Describing the phenotype in Rett syndrome using a population database

L Colvin, S Fyfe, S Leonard, T Schiavello, C Ellaway, N de Klerk, J Christodoulou, M Msall, H Leonard

Rett syndrome (RTT) is a neurodevelopmental disorder, first recognised in 1966 by Dr Andreas Rett, and published in the English speaking literature in 1983. Following publication of several case series and in the absence of a biological marker, the Tretban diagnostic criteria were developed by an international working group. Necessary criteria for a diagnosis of classical RTT are: apparently normal prenatal and perinatal period; apparently normal psychomotor development between birth and 6 months of age; normal head circumference at birth; deceleration of head growth between 5 months and 4 years of age; loss of purposeful hand skills between 6 and 30 months of age; development of severely impaired receptive and expressive language; development of stereotypic hand movements; and the appearance of gait ataxia and truncal apraxia/ataxia between the ages of 1 and 4 years. Supportive criteria including breathing dysfunction, EEG abnormalities, seizures, spasticity, peripheral vasomotor disturbance, scoliosis, growth retardation, and hypotrophic small feet were developed to assist further with diagnosis and describe the phenotype. The presence of all necessary criteria is required for a diagnosis of classical RTT. However, it has become clear that the phenotypic spectrum extends beyond this. Individuals with some, but not all characteristics have been categorised as atypical or as one of six variant forms. However, with the exception of a Swedish series where forme fruste represented 11.5% of cases, there is little population based data on the prevalence of different RTT variants. Moreover, there have been no large systematic longitudinal studies comparing the clinical course of classical and atypical cases over time, although follow up data on a series of RTT females representing just under a quarter of cases in the Swedish register were recently published.

The most comprehensive data relating to the clinical profile of RTT come from a Swedish study of 105 girls and women with classical and atypical RTT. This study documented aspects of development, neuromotor impairment and disability, epilepsy, and breathing irregularities. The previously defined staging system was applied to and evaluated for a group of these cases. In a subgroup with classical RTT aged between 22 and 44 years (n = 30), six had never walked and 18 had lost the ability to walk. In the entire series, seizures were reported in 65% of girls aged 3–10 years and 94% of those aged 11–20 years. These findings are consistent with a more recent study of 53 females from Western Sweden where a history of epilepsy was reported in 94%.

There has been no population based research relating to activities of daily living, cognitive function, and communicative abilities in RTT. In a recent clinic based series, the language development of 99 cases was documented and showed that just over half (55.5%) did develop meaningful words, with eight cases speaking two word sentences. However, most (87.5%) ceased these utterances and this generally occurred before 40 months. In other studies, relative preservation of gross motor and daily living skills according to the developmental level achieved before regression has been reported in 15 girls and correlation of skills with age of onset in 28 girls. To gain accurate information about prevalence of functional ability, behavioural, and other features, studies using well validated measures, systematic protocols, and large numbers of subjects are needed.

Now that mutations in the MECP2 gene have been found to be associated with RTT, there have been a number of attempts to correlate genotype with phenotype. Individual research groups have developed their own severity scales to describe the phenotype. The recently published guidelines...
for reporting clinical cases of Rett could also be adapted for this purpose. However, the absence of a universally accepted scale for determining phenotypic variation or clinical severity is clearly a problem. The Australian Rett Syndrome Database (ARSD) is a unique population based registry of cases born since January 1976. This database has collected phenotype data on cases reported to the study since 1993. It has already been used as a source of cases for a number of epidemiological, clinical, and radiological studies. Inclusion of molecular genetic data now enhances the registry and enables differentiation of those females with an identified MECP2 mutation. This paper describes the clinical characteristics in a subset of the registry, alive in 2000, and representing 80% of cases ever enrolled in the study.

METHODS
Case classification
The original cohort of 134 Australian cases aged less than 18 years was established in 1993 using multiple sources, including the Australian Paediatric Surveillance Unit (APSU). Ongoing ascertainment of cases continued, with the parent support group, the Rett Syndrome Association of Australia, being the major source of cases from 1995 to 1999 coupled again with the APSU since 2000. Presence or absence of necessary criteria is generally determined by combined use of information from clinicians and families. Cases are categorised, as previously described, with verification of a classical case requiring all eight necessary criteria. A case is considered atypical if at least six of the eight are present, provided that the six criteria included loss of hand skills or development of hand stereotypes. An atypical categorisation also requires the presence of a minimum of three primary inclusion criteria from the variant delineation model. We have been less stringent in the classification of children under 5 years (n = 11) who at presentation did not meet criteria for classical or atypical Rett and, according to Hanefeld and colleagues, should be considered potential. Cases are designated atypical generally because they are more severe and less likely to have normal early development, or milder (previously categorised as formé fruste) with neither ambulatory ability nor head growth compromised. Therefore, we decided to group our cases into three categories: “classical,” “mild atypical” (most likely to have remained ambulant and not to have experienced head growth deceleration), and “early onset atypical” (less likely to have had a normal prenatal and perinatal period or normal development in the first six months of life).

Phenotype data collection and scoring
During 2000, all families in the study were invited to complete a follow up questionnaire while DNA samples were collected on as many cases as possible. The questionnaire consisted of 15 sections, some of which had already been piloted on a mostly US parent population using the internet. These particular sections included a questionnaire version of the WeeFIM (the Functional Independence Measure for Children). The WeeFIM measures typical performance (by observation or parental report) of children in essential self care, mobility, and communication-social learning tasks. The questionnaire also contained questions on the child’s medical conditions, hospital admissions, use of medical and therapy services, education, and accommodation options. Behavioural items modified for use in Rett and the symptom index score, which had previously been developed and used, were also included.

After pilot testing, questionnaires were mailed to families. Follow up phone calls were used to improve response. The option of response through an internet based form was made available and taken up by 16 of the respondents. All questionnaire information was entered into a computer database. Data from the original family and clinician questionnaires, which had been used to define the Trevathan criteria, were also used for historical details such as age at amputation and regression.

Within the constraints of the available data, we assigned a severity level for each item outlined in three scales:

1. A scale suggested by Kerr and colleagues: Kerr scale (table 1)
2. A scale modified by Percy (personal communication, 11 September 2001) and Schanen from Amir and Zoghbi: Percy scale (table 2)
3. A scale modified by Pineda (personal communication, 11 September 2001) from that designed by Monros and colleagues: Pineda scale (table 3). (All tables can be viewed on the ADC website: www.archdischild.com)

We were able to use 19 of 20 items suggested by Kerr and colleagues to produce a severity score with a maximum of 37.

The Percy scale uses more developmental data to contribute to the phenotype severity than the Kerr guidelines. It includes age at regression, milestones for sitting, crawling, and walking, onset of stereotypes, growth changes, and age at which communication skills were lost. The Percy scale also includes non-verbal communication abilities as well as language. The 15 questions in the score produce a maximum severity score of 47.

The Pineda score also increases with increasing severity, with a maximum score of 31. It contains 10 items which are similar to those used in the Percy scale, but does not include information about crawling, non-verbal communication, feeding, somatic growth, autonomic dysfunction, and scoliosis. As a measure of regression, the Pineda scale uses “loss of social interaction” for which we used the response to the question on age at loss of communication skills. The only additional item not included in any other scale is air swallowing.

We also calculated a WeeFIM score from the questionnaire items in the follow up questionnaire. The WeeFIM produces a maximum score of 126 for independent functioning in a normal 8 year old child. Unlike the other three measures, the higher the WeeFIM score, the higher the level of independent functioning.

Algorithms were generated which compiled scores from multiple items on the original family questionnaire, the clinician questionnaire, and the follow up questionnaire. If height and weight questions were unanswered on the follow up questionnaire, data from previous measures were used if available. Z scores and percentiles were determined using NIH norms. Birth head circumference Z scores were calculated using New South Wales norms as our best proxy for “head circumference during the first year”, but even so were not available for all cases.

Some items were associated with greater proportions of missing data: information on muscle tone and intellectual disability (table 1) were only obtainable from clinicians who, sometimes, could not, or did not feel it appropriate, to respond to these questions, especially to the latter in young children. Questions relating to age at onset of stereotypes and autonomic dysfunction (table 2) were not in the first version of the questionnaire used for the original local Western Australian cases but added later. Age at regression (table 2) and loss of social interaction (table 3) were based on age at loss of communication. However, the question on age at loss of communication was not always answered, especially in children who had never actually developed normally. There was not a specific question on crawling; this information had to be inferred from responses to other questions (table 2) so that data were incomplete. Similarly, some parents did not provide information needed to quantify the level of non-verbal communication. Where data were missing for particular scale...
items, the average score for that item for the classification group (classical, mild atypical, or early onset atypical) was used.

Mutation screening
DNA was extracted from blood samples. Polymerase chain reactions (PCR) amplifying the MECP2 coding region were performed and the majority sequenced using previously published methods, except that PCR products were cleaned using shrimp alkaline. For about one quarter of the samples, screening was performed by denaturing high performance liquid chromatography, followed by confirmatory direct sequencing.

Ethical approval for this and the epidemiological component of the study was obtained from local institutional committees.

RESULTS
Total cases by birth year and age at ascertainment
Cases have been registered with the Australian RTT Study for births since 1976 (fig 1, table 4). Age at ascertainment is defined as age when the clinician or family questionnaire is received, whichever is greater. Mean age at study ascertainment has fallen considerably by birth year, with a mean of 18.4 years for 1976, 4.8 years for 1994, and 2.8 years for 1997 births. The decrease in mean age prior to 1993 is caused by the “catch up” effect of establishing the registry. For those births after 1993, the decrease is probably a reflection of the close follow up by the study group once the APSU and parent support notifications are received.

By July 2000, 199 verified cases had been reported, with a further 19 cases pending verification (fig 2). Twelve cases died prior to July 2000. In nine of the 187 cases, information had only been provided by the clinician. Of the remaining 178 families, there were 161 with whom current contact had been maintained and were thus eligible for the follow up study. Questionnaires were returned by 152/161 and this is the subset described here.

Of the 152 cases in the follow up cohort, 84 (55%) were categorised as “classical”, 40 (26%) as “early onset atypical”, and 28 (19%) as “mild atypical”. In the follow up cohort, the proportion of classical cases in the oldest group was nearly double (73%) that in the youngest age group (42%). Mutation screening has now been completed in 134 of these 152 cases (88%) with a pathogenic mutation detected in 96 (72%). Mutations were found in 73% (55/75) of classical cases, 62% (21/34) of the early onset atypical cases, and 80% (20/25) of mild atypical cases.

Kerr scale
There was considerable variation in severity across the items in the Kerr scale (table 5). The majority of cases showed severe abnormality (severity level = 2) in items reflecting the necessary criteria—that is, gross motor function (64%), hand stereotypes (91%), intellectual disability (97%), and speech (95%). Overall there was a lower proportion of severe abnormality in spinal posture (21%), oromotor difficulties (12%), and epilepsy (18%).

For present head circumference, weight, height, muscle tone, and spinal posture there was variation with age with a greater proportion in the more severe category in the oldest age group (table 5). The majority of cases showed severe abnormality (severity level = 2) in items reflecting the necessary criteria—that is, gross motor function (64%), hand stereotypes (91%), intellectual disability (97%), and speech (95%). Overall there was a lower proportion of severe abnormality in spinal posture (21%), oromotor difficulties (12%), and epilepsy (18%).

For present head circumference, weight, height, muscle tone, and spinal posture there was variation with age with a greater proportion in the more severe category in the oldest age group (table 5). The majority of cases showed severe abnormality (severity level = 2) in items reflecting the necessary criteria—that is, gross motor function (64%), hand stereotypes (91%), intellectual disability (97%), and speech (95%). Overall there was a lower proportion of severe abnormality in spinal posture (21%), oromotor difficulties (12%), and epilepsy (18%).

For present head circumference, weight, height, muscle tone, and spinal posture there was variation with age with a greater proportion in the more severe category in the oldest age group (table 5). The majority of cases showed severe abnormality (severity level = 2) in items reflecting the necessary criteria—that is, gross motor function (64%), hand stereotypes (91%), intellectual disability (97%), and speech (95%). Overall there was a lower proportion of severe abnormality in spinal posture (21%), oromotor difficulties (12%), and epilepsy (18%).

The mean Kerr score increased with age from 20.5 (SD 2.9) in the under 7 years age group to 24.6 (SD 4.2) in those between 12 and 17 years, and 24.2 (SD 4.8) in those over 17 years (p < 0.001). The pattern of the profile also varied by age group, with severe sleep and mood disturbances, oromotor difficulties, and involuntary movements contributing less, and present head circumference and height abnormalities contributing more, to the overall score in the oldest age group.
Phenotype in Rett syndrome

Percy scale
For all cases with available data, 37% (51/136) regressed before 18 months and 76% (103/136) regressed before 30 months (table 6). While almost all (146/148) children could sit and maintain their position by or after eight months, only just under half (71/152) walked by 18 months. Another 14% (21/152) walked later or with help but 39% (60/152) never walked.

In about four fifths (81%) of cases, onset of stereotypies occurred before 36 months. Although hand use had been acquired and conserved in 19 children (12%), only one was reported to feed independently. Of the 152 cases, nearly half had acquired and lost (n = 65) or never acquired hand use (n = 5), and three quarters could not attempt to feed themselves. Somatic growth was severely affected with height-for-age two or more SD below the mean in over half (74/136) of the cases, with only 10% (14/136) of cases above the mean height-for-age.

Nearly 40% (50/129) of cases were reported to have never acquired any non-verbal communication, with the remainder showing varying levels of eye pointing, eye gaze, or body language to show their intent. However, only two of 149 cases (1%) had speech that was more than single words and a further six (4%) used single words. While 67% (100/149) babble and vocalised, 28% (41/149) used no utterances.

As with the Kerr scale, some scores on the Percy scale changed with age. The parameters of head growth, epilepsy, somatic growth, autonomic dysfunction, and scoliosis varied significantly with age. Onset of stereotypies was also more likely to be of later onset in the older age groups compared with younger aged children.

Pineda scale
The mean Pineda score was 16.3 (SD 4.3) for all cases and did not differ by age group (p = 0.2) (table 7). However, the individual parameters of head growth (p = 0.04) and epilepsy (p < 0.001) varied significantly with age. Just under half of all cases (68/151) were reported to swallow air. Only one case was reported as having lost social interaction before the age of 6 months. Over half (78/135) lost interaction between the ages of 6 and 18 months, with 41% (56/135) showing this loss after the age of 18 months.

WeeFIM score
The WeeFIM provides a composite score with a maximum possible of 126 for independent functioning and also subscores for the areas of self care (max = 56), mobility (max = 35), and cognition (max = 35). The mean composite score for all cases was 29.0 (SD 11.9) (table 8). Performance was greatest in the area of mobility where 58/148 (39%) of cases were mobile, 12% (18/149) were able to manage stairs independently or with assistance, 22% (33/147) were able to get in and out of a chair alone or with assistance, and 11% (17/150) were able to transfer in and out of a bath or shower, either independently or with assistance.

“Classical” cases had a mean score of 25.7 (SD 8.0) with a range of 18 to 66. The “mild atypical” group had a higher mean WeeFIM score of 40.7 (SD 15.3) and a range of 19 to 75, whereas the “early onset atypical” group was more similar to the “classical” group with mean score of 27.8 (SD 11.2). While there were no significant differences in WeeFIM score with age, children aged 7–12 years had higher mean WeeFIM scores (32.2, SD 14.6) than younger children (29.3, SD 9.1), children aged 12–17 years (26.8, SD 9.6), or children aged over 17 years (28.2, SD 12.2). However, this group also displayed more variation in the score distribution.

The proportion of atypical cases (45%) was higher than in our original cohort, where our finding of 32% was consistent with the proportion (30%) found by Hagberg et al in a similar age group. However, in the follow up study we had a higher participation of parents of recently enrolled younger children who were less likely to be classical. While previously RTT diagnosis was considered tentative under 5 years, it is now made confidently at younger ages. Thus, some children not meeting all the criteria at initial presentation, may do so over time. The high proportion of the early onset variant may both reflect an increasing awareness of the disorder and also the fact that the early development may not always be normal.

DISCUSSION
This paper has expanded on the descriptive epidemiology of RTT by describing the range of clinical severity in a population based cohort of affected children. Despite the major activity relating to the genetic investigation of RTT, the clinical profile has not been further examined since the Swedish cross sectional study over a decade ago. While Opitz and Lewin warned about the premature truncation of phenotype to fit with artificial criteria, consideration of the existence of atypical forms led to Hagberg and Skjeldahl defining criteria for RTT variants and fitting these to a small case series.

The need for revision of some of the essential criteria has become even clearer since the identification of the gene and is currently in progress (Alan Percy, personal communication, 11 September 2001). Hagberg’s recent work has already shown that head growth deceleration is not mandatory. For our work, it was necessary, using the Trevathan criteria, to develop algorithms to classify our cases. However, as previously commented, the supportive and pathognomonic features may have been so obvious at times that clinicians could not help but make a diagnosis of classical RTT even in the absence of all the necessary criteria. Thus, it is possible that our epidemiological diagnosis is more stringent than a clinical diagnosis and may account for our higher proportion of atypical cases.

The proportion of atypical cases (45%) was higher than in our original cohort, where our finding of 32% was consistent with the proportion (30%) found by Hagberg et al in a similar age group. However, in the follow up study we had a higher participation of parents of recently enrolled younger children who were less likely to be classical. While previously RTT diagnosis was considered tentative under 5 years, it is now made confidently at younger ages. Thus, some children not meeting all the criteria at initial presentation, may do so over time. The high proportion of the early onset variant may both reflect an increasing awareness of the disorder and also the fact that the early development may not always be normal. Alternatively there may be a few children, labelled as potential cases, who in time turn out not to have RTT.

Overall, the proportion of our cases with mutations identified is no different from other studies that do not differentiate classical and atypical cases, except that with our larger population our estimate is more precise. Our results relating specifically to classical cases are also similar to other studies. We were surprised to find such a high prevalence (80%) of identified mutations in our “mild atypical” group, although in the “early onset atypical” group the prevalence was lower (62%). Nevertheless the prevalence in our atypical cases overall was still higher than in the few other case series which have examined this. These findings could partly reflect our stringent case definition, which labels as atypical some cases regarded by others as classical. The overall level of mutation identification in our total population based cohort was comparable with other studies. Thus, we are confident that case verification was more than adequately achieved despite study investigators not clinically examining all cases. Therefore, the descriptive epidemiology of the distribution of the main features encompassing the clinical spectrum of RTT is lacking on a population basis. Behavioural characteristics and features such as the range and repertoire of hand stereotypes, have not been comprehensively studied. Thus, there is a real concern that clinicians are not being provided with the information they need to diagnose and prognosticate about this disorder. Irrespective of what we still have to learn about genotype-phenotype relations, there is little quantitative epidemiological evidence on the proportion of children likely to be affected by the various features and at what ages these effects, including death, are likely to occur.

Although the results in this report relate to cross sectional data, the clinical profile, as expected, changes with age. Deceleration of somatic growth and microcephaly are found more
commonly in older females. Older girls and young women with RTT are also more likely to suffer from epilepsy, periphereal circulatory problems, and scoliosis. Stereotypies that generally begin before 30 months of age are maintained throughout childhood but are slightly less prevalent in young women as are involuntary movements. However, the ability to walk declines with age and contributes more to the low WeeFIM score in young women than in children under the age of 12. Although the overall WeeFIM score does not change significantly with age, children between the ages of 7 and 12 have the highest mean scores, suggesting that their functional abilities are maximal at this age. In our previous study using an opportunistic sample of 86 children and adults, the mean WeeFIM score was slightly higher at age 34.5 and in contrast to the current study, did improve with age. However, our earlier study was not population based, and the older women included were likely to be a biased group living at home and receiving optimal therapy and stimulation. Nevertheless, in both studies the functional abilities of these children and adults generally were very low with all dependent, and most completely dependent, on others for self care.

In the cross sectional evaluation of age related features in RTT, there is likely to be a selection towards survivors, with the most severely affected dying earlier. This could account for our findings of less severe epilepsy and circulation problems in the over 17 years compared with the 12–17 year age group. It might also account for the fact that age at development of hand stereotypies tended to be earlier in the older compared with the younger age groups. We previously showed that age at development of stereotypies was significantly later in missense than truncating mutations. If truncating mutations are more severe, it is feasible that children with truncating mutations might have a shorter lifespan than those with missense mutations, and this could lead to a bias in the older age group. There is an urgent need for longitudinal studies such as we are currently undertaking, which will follow a cohort over time and account accurately for loss to follow up by death or other reasons in the analysis and interpretation of findings. This is particularly so in genotype-phenotype correlation studies, where analyses limited to survivors could be seriously flawed.

Of the four scales we used, only the WeeFIM has undergone formal analysis of its psychometric properties. In tandem with the age related changes in specific characteristics, total severity scores also worsen with age in the Kerr and Percy scales but not in the Pineda scale. Thus in research examining how the phenotype is affected by the genotype, it is essential to account for the confounding effect of age either in the statistical analysis or by using an age independent scale such as the Pineda scale.

The main limitation of this study is our forced adaptation of the scales in keeping with the nature of the retrospectively collected data and the fact that for the occasional item we did not have an up to date severity level and were forced to rely on information provided at enrolment. In the process of developing an international phenotype database, fewer modifications would be required as we could ensure that sufficient questions to provide all this information were included and missing data would be minimised. Nevertheless, items suggested in some of these scales may still not be available universally. For instance, although in the USA healthy children under 2 years may have regular and frequent paediatric assessments that include growth monitoring, this is not standard practice in Australia or in many other countries.

The major strength of this study is its population based nature. Because it uses multiple sources and active ascertainment, it should not be biased against particular sociodemographic groups, a problem that may often occur in clinic based samples. As has not yet been possible elsewhere, the phenotype-genotype correlations which we are undertaking will be enhanced by the use of population based data. This cohort also provides the necessary framework for longitudinal follow up studies. Such research, although extremely important, is still, after nearly 20 years, not yet informing treatment or intervention for RTT patients.

ACKNOWLEDGEMENTS
We would like to thank the clinicians who have reported cases and the families for their ongoing participation in our study. We would particularly like to thank Bill Callaghan and the Lillis and Aitken families. We would also like to acknowledge the Australian Paediatric Surveillance Unit and the Rett Syndrome Association of Australia for their support. We would also like to thank the International Rett Syndrome Association (IRSA) for organising the meeting in Baden-Baden that discussed diagnostic criteria and severity scales. We would also like to acknowledge all those who contributed severity scales at the meeting. Funding for this project has been provided by the Financial Markets Research and Development Grant 003209. We would also like to acknowledge Linda Raffaele, Mark Davis, Bruce Bennetts, Sarah Williamson, and Catherine Stevenson who did the molecular laboratory work.

The tables can be viewed on the ADC website (www.archdischild.com)
Effect of a fourth Haemophilus influenzae type b immunisation in preterm infants who received dexamethasone for chronic lung disease. P Clarke, P J Powell, D Goldblatt, et al