Research article

Validation of the diagnosis of autism in general practitioner records

Eric Fombonne*[1,2], Lisa Heavey[2], Liam Smeeth[3], Laura C Rodrigues[3], Claire Cook[3], Peter G Smith[3], Linyan Meng[1] and Andrew J Hall[3]

Address: [1]McGill University, Montreal Children's Hospital, 4018 St. Catherine St. West, Montreal, QC, Canada, H3Z 1P2, [2]Institute of Psychiatry, 16 De Crespigny Park, Denmark Hill, London, SE5 8AF, UK and [3]London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Email: Eric Fombonne* - eric.fombonne@mcgill.ca; Lisa Heavey - l.reidy@shu.ac.uk; Liam Smeeth - liam.smeeth@lshtm.ac.uk; Laura C Rodrigues - laura.rodrigues@lshtm.ac.uk; Claire Cook - claire.cook@lshtm.ac.uk; Peter G Smith - peter.smith@lshtm.ac.uk; Linyan Meng - linyan.meng@muhc.mcgill.ca; Andrew J Hall - andy.hall@lshtm.ac.uk

* Corresponding author

Abstract

Background: We report on the validity of the computerized diagnoses of autism in a large case-control study investigating the possible association between autism and the measles, mumps and rubella vaccine in the UK using the General Practitioner Research Database (GPRD). We examined anonymized copies of all relevant available clinical reports, including general practitioners' (GP) notes, consultant, speech therapy and educational psychologists reports, on 318 subjects born between 1973 and 1997 with a diagnosis of autism or a related disorder recorded in their electronic general practice record.

Methods: Data were abstracted to a case validation form allowing for the identification of developmental symptoms relevant to the diagnosis of pervasive developmental disorders (PDDs). Information on other background clinical and familial features was also abstracted. A subset of 50 notes was coded independently by 2 raters to derive reliability estimates for key clinical characteristics.

Results: For 294 subjects (92.5%) the diagnosis of PDD was confirmed after review of the records. Of these, 180 subjects (61.2%) fulfilled criteria for autistic disorder. The mean age at first recording of a PDD diagnosis in the GPRD database was 6.3 years (SD = 4.6). Consistent with previous estimates, the proportion of subjects experiencing regression in the course of their development was 19%. Inter-rater reliability for the presence of a PDD diagnosis was good (kappa = .73), and agreement on clinical features such as regression, age of parental recognition of first symptoms, language delay and presence of epilepsy was also good (kappas ranging from .56 to 1.0).

Conclusions: This study provides evidence that the positive predictive value of a diagnosis of autism recorded in the GPRD is high.

Background

Of 32 epidemiological surveys of autism and pervasive development disorders (PDDs) included in a recent review [1], 13 were published within the last 5 years. Increased research activity in this field of neuropsychiatry has led to a refinement of the definition of autism that involves a combination of qualitative impairments in language/communication, in social interaction and in patterns of play behaviours and interests. Improved operationalisation of diagnostic definitions within nosographies has occurred (American Psychiatric Association, 1994; World Health Organization, 1992), in parallel with the development of more precise diagnostic instruments such as the Autism Diagnostic Interview [2] and the Autism Diagnostic Observational Schedule [3]. There has also been an increasing concern by the public about this group of disorders, prompted, in part, by concern that the rates of PDD may have increased in recent decades [4-6]; and that the cause of the increase may be due to the side effects of vaccination [7-9] or increased exposure of young infants to neurotoxins such as methylmercury or thimerosal [10].

The investigation of risk factors for autism in epidemiological surveys has been limited by the small size of many studies. The median number of cases identified in the 32 surveys reviewed was 50 children [1]. Some investigators examined the effects of specific environmental exposures using large samples of subjects obtained from educational or hospital services [11-13]. These included computerized databases obtained through research networks of general practitioners, such as the General Practitioner Research Database (GPRD) [14], the Doctor Independent Network Database in the UK [15], national registers [16] and memberships of consumer associations [17]. There are problems with the use of these databases because of uncertainties about the validity of the diagnosis of the cases. Often there is no information on specific clinical characteristics which might allow identification of subgroups of individuals within the same diagnostic category (e.g. children with autism who have regressed in the course of their development), precluding investigation of hypotheses proposing an association between subtypes of PDDs and specific exposures [18].

To test the hypotheses of a link between autism and exposure to combined measles, mumps and rubella vaccines, or to other infectious agents, we designed a study based on cases identified in the GPRD database in the UK [19,20]. The research protocol included evaluating the quality of diagnoses in the GPRD database by obtaining clinical reports on a sub-sample of children included in the study. We report the results of this validation study on a subset of 318 cases of autism, based on the reports in the medical files of the general practitioners through whom these cases were identified.

Methods

The General Practice Research Database

The GPRD (previously called the VAMP (Value Added Medical Products) Research Bank) was set up in 1987 and is now held on behalf of the Department of Health by the Medicines Control Agency [21]. It consists of the computerised general practice medical records for around 3 million people in the United Kingdom. The electronic record includes demographic information such as age and sex, details of every consultation with a general practitioner, all prescribed drugs and vaccinations given, and details of referrals to hospital or specialist services. It is possible to obtain from general practitioners copies of hospital letters regarding specific patients (in anonymised form), although not all participating general practices provide this service.

Practices in the GPRD originally used a modification of the Oxford Medical Information Systems coding system to record diagnoses [22]. Through the 1990s an increasing number of practices changed to the READ coding system, which is now used throughout the United Kingdom National Health Service [23].

Selection of cases and data obtained

Of patients with a recorded diagnosis of PDD in the GPRD, including prevalent cases when first registered, 446 were registered with 203 general practices willing to provide copies of patient records. For 80 of these patients, records were not available as the patient was no longer registered with the general practitioner. We obtained complete case records including copies of hospital clinic letters and specialist reports for 318 (87%) of the remaining 366 patients.

Abstraction of clinical data

The case validation form included a section on socio-demographic data, an assessment of current level of language and of educational status, together with estimates of associated levels of learning disabilities. The assessment of learning disabilities was based on available psychometric data and, when not available, on a best estimate of intellectual functioning, classified into broad bands after review of all the available information. A section on health covered the lifetime occurrence of epilepsy, of treatments with psychoactive drugs, of associated medical disorders, the body measurements of head circumference, weight and height, of dysmorphic syndromes and the report of any significant non-autistic symptom in the course of the development of the child (such as: gastrointestinal symptoms, infections, sleeping difficulties or immune medical conditions). Symptoms were rated as

being either reported or not reported since information allowing a more detailed coding (based on severity of the symptom, age of onset and of offset) was not available for most cases where symptoms were reported. A section on pregnancy and birth covered the incidence of maternal illness and infection during pregnancy, the mode of delivery, length of labour, birth weight and birth order. A section on the early development of the child covered major milestones (coded as normal versus delayed), the age of first words and phrases (coded either as an age in months or as an approximate age band), language delay (defined as single words not occurring until after 24 months of age or phrase speech not occurring until after 36 months of age), any regression or loss of skill at any point in the course of development and, when present, the type of skill lost. For those cases with some regression/loss of skills, a global judgment was provided by the rater on whether the developmental pattern was suggestive of a definite regression/loss of acquired skills as opposed to fluctuating development with an uneven rate of skill acquisition. As age on onset of first symptoms is a key diagnostic criterion, this was operationalised in three different ways. First, we recorded the age at which the parents first recognized signs of developmental delay or variation in their child and the type of symptoms that first triggered their concerns. Second, we recorded age at the date of the first letter on file with concerns about a developmental problem in the child (e.g. a referral letter from the GP to a specialist to gain an opinion). Third, rater's assessment of age of onset was based on the rater's judgment of age of first symptom onset, irrespective of actual parental and or professional recognition of these symptoms at the time. Specific or global developmental disorders were identified in the first and second-degree relatives, along with specific medical (especially autoimmune) and psychiatric disorders.

The overall diagnostic rating of a child was made with two approaches, one algorithmic and one based on judgment of the rater. First, reports were searched for evidence of 12 specific DSM-IV symptoms for PDDs. A computer diagnostic algorithm using DSM-IV symptoms ratings was devised. Instead of using the typical DSM-IV algorithm (2 social symptoms, 1 communication/language symptoms, 1 repetitive behaviour symptom, together with at least 6 symptoms out of the 12 possible), we generated an algorithm to take account of the uneven quality of the data across the subjects. The algorithm generated a PDD diagnosis when at least three out of the 12 DSM-IV symptoms were scored, with the further constraint that there would be at least 1 symptom in the social domain and 1 symptom in either the communication/language or the repetitive behaviour domain. This algorithm is consistent with that used in a recent survey using a comparable record review approach [5]. Second, when all the documentation

had been reviewed, the rater was asked to make a global judgement regarding the presence or absence of a PDD in the child and, when present, to provide a diagnosis for the specific subtype of PDD whenever possible. A global index of confidence in the rater's judgment about the PDD diagnosis was also derived.

Raters and inter-rater reliability
The two raters were a child psychiatrist (EF) and a psychologist (LH) both with long experience in the field of autism. LH reviewed and coded all the files. In order to obtain reliability estimates on the rating procedure, a subset of 50 medical notes chosen at random amongst the 318 records was rated blindly by EF. Records that posed particular coding difficulties were identified and consensus ratings were derived by the two raters at the end of the study. These records were eligible to be selected for the inter-rater reliability study. All ratings were made blind to the child's history of immunization.

Statistical analysis
All data were analyzed with SPSS and SAS with conventional chi-square and Fisher exact tests for categorical variables and Student's t test for continuous variables. Inter-rater reliability was measured with the kappa coefficient for categorical ratings and with the intraclass correlation coefficient (ICC) for continuous measures [24]. Throughout, a p value of .05 was chosen as the level of statistical significance. Missing data occurred at high rates for many variables included in the case validation forms and, as a result, we report both absolute and relative frequencies.

Results
Sample characteristics
Medical notes for 318 subjects were obtained. They varied in quality and exhaustiveness. For some children, GP records included several consultant reports, speech and language assessments, and educational psychology reports. For other children, the information available was scanty, with sometimes the only available data consisting of one, or a few, letters between the GPs and consultants. A high proportion of records had missing data on parental age, socio-economic status, and detailed psychometric assessment of the child and therefore the frequencies of these variables are not described here. Of the 318 children whose medical forms were obtained, the raters confirmed a diagnosis of PDD in 294 children (92.5%). Compared to children with a confirmed PDD diagnosis, children for whom the diagnosis was not confirmed (n = 24) had significantly fewer PDD symptoms (2.1 vs 6.2; p < .001), higher language level (phrase speech: 80% vs 45%; p=.051), and more frequent parental concern arising for the first time after the age of 3 years (20% vs 2.9%; p=.024). No significant differences were found with respect to gender, birth year, presence of epilepsy or

Table 1: Characteristics of 294 Children with PDD

		N	%
Male		238	81.0
Diagnosis subtype	Autism	180	61.2
	PDD NOS	19	6.5
	Asperger	18	6.1
	Unknown	77	26.2
Birth year	1973 – 1977	19	6.5
	1978 – 1982	31	10.7
	1983 – 1987	65	22.3
	1988 – 1992	115	39.5
	1993 – 1997	61	21.0
Past or current epilepsy		53	18.0
Overall level of language[1]	Mute	45	16.9
	Single words	52	19.5
	Few phrases	50	18.7
	Phrase speech	120	44.9
Estimated intellectual functioning	Normal IQ	89	34.5
	Mild MR	99	38.4
	Moderate MR	49	19.0
	Severe MR	21	8.1
Age of parental recognition (months)	0 – 12	82	39.6
	13 – 24	94	45.4
	25 – 36	25	12.1
	36+	6	2.9
Regression		55	19.0
	N	Mean	SD
Age at first GPRD diagnosis (years)	278	6.25	4.52
Age at first letter (years)	282	4.50	3.84
Age at last report on note (years)	289	10.66	5.65
Mean number of DSM-IV PDD symptoms			
–Total	294	6.23	4.17
–Language/communication	294	1.66	1.43
–Social interaction	294	2.53	1.82
–Repetitive behaviours	294	2.05	1.76

[1]: Mean age when language level was recorded was: 7.86 years (SD: 4.50).

regression or in the average age at first diagnosis in the GPRD database.

The main characteristics of the 294 children with a confirmed diagnosis of PDD are shown in Table 1. The male/female ratio was 4.25:1. A third of the children had no phrase speech when language level was recorded (at a mean age of 7.9 years). About a third of children had estimated intellectual skills falling into the normal range. 55 (19%) children showed clear evidence of regression and loss of acquired skills, and a further 34 had a developmental pattern consistent with an uneven and slow rate of acquisition of new skills as they grew up. The rate of regression and of epilepsy (18%) are consistent with those described in other surveys of autistic children. The mean number and pattern of DSM-IV symptoms was consistent with the diagnostic concepts of autism, especially as

symptoms of social deficits appeared to be reported more frequently (Table 1). The computer-based algorithm identified 237 (80.6%) of the 294 cases as having a PDD.

It was possible to allocate a more specific diagnosis to 217 of the 294 children with PDD. This was autistic disorder in 180 children (82.9%), Asperger Disorder (AD) in 18 children (8.3%), and PDDNOS (Pervasive Developmental Disorder Not Otherwise Specified) in 19 children (8.8%). The confidence level in the diagnostic subtype was generally high (high in 67.1%; medium in 19.7%, and low in 13.1%). In the remaining 77 children (26.2%), the quality of the data did not allow for the diagnosis of a specific PDD subtype. A comparison of the PDD children with and without a more specific diagnosis showed that children without a PDD subtype were comparable to children with autism with respect to language level and intel-

Table 2: Comparison of Children with Autism or other PDD (N = 217)

		Autism		PDDNOS & Asperger		P
		(N = 180)		(N = 37)		
		N	%	N	%	
Male Sex		146	81.1	31	83.8	.703
Birth year	1973 – 1977	11	6.2	-	-	.217
	1978 – 1982	15	8.4	3	8.3	
	1983 – 1987	43	24.2	11	30.6	
	1988 – 1992	68	38.2	18	50.0	
	1993 – 1997	41	23.0	4	11.1	
Past or current epilepsy		29	16.1	5	13.5	.692
Overall level of language[1]	Mute	31	17.7	-	-	.000
	Single words	38	21.7	1	2.9	
	Few phrases	39	22.3	5	14.7	
	Phrase speech	67	38.3	28	82.4	
Estimated intellectual Functioning	Normal IQ	42	25.3	27	81.8	.000
	Mild MR	78	47.0	4	12.1	
	Moderate MR	34	20.5	2	6.1	
	Severe MR	12	7.2	-	-	
Age of parental recognition (months)	0 – 12	60	41.1	8	32.0	.176
	13 – 24	68	46.6	11	44.0	
	25 – 36	16	11.0	4	16.0	
	36+	2	1.4	2	8.0	
Regression[2]		43	24.0	3	8.3	.036
Language delay	Yes probable/definite	162	96.4	13	44.8	.000

	Autism			PDD NOS & Asperger			P
	N = 180			N = 37			
	N	Mean	SD	N	Mean	SD	
Age at first GPRD diagnosis (years)	167	5.45	4.06	36	7.77	3.68	.002
Age at first letter (years)	174	3.45	2.48	36	5.08	3.39	.009
Age at last report on note (years)	176	10.50	5.61	36	10.51	4.28	.992
Mean number of DSM-IV PDD symptoms							
–Total	180	7.67	3.92	37	7.24	3.81	.543
–Language/communication	180	2.14	1.41	37	1.62	1.28	.040
–Social interaction	180	3.04	1.76	37	3.24	1.57	.525
–Repetitive behaviours	180	2.49	1.73	37	2.38	2.06	.732

[1]: Mean age when language was recorded was: 7.77 years (SD: 4.31). [2]: 9 (12%) out of 77 children in unknown diagnostic group had regression

lectual functioning but closer to the children with either PDDNOS or AD with respect to age at first electronic diagnosis and rate of regression. Compared to both other groups, they had significantly fewer PDD symptoms, most certainly reflecting the poorest quality of the notes that precisely precluded a final sub-typing to be attained by raters.

The mean age at first parental concern regarding their child's development was 16.8 months (SD = 9.8) in 142 children where a precise age could be estimated. Age of first recognition of symptoms in medical records could be estimated in broad age bands in 207 subjects and occurred before age 3 years in 201 subjects (97.1%). Onset of first symptoms was also determined by the rater's

judgment, based on the medical records, in 91 subjects and was 12 months (SD = 8.5). In 88 subjects where both a parental and a rater age of onset were available, the rater mean age of onset was significantly younger than the age at parental concern (12.1 months vs 13.3 months; paired t-test; p = 0.02). Finally, the presence of a PDD in a first-degree relative of the index child was reported in 7.8% of the sample, consistent with other surveys of PDDs [4].

We compared children with an autistic disorder diagnosis with children with another PDD diagnosis (Table 2). The PDDNOS/AD group had significantly fewer language and intellectual impairments and were on average 2.3 years older than their autistic counterparts when recorded in the

Table 3: Autistic Children with or without Regression (N = 179¹)

		Regression				
		No (N = 136)		Yes (N = 43)		P
		N	%	N	%	
Male Sex		114	83.8	32	74.4	.166
Birth year	1973 – 1977	8	5.9	3	7.0	.789(trend)
	1978 – 1982	12	8.9	3	7.0	
	1983 – 1987	31	23.0	12	27.9	
	1988 – 1992	56	41.5	12	27.9	
	1993 – 1997	28	20.7	13	30.2	
Past or current epilepsy		21	15.4	7	16.3	.895
Overall level of language²	Mute	21	15.9	10	23.3	.065 (trend)
	Single words	27	20.5	11	25.6	
	Few phrases	28	21.2	11	25.6	
	Phrase speech	56	42.4	11	25.6	
Estimated intellectual functioning	Normal IQ	35	27.8	7	17.5	.054 (trend)
	Mild MR	61	48.4	17	42.5	
	Moderate MR	22	17.5	12	30.0	
	Severe MR	8	6.3	4	10.0	
Age of parental recognition (months)	0 – 12	45	43.3	15	35.7	.897 (trend)
	13 – 24	45	43.3	23	54.8	
	25 – 36	12	11.5	4	9.5	
	36+	2	1.9	-	-	
Language delay	Yes probable/definite	122	97.6	40	93.0	.163

		Regression					
		No (N = 136)			Yes (N = 43)		P
		N	Mean	SD	N	Mean	SD	
Age at first GPRD diagnosis (years)	127	5.39	4.01	40	5.63	4.29	.746
Age at first letter (years)	133	3.66	2.75	41	2.76	1.02	.002
Age at last report on note (years)	134	10.61	5.65	42	10.15	5.56	.646
Mean number of DSM-IV PDD symptoms							
–Total	136	7.51	3.94	43	8.35	3.67	.221
–Language/communication	136	2.08	1.43	43	2.37	1.29	.236
–Social interaction	136	3.04	1.66	43	3.14	2.02	.738
–Repetitive behaviours	136	2.40	1.73	43	2.84	1.68	.144

¹: Autism children only. One child had no information about regression. ²: Mean age when language was recorded was: 7.65 years (SD: 4.39).

GPRD database. Regression was less often reported in the PDDNOS/AD group.

Since regression and loss of skills is a clinical feature of potential interest for our main study, we examined further the clinical correlates of regression. As regression was infrequent in the PDDNOS/AD group and as these children were different from children with autism with respect to age at diagnosis and severity, we restricted this analysis to those children with a diagnosis of autistic disorder (Table 3). The regressive and non regressive groups differed with respect to language level and intellectual functioning where children with regression exhibited lower levels of functioning at the final assessment. They also had a significantly lower age at the first referral letter on file mentioning a developmental problem.

Trends over time in clinical features that are known to indicate autism severity were also examined for the autism group (Table 4). Birth years were grouped into 5-year intervals. There was a significant trend for decreasing levels of mental retardation and for an increasing proportion of males, suggesting that clinical presentation became less severe over time. Age differences in the birth cohorts made the interpretation of trend for phrase speech and for epilepsy more difficult to interpret. Other clinical features, including regression, did not change significantly with time.

Table 4: Characteristics of 178[1] Children with Autism over five year intervals

| | \multicolumn{10}{c}{Age interval} | P (for trend) |
| | 73 – 77 | | 78 – 82 | | 83 – 87 | | 88 – 92 | | 93 – 97 | | |
	N	%	N	%	N	%	N	%	N	%	
Male Sex	8	72.7	12	80.0	32	74.4	55	80.9	38	92.7	.058
Past or current epilepsy[2]	4	36.4	4	26.7	5	11.6	11	16.2	4	9.8	.046
Phrase speech	3	27.3	10	66.7	19	47.5	24	35.8	10	24.4	.044
Normal/mild moderate Functioning	4	36.4	9	60.0	30	75.0	45	71.4	31	86.1	.003
Age of parental recognition <24 months	8	80.0	9	81.8	30	85.7	47	88.7	34	91.9	.198
Regression	3	27.3	3	20.0	12	27.9	12	17.6	13	31.7	.789
Probable/definite Language delay[3]	9	81.8	11	91.7	41	100.0	62	96.9	38	97.4	.069
DSM-IV Symptoms:											
–Mean	6.91		6.00		8.16		7.62		8.29		.127
–SD	4.13		3.70		4.48		3.19		4.24		

[1]: Two children had no information about date of birth. [2]: Combines definite and possible cases. [3]: Mean age when language was recorded was: 7.65 years (SD: 4.39).

Interrater reliability

Interrater reliability was examined on the subset of 50 randomly selected children. Agreement between the two raters was good for the presence/absence of a PDD in the child (Kappa = .73), and there were only 2 cases where raters originally disagreed. The agreement on the number of DSM-IV symptoms was excellent (ICC = .92). PDD symptom scores for each of the three domains separately showed high intra-class correlations as well, with ICC values of .87 for the social domain, .75 for the communication/language domain, and of .91 for repetitive behaviours. The agreement was also good to excellent on the presence/absence of language delay (Kappa = 1.0), of regression or loss of skills in the course of development (Kappa = .58), of epilepsy (Kappa = .84), on overall language level (Kappa = .62), on estimate of intellectual functioning coded on 3 levels (normal range, mild retardation, moderate to profound retardation) (Kappa = .72), and on the presence/absence of any developmental disorder amongst first degree relatives (Kappa = .74). Reliability was lower for regression due to the difficulty in differentiating loss of skills from developmental stagnation and to establish language level before the reported loss occurred.

Discussion

We have shown that the positive predictive value of a diagnosis of autism recorded in the electronic health record of patients in the General Practice Research Database is high, and higher than in a previous study where the diagnosis of autism was confirmed in 80% of 83 subjects

with a GPRD computer record of autism [25]. A high positive predictive value for other morbidity data recorded in the GPRD has been found for a range of other conditions. For example 94% of cases of cataract identified had their diagnosis confirmed by a review of hospital eye service discharge summaries [26] and a recorded diagnosis of myocardial infarction was confirmed in over 90% of cases [27]. In our study, the diagnosis of PDD was confirmed by expert review of the notes in 92.5% of the cases. Amongst the unconfirmed cases were several records with poor quality data which precluded a positive confirmation of the diagnosis. A North London study based on a disability register identified a similar proportion of confirmed cases (89%) in their diagnostic validation [28]. The study design precluded an estimate of the sensitivity of a GP's diagnosis of a PDD, that is the percentage of children with a PDD who did not have this recorded in the GP records. This would have required a much more extensive study.

The characteristics of the children with PDD were consistent with those of published studies on autistic samples. Thus, for children with an autistic disorder from this study, the male/female ratio was 4.3:1 (146/34) and the proportion of subjects without intellectual impairments was 25.3%; these results compare well to respective figures of 4.3:1 and 30% deriving from recent reviews of epidemiological surveys of autism [1]. Equally, the 7.8% risk of PDD in the first-degree relatives is in line with current estimates [29]. Similarly, the rate of 19% (24% in the autism group) for regression in the course of the development are consistent with the rates reported in earlier ([30]:

31.3%) or more recent ([31]: 25%; [4]: 15.6%) epidemiological surveys or in various clinical samples ([32]: 37.2%; [33]: 29.6%; [34]: 30%). The relatively wide range of regression rates across studies reflects the different definitions and methods of data collection used in these studies. Therefore, on a range of indices, our sample characteristics were typical of studies including well characterized PDD children.

Within the PDD spectrum, a relatively small proportion was identified as having a PDDNOS or Asperger Disorder. This would not have been surprising in earlier years, since the diagnosis of Asperger Disorder was not defined until 1992, and therefore many of these children will not have received a diagnosis that lead to ascertainment in the sample. However, although their number increased, there were still few children with these diagnoses among the most recent birth cohorts in the GPRD. This was unexpected: evidence from epidemiological surveys suggests that the prevalence of PDDNOS is higher than that of autistic disorder [1,4]. This is consistent with three alternative explanations: first, a precise differentiation between autism and PDDNOS has not been possible in our study (maybe due to our particular mode of data collection), leading to an inclusion of children with atypical forms of autism in the autistic disorder group; second, that some children with atypical autism were not diagnosed as having a PDD at all; and third, that among those with a PDD diagnosis, the recording of diagnosed PDDNOS into GPRD is less complete than of autism. The fact that the severity of autism, as indicated by gender ratio, intellectual and language levels, decreased over time supports the first interpretation. It could be the case, however, that this trend also reflects a genuine change in the association between autistic disorder and mental retardation, possibly due to earlier diagnosis and intervention.

Conclusions

We conclude from this validation study that the positive predictive value of the diagnoses of autistic spectrum disorders or of the broad category of PDDs is good in the GPRD. The differentiation of PDD subtypes within this broad PDD class was, however, less good.

Authors' contributions

LH and EF coded all the medical notes. LM, EF and CC analyzed the data. EF, LS, AH, LR, PS took part in study design, interpretation of the results and writing the paper. All authors read and approved the final manuscript.

Acknowledgments

References

1. Fombonne E: **Epidemiological surveys of autism and other pervasive developmental disorders: an update.** *Journal of Autism and Developmental Disorders* 2003, **33**:365-381.
2. Lord C, Rutter M, Le Couteur A: **Autism Diagnostic Interview: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders.** *Journal of Autism and Developmental Disorders* 1994, **24(5)**:659-85.
3. Lord C, Risi S, Lembrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M: **The Autism Diagnostic Observation Schedule-Generic: a standard measure of social and communication deficits associated with the spectrum of autism.** *Journal of Autism and Developmental Disorders* 2000, **30**:205-223.
4. Chakrabarti S, Fombonne E: **Pervasive developmental disorders in preschool children.** *JAMA* 2001, **285(24)**:3093-3099.
5. Yeargin-Allsopp M, Rice C, Karapurkar T, Doernberg N, Boyle C, Murphy C: **Prevalence of autism in a US metropolitan area.** *JAMA* 2003, **289(1)**:49-55.
6. Fombonne E: **The prevalence of autism.** *JAMA* 2003, **289(1)**:1-3.
7. Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, Berelowitz M, Dhillon AP, Thomson MA, Harvey P, Valentine A, Davies SE, Walker-Smith JA: **Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children.** *Lancet* 1998, **351**:637-641.
8. Wakefield A: **MMR vaccination and autism – letter.** *Lancet* 1999, **354(11.09.99)**:949-950.
9. Wakefield AJ, Anthony A, Murch SH, Thomson M, Montgomery SM, Davies S, O'Leary JJ, Berelowitz M, Walker-Smith JA: **Enterocolitis in children with developmental disorders.** *American Journal of Gastroenterology* 2000, **95(9)**:2285-2295.
10. Bernard S, Enayati A, Roger H, Binstock T, Redwood L: **The role of mercury in the pathogenesis of autism.** *Molecular Psychiatry* 2002, **7**:S42-S43.
11. Taylor B, Miller E, Farrington CP, Petropoulos MC, Favot-Mayaud I, Li J, Waight PA: **Autism and measles, mumps, and rubella vaccine: no epidemiological evidence for a causal association.** *Lancet* 1999, **353(June 12)**:2026-2029.
12. Dales L, Hammer S, Smith N: **Time trends in autism and MMR immunization coverage in California.** *Journal of the American Medical Association* 2001, **285**:1183-1185.
13. Croen LA, Grether JK, Hoogstrate J, Selvin S: **The changing prevalence of autism in California.** *Journal of Autism and Developmental Disorders* 2002, **32(3)**:207-215.
14. Kaye J, Melero-Montes MdM, Jick H: **Mumps, measles, and rubella vaccine and the incidence of autism recorded by general practitioners: a time trend analysis.** *BMJ* 2001, **322**:0-2.
15. DeWilde S, Carey IM, Richards N, Hilton SR, Cook DG: **Do children who become autistic consult more often after MMR vaccination?** *Br J Gen Pract* 2001, **51(464)**:226-7.
16. Madsen KM, Hviid A, Vestergaard M, Schendel D, Wohlfahrt J, Thorsen P, Olsen J, Melbye M: **A population-based study of measles, mumps, and rubella vaccination and autism.** *N Engl J Med* 2002, **347(19)**:1477-82.
17. Chen W, Landau S, Sham P, Fombonne E: **No evidence for links between autism, MMR and measles virus.** *Psychological Medicine* 2004, **34**:543-53.
18. Spitzer WO: **Measles, mumps, and rubella vaccination and autism.** *N Engl J Med* 2003, **348(10)**:951-4.
19. Smeeth L, Hall AJ, Fombonne E, Rodrigues LC, Huang X, Smith PG: **A case-control study of autism and mumps-measles-rubella vaccination using the general practice research database: design and methodology.** *BMC Public Health* 2001, **1**:2.
20. Smeeth L, Cook C, Fombonne E, Rodrigues LC, Smith PG, Hall AJ: **Rate of first recorded diagnosis of pervasive developmental disorders in United Kingdom general practice, 1988 to 2001.** *BMC Medicine* in press.
21. Walley Mantgani A: **The UK General Practice Research Database.** *Lancet* 1997, **350**:1097-9.
22. Perry J, ed: *OXMIS problem codes for primary medical care Oxford OXMIS Publications*; 1978.
23. Department of Health: *New clinical classification system will streamline computerised medical records. London* 1990.
24. Bartko JJ: **On various intraclass correlation reliability coefficients.** *Psychological Bulletin* 1976, **83(5)**:762-765.
25. Black C, Kaye JA, Jick H: **Relation of childhood gastrointestinal disorders to autism: nested case-control study using data**

    from the UK General Practice Research Database. *BMJ* 2002, **325(7361):**419-21.

26. Ruigomez A, Garcia Rodriguez LA, Dev VJ, Arellano F, Raniwala J: **Are schizophrenia or antipsychotic drugs a risk factor for cataracts?** *Epidemiology* 2000, **11:**620-3.

27. Jick H, Derby LE, Gurewich V, Vasilakis C: **The risk of myocardial infarction associated with antihypertensive drug treatment in persons with uncomplicated essential hypertension.** *Pharmacotherapy* 1996, **16:**321-6.

28. Lingam R, Simmons A, Andrews N, Miller E, Stowe J, Taylor B: **Prevalence of autism and parentally reported triggers in a north east London population.** *Arch Dis Child* 2003, **88(8):**666-70.

29. Szatmari P, Jones M, Zwaigenbaum L, MacLean J: **Genetics of Autism: Overview and New Directions.** *Journal of Autism and Developmental Disorders* 1998, **28(5):**351-368.

30. Lotter V: **Epidemiology of autistic conditions in young children: I. Prevalence.** *Social Psychiatry* 1966, **1:**124-137.

31. Taylor B, Miller E, Lingam R, Andrews N, Simmons A, Stowe J: **Measles, mumps, and rubella vaccination and bowel problems or developmental regression in children with autism: population study.** *BMJ* 2002, **324(7334):**393-6.

32. Kurita H: **Infantile autism with speech loss before the age of thirty months.** *Journal of the American Academy of Child Psychiatry* 1985, **24:**191-196.

33. Kobayashi R, Murata T: **Setback phenomenon in autism and long-term prognosis.** *Acta Psychiatrica Scandinavica* 1998, **98(4):**296-303.

34. Tuchman R, Rapin I: **Regression in pervasive developmental disorders: seizures and epileptiform electroencephalogram correlates.** *Pediatrics* 1997, **99(4):**560-566.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2458/4/5/prepub