohort	Birth cohort	Birth cohort
Data source of study population	Danish Medical Birth Registry	Finnish Medical Birth Registry	Israel Birth Register	Medical Birth Registry of Norway	Swedish Patient Registry	Midwives' Notification System
<i>Characteristics of data source of autism spectrum disorders (ASDs) diagnosis for individuals in iCARE birth cohort study populations</i>						
Name	Danish Psychiatric Central Registry	Finnish Hospital Discharge Registry	Israeli Ministry of Social Affairs	Norwegian National Insurance Scheme (Norwegian Patient Register to be used when available)	Swedish Hospital Discharge Register	1 Disability Services Commission of WA 2 WA Register for ASD
Type	Medical discharge registry	Medical discharge registry	Service registry	Benefit registry	Medical discharge registry	1 Service registry 2 Diagnosis registry
Steward	Government agency	Government agency	Government agency	Government agency	Government agency	1 Government agency 2 Telethon Institute for Child Health Research
Diagnostic system for iCARE cohorts	ICD-8: 1987-1993 ICD-10: 1994-present	ICD-9: 1987-1995 ICD-10: 1996-present	DSM-IV	ICD-9: 1987-1992 ICD-10: 1992-present	ICD-9: 1987-96 ICD-10: 1997-present	DSM-IV (individuals diagnosed under earlier diagnostic systems were reclassified to DSM IV by the data sources)
Types of contact reported to source	Inpatient: 1987-present Outpatient: 1995-present Does not include contacts with general practitioners (the primary care providers)	Inpatient: 1987-present Outpatient: 1998-present Does not include contacts within primary health care (before 2011) and private outpatient visits outside hospitals contacts	All registrations for services; likely includes more severely impaired individuals with ASD not receiving benefits are not recorded	All registrations for benefits Individuals with ASD not receiving benefits are not recorded.	Inpatient: 1987-present Outpatient: 2000-present Does not include contacts with general practitioners (the primary care providers)	All registrations for services

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

*ICD* international classification of diseases, *DSM* diagnostic and statistical manual (of Mental Disorders)

Western Australia ASD data sources supplemented by a clinical cohort of ASD cases born 1983–1995 and diagnosed by 1999 (Nassar et al. 2009)

Table 3

Referral, evaluation and reporting of autism spectrum diagnoses to data sources at sites in the International Collaboration for Autism Registry Epidemiology (iCARE)

If a child is suspected of having a developmental problem						
Across-site summary	Step 1: Referrals are made by health or education providers or by parent referral	Step 2: Child is evaluated at a specialist care unit in child psychiatry, child development, or child assessment	Step 3: The evaluation is performed by a multi-disciplinary team or single specialist	Step 4: At 5 sites, the diagnosis is reported by the specialist to the data source. In Norway, the diagnosis is recorded in the data source based on parent application for benefits; an accompanying specialist report of the evaluation is mandatory		
Site-specific details	Denmark	Finland	Israel	Norway	Sweden	Western Australia (WA)
Step 1: Who can refer?	General practitioner, pediatrician, child psychologist	General practitioner, pediatrician	Public well-child care clinics	Primary health care providers, child psychologist	General practitioner, pediatrician, child psychologist	Health care providers, parent request via health care provider
Step 2: Where do they go?	Child psychiatric clinic	Specialist care facility at a central or university hospital	Child Development Centers	Child psychiatry or pediatric clinics	Child psychiatric clinic	Government-based assessment agency or a non-government private practice
Step 3: Who performs the evaluation?	Typically, multidisciplinary assessment led by child psychiatrist	Multidisciplinary assessment led by child neurologist or psychiatrist	Typically, multidisciplinary assessment led by child psychiatrist or child neurologist with expertise in child development	Child psychiatrist, pediatrician, or specialist in clinical psychology	Multidisciplinary assessment	Standardized procedure for multidisciplinary evaluation (pediatrician or psychiatrist, psychologist and speech pathologist) for ASD established in early 1990's
Step 4: Who reports the diagnosis to the data source?	Only child psychiatrist	Only health care personnel: attending health care specialist	Only board-certified specialists in child and adolescent psychiatry or pediatric neurology approved by the Ministry of Health and Ministry of Welfare	Registration occurs when application for benefits is approved and benefits received; an accompanying medical specialist report is required. Staff at evaluation site assists in process	Only attending health care specialist	Consensus diagnosis of evaluating multidisciplinary team
Validation studies of reported ASD diagnosis	Expert review of behavioral information in medical records to determine if number and type of behaviors	Children with <i>childhood autism</i> re-assessed using the Autism Diagnostic Interview-Revised (ADI-R)	Children with <i>childhood autism</i> re-assessed using the Autism Diagnostic Interview-Revised (ADI-R) +	None for the National Insurance Scheme Validation study of	Expert review of behavioral and diagnostic information in medical records for a random sample of cases.	None

**If a child is suspected of having a developmental problem**

**Across-site summary**

Step 1: Referrals are made by health or education providers or by parent referral

Step 2: Child is evaluated at a specialist care unit in child psychiatry, child development, or child assessment

Step 3: The evaluation is performed by a multi-disciplinary team or single specialist

Step 4: At 5 sites, the diagnosis is reported by the specialist to the data source. In Norway, the diagnosis is recorded in the data source based on parent application for benefits; an accompanying specialist report of the evaluation is mandatory

Site-specific details	Denmark	Finland	Israel	Norway	Sweden	Western Australia (WA)
Sources of variation in ASD reporting within sites	<p>One main hospital didn't report ASD to the data source until 1992</p> <p>Addition of out-patient, as well as in-patient, contacts in 1990s</p> <p>Changes in ICD diagnostic system versions</p>	<p>96 % met criteria in all three ADI-R domains and onset of symptoms before age 3 years (80 assessed cases)</p> <p>Lauritsen et al. 2010</p>	<p>Expert review of medical records to confirm reported diagnosis of all registered cases</p> <p>ongoing</p>	<p>children in the Autism Birth Cohort with an ASD diagnosis in the Norwegian Patient Registry. Validation diagnosis comprised of clinician best estimate based on all information (clinical evaluation, parent/teacher interviews).</p> <p>97 % met DSM IV criteria for ASD (of 60 assessed cases to date)</p> <p>ongoing</p>	<p>94 % met criteria for autistic disorder based on DSM IV (of 88 cases assessed to date).</p> <p>Ongoing</p>	<p>Waiting lists for referral can be lengthy, especially for school-aged children on government agency waiting lists – this can delay registration to older age</p> <p>Services vary by age and diagnosis; fewer benefits after 5 years of age</p> <p>more benefits for autistic disorder</p> <p>Intentional trend for clinicians to diagnosis autistic disorder before Asperger syndrome</p>

**If a child is suspected of having a developmental problem**

**Across-site summary**

Step 1: Referrals are made by health or education providers or by parent referral

Step 2: Child is evaluated at a specialist care unit in child psychiatry, child development, or child assessment

Step 3: The evaluation is performed by a multi-disciplinary team or single specialist

Step 4: At 5 sites, the diagnosis is reported by the specialist to the data source. In Norway, the diagnosis is recorded in the data source based on parent application for benefits; an accompanying specialist report of the evaluation is mandatory

**Site-specific details**

Denmark

Finland

Israel

Norway

Sweden

Western Australia (WA)

Changes in ICD diagnostic system versions  
 Changes in ICD diagnostic system versions  
 Changes in ICD diagnostic system versions  
 Changes in ICD diagnostic system versions

Table 4

Sample size, prevalence (per 1,000), and sex ratio of individuals with autism spectrum disorders (ASDs) in each birth cohort in the International Collaboration for Autism Registry Epidemiology (iCARE)

Final year of follow-up period	Denmark		Finland		Israel <sup>a</sup>		Norway		Sweden		Western Australia <sup>b</sup>	
	N	Prevalence	N	Prevalence	N	Prevalence	N	Prevalence	N	Prevalence	N	Prevalence
<i>1987–1996 births</i>												
All ASD	6,120	9.5	4,712	7.4	438	1.4	985	1.7	10,289	9.2	770	3.3
Autistic disorder	1,277	2.0	879	1.4	Not available		455	0.8	3,214	2.9	531	2.3
		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)
All ASD		3.7:1		3.6:1		5.0:1		3.9:1		2.2:1		5.9:1
Autistic disorder		4.0:1		3.6:1		Not available		3.1:1		2.6:1		5.9:1
<i>1997–2004 births</i>												
All ASD	4,218	8.0	2,676	5.9	1,653	2.2	299	0.6	4,212	5.8	364	5.2
Autistic disorder	1,633	3.1	561	1.2	Not available		197	0.4	2,172	3.0	279	4.0
		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)		Sex ratio (M:F)
All ASD		4.7:1		3.4:1		4.4:1		4.9:1		3.4:1		4.9:1
Autistic disorder		4.8:1		3.3:1		Not available		5.4:1		3.5:1		4.6:1

<sup>a</sup> 1993–2004 births. Jewish population only included in iCARE dataset due to suspected under-ascertainment of ASD in non-Jewish population. Service registry likely includes more severely impaired individuals and a high proportion of individuals with autistic disorder

<sup>b</sup> Non-Aboriginal population only included in iCARE dataset due to suspected under-ascertainment of ASD in Aboriginal population. Benefits structure in health care system may contribute to a disproportionate number of autistic disorder diagnoses reported than other ASD diagnoses

**Table 5**

Harmonization of autism spectrum disorder (ASD) diagnoses across different diagnostic systems

Diagnostic System and Associated Codes	ICD-8 299.00/01/02/03 (Psychosis; used in Denmark to indicate autism)	ICD-9 299.0 (Infantile Autism) 299.1 (Disintegrative psychosis) 299.8 (Other) (note: should include Asperger syndrome and Other PDD) 299.9 (unspecified) (note: should include PDD-NOS).	ICD-10 F84.0 (Childhood Autism) F84.1× (Atypical Autism) F84.5 (Asperger Syndrome) F84.8 (Other PDD) F84.9 (PDD-NOS) F84.2 (Rett Syndrome)	DSM-IV 299.0 (Autistic Disorder) 299.1 (Childhood disintegrative disorder (CDD)) 299.8 (Rett syndrome) 299.8 (Asperger syndrome) 299.8 (PDD-NOS)
<i>iCARE ASD categories</i>				
Autism (infantile autism, autistic disorder (AD))	299.00	299.0	F84.0	299.0
Asperger syndrome (ASP)	299.02	299.8	F84.5	299.8
Pervasive developmental disorder—not otherwise specified (PDD-NOS) (or, other ASDs and not Autism and not Asperger Syndrome)	299.01, 299.03	299.9	F84.9 F84.1x, F84.8	299.8

When multiple diagnoses are available for a given individual

if ever Rett Syndrome or CDD, then classify as not ASD, regardless if also ever had an ASD diagnosis

if never Rett Syndrome or CDD, then classify as

AD: Autistic disorder/childhood autism if ever received this diagnosis (i.e., disregard other ASD subtype diagnoses)

ASP: If never autistic disorder/childhood autism AND ever had Asperger syndrome (i.e., disregard other ASD subtypes)

PDD-NOS: if never AD and never ASP and ever (PDD-NOS OR ATYPICAL AUTISM OR OTHER PDD)

If a person has both ICD9 and ICD 10 diagnosis codes (for final subtype) then use the ICD 10 code for the subtype diagnosis