

## Abstract

We set out to replicate the work of Wright et al, demonstrating the link between early years foundation stage profile (EYFSP) scores and autism spectrum diagnosis, using a larger cohort taken from the Connected Bradford project. We conducted a retrospective data linkage study, using the post-2013 EYFSP scores from the Connected Bradford Department for Education data, and linked these with education census and primary care (GP records). The outcome measures were diagnosis of ASD using associated primary care (GP) practice codes (SNOMEDs). As with the Wright et al analysis, we used a “total EYFSP score” and the same subscore consisting of five key learning goals identified using a panel of early-years autism experts.

This study validates the results of Wright et. al, and further demonstrates the feasibility of linking education and healthcare records using the Connected Bradford platform. A total of 70,277 children had linked primary care and education data, compared with a cohort of 8,935 children in the Wright et. al. study taken from Born In Bradford study. 17.3% of the cohort scoring  $<25$  on the total EYFSP score, and 15.4% scoring  $<8$  on the EYFSP subscore proposed in the Wright et al. analysis - individuals in these groups had a 11.09/16.5-fold increased odds of receiving an autism diagnosis respectively, when compared to baseline individuals scoring above these bounds.

Notes:

Comment on comparison of stats with Wright analysis Make point about using EYFSP to identify children at need of educational help

# 1 Introduction

introduction here

## 2 Methods

### 2.1 Design

This was a retrospective data linkage study. Our aim was to replicate the work of Wright et al., demonstrating the link between EYFSP scores and ASD and, in general, the utility of large connected datasets. This study expands the cohort beyond participants in the Born in Bradford study, to the entire Bradford district using Connected Bradford data. Connected Bradford hosts post-2013 EYFSP scores and education census data up to the 2019 academic year, comprising the 70,277 unique individuals assessed for the EYFSP in the Bradford district during that period, inclusive of the 8,935 children from the Wright study. Education data were linked with primary care records collected from 86 GP surgeries in the Bradford district.

### 2.2 EYFSP Observations

The Early Year's Foundation Stage Profile - EYFSP - summarises and describes children's learning and development in accordance with 7 areas of learning, subdivided into 17 early learning goals (Figure 1). Assessments are conducted in the final term of the year in which a child reaches age 5 and are used to support the transition into the national curriculum Key Stage 1. Scoring is based on an observational assessment conducted by teaching practitioners and guided by a framework set out by the UK Department for Education[1]. Scores are recorded for each goal as "emerging" (not meeting the expected level of development for this goal), "expected" (meeting the expected level of development), "exceeding" (exceeding the expected level of development), or "absent for long periods or recently arrived". For the purposes of this study the "absent for long periods or recently arrived" is considered a missing or NA result.

---

<b>Communication and language development</b>	<ul style="list-style-type: none"><li>• Listening and attention</li><li>• Understanding</li><li>• Speaking</li></ul>
<b>Physical development</b>	<ul style="list-style-type: none"><li>• Moving and handling</li><li>• Health and self-care</li></ul>
<b>Personal, social and emotional development</b>	<ul style="list-style-type: none"><li>• Self-confidence and self-awareness</li><li>• Managing feelings and behaviour</li><li>• Making relationships</li></ul>
<b>Literacy</b>	<ul style="list-style-type: none"><li>• Reading</li><li>• Writing</li></ul>
<b>Mathematics</b>	<ul style="list-style-type: none"><li>• Numbers</li><li>• Shape, space and measures</li></ul>
<b>Understanding of the world</b>	<ul style="list-style-type: none"><li>• People and communities</li><li>• The world</li><li>• Technology</li></ul>
<b>Expressive arts and design</b>	<ul style="list-style-type: none"><li>• Exploring and using media and materials</li><li>• Being imaginative</li></ul>

---

Figure 1: The 17 early learning goals as defined in the Early Years Foundation Stage profile

As with the Wright et al. methodology, we used only post-2013 EYFSP results and aggregated the results according to the following method: goals assessed as emerging were scored 1, expected were scored 2 and exceeding 3. Total scores were calculated for each individual, as well as the same 5-item subscore used by Wright et al. developed by a group of ASD assessment experts (Figure 2).

---

<b>EYFSP Weighted Subscore:</b>	
Using a weighted subscore, where four aspects of childhood autism (social, language and communications, imagination and repetitive behaviour) were mapped onto EYFSP elements	
<b>Speaking</b>	<ul style="list-style-type: none"> <li>• Personal, social and emotional: managing feelings and behaviour</li> <li>• Personal, social and emotional: making relationships</li> </ul>
<b>Social</b>	<ul style="list-style-type: none"> <li>• Communication and language: listening and attention</li> </ul>
<b>Imagination</b>	<ul style="list-style-type: none"> <li>• Expressive arts and design: being imaginative</li> </ul>
<b>Repetitive Behaviour</b>	<ul style="list-style-type: none"> <li>• Physical development: health and self-care</li> </ul>

---

Figure 2: Five item EYFSP subscore. The five learning goals used in the subscore were chosen from the four main symptom areas defined in the WHO (1992) research diagnostic criteria for ASD - social reciprocity, language and communication, imagination delays and repetitive and stereotyped patterns of behaviour. Because the social reciprocity domain is given more weight in this classification system, it was decided to include two items from the EYFSP to reflect this weighting

## 2.3 SNOMED coding of ASD

Unlike the "read codes" used in the primary care records in Born in Bradford, the Connected Bradford data use the more up-to-date SNOMED-CT codings currently in use by GP practices. As such, we compiled a list of relevant SNOMED codes to identify autism spectrum disorders.

SNOMED-CT is a hierarchical coding system. Conditions are organised in a tree-like structure with "parent" conditions and descendant "child" conditions. The "children" are considered to be more specific sub-divisions of the "parent" conditions from which they descend. For example, a patient may be diagnosed with "Epilepsy, SCTID: 84757009", whilst another patient may be diagnosed with the more specific child condition "Acquired epileptic aphasia, SCTID: 230438007" - the second patient could also be considered to have "Epilepsy", as it is a "parent" of the condition they've been diagnosed with.

We identified ASD in the SNOMED-CT codings as "Pervasive Developmental Disorder, SCTID: 35919005" ("Autism Spectrum Disorders" are listed as a secondary preferred term under this code). We then proceeded to collect 49 descendant conditions of this code, and used this list as our indicator of an ASD diagnosis[REFERENCE TO SNOMEDS TABLE - PROBS IN APPENDIX]. This method identified 1227 individuals, 1.75% of our cohort, as having an ASD diagnosis; this proportion is in line with expected population levels.

## 2.4 Analysis

Following the same methodology as Wright et al. we used logistic regression to model the association between EYFSP scores and ASD diagnoses, using the python Statsmodels package[2]. We included the same covariates: sex (male/female), free school meal status, ethnicity (grouped into White British/Pakistani/other ethnicity) and age of child at date of extract. We also chose to add an interaction term between sex and EYFSP score, as we hypothesized that sex and EYFSP performance are likely not independent of one another. EYFSP scores ranged from a total of 17 to 51, with mean 32.87 and SD 8.01, and a 5-item subscore of 5 to 15, mean 9.94 and SD 2.40. As in the previous study, the scores show a bimodal distribution with a distinct group of children achieving the expected level of

development and another somewhat distinct group with a majority of “emerging” development levels (Figure 3). As such, we chose to use the same approach to dichotomising the EYFSP scores, defining an individual score as “low” or “not low” according to a cutoff  $<1$  SD below the mean - this was a total score of less than 25 and a 5-item subscore of less than 8.

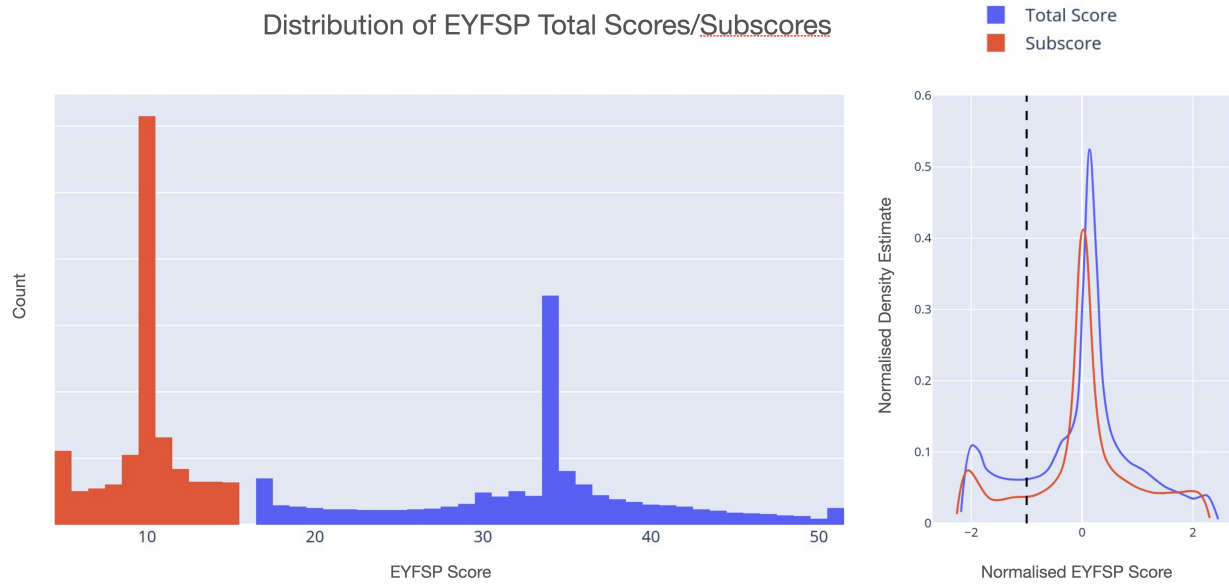


Figure 3: Histogram of EYFSP total and sub-scores (left) and normalised Kernel Density Estimates for distribution of Total Scores/Subscores (right). The black dashed line on the right plot represents the  $<1$  SD boundary for defining “low” scores/subscores

### 3 Results

1.75% (1227) of the 70277 individuals in the matched cohort had a diagnosis of ASD. Of the children with autism diagnoses 64.1% (786) had a “low” EYFSP total score and 68.5% (841) had a “low” 5-item subscore, compared with 16.5% (11359) and 14.5% (9979) respectively from the individuals without an ASD diagnosis (Table 1).

As a baseline for odds comparison, we used a female individual, with a “not low” EYFSP score/subscore, not eligible for free school meals, of white british ethnicity and of < 10 years of age. Our regression models showed that children with a low EYFSP subscore were approximately 16 times more likely to be diagnosed with ASD than our reference baseline (Table 2). Males were over 3 times more likely to be diagnosed with ASD than the baseline, and individuals of Pakistani or Other ethnicity were 0.62/0.47 times as likely to be diagnosed with an ASD compared to the baseline. The free school meals and age variables (excluding age 11, which given the effect size and confidence intervals is likely a statistical anomaly) were not statistically significant, .

The interaction between gender and EYFSP scores proved to be statistically significant. A male individual with a low EYFSP subscore was 36.3 times more likely to have an ASD diagnosis than baseline; females with a low subscore had a comparatively lower odds of 16.3 of being diagnosed with ASD. The greater odds for males with low scores was to be expected, given the increase in risk seen in males and individuals with low EYFSP scores individually; the interaction coefficient itself was negative. Thus, the interaction coefficient, male and low EYFSP score, served to decrease the estimated odds that would result were we to assume sex and EYFSP scores were independent.

			Grouped by ASD Diagnosis, n(%)		
		Missing	Overall	No	Yes
n			70277	69050	1227
Male	No	639	33607 (48.3)	33354 (48.8)	253 (20.6)
	Yes		36031 (51.7)	35062 (51.2)	969 (79.4)
Free School Meals Eligible	No	0	55725 (79.3)	54816 (79.4)	909 (74.1)
	Yes		14552 (20.7)	14234 (20.6)	318 (25.9)
Ethnicity,	White British	4528	35901 (54.6)	35129 (54.4)	772 (64.8)
	Pakistani		20509 (31.2)	20248 (31.4)	261 (21.9)
	Other		9339 (14.2)	9181 (14.2)	158 (13.3)
Age At Extract	<10	1	15857 (22.6)	15610 (22.6)	247 (20.1)
	10		10037 (14.3)	9901 (14.3)	136 (11.1)
	11		10058 (14.3)	9865 (14.3)	193 (15.7)
	12		10299 (14.7)	10107 (14.6)	192 (15.6)
	13		10281 (14.6)	10071 (14.6)	210 (17.1)
	>13		13744 (19.6)	13495 (19.5)	249 (20.3)
Has Low Total Score	No	0	58132 (82.7)	57691 (83.5)	441 (35.9)
	Yes		12145 (17.3)	11359 (16.5)	786 (64.1)
Has Low Subscore	No	0	59457 (84.6)	59071 (85.5)	386 (31.5)
	Yes		10820 (15.4)	9979 (14.5)	841 (68.5)

Table 1: Sample characteristics broken down by ASD diagnoses

	EYFSP Total Score		EYFSP 5 Item Sub-Score	
	Odds Ratio	95% CI (low, high)	Odds Ratio	95% CI (low, high)
<b>Univariate - EYFSP Score:</b>				
<b>EYFS Score:</b>				
Not Low (reference)	-	-	-	-
Low	9.1211	(8.085, 10.29)	12.943347	(11.427, 14.661)
<b>Multivariable - EYFSP Score and covariates:</b>				
<b>EYFS Score:</b>				
Not Low (reference)	-	-	-	-
low	11.090676	(8.557, 14.374)	16.504568	(12.676, 21.49)
<b>Gender:</b>				
Female (reference)	-	-	-	-
Male	3.454496	(2.758, 4.326)	3.317918	(2.618, 4.206)
<b>Interaction - Male and low Score*:</b>				
No (reference)	-	-	-	-
Yes	27.643759	(12.718, 60.099)	36.342941	(16.33, 80.883)
<b>Free School Meals Eligible:</b>				
No	-	-	-	-
Yes	0.943423	(0.823, 1.081)	0.902225	(0.787, 1.035)
<b>Ethnicity:</b>				
White British (reference)	-	-	-	-
Pakistani	0.473528	(0.409, 0.548)	0.476529	(0.412, 0.551)
Other	0.621275	(0.521, 0.741)	0.616304	(0.516, 0.736)
<b>Age:</b>				
< 10 (reference)	-	-	-	-
10	0.866965	(0.698, 1.077)	0.84456	(0.711, 1.1)
11	1.256707	(1.032, 1.531)	1.269895	(1.041, 1.549)
12	1.156423	(0.95, 1.408)	1.170753	(0.96, 1.428)
13	1.134527	(0.936, 1.375)	1.104517	(0.91, 1.34)
>13	0.985703	(0.819, 1.186)	0.953314	(0.792, 1.148)

Table 2: Odds ratios from logistic regression models. Total effects of EYFS scores are reported in the univariate models. Effects when controlling for covariates are reported in Multivariable models. \*Interaction effect odds are calculated using a combination of the individual “male”, “low score” and “male&low score” coefficients.

## 4 Discussion

discussion here

## References

- [1] Department for Education. Early years foundation stage profile handbook. <https://www.gov.uk/government/publications/early-years-foundation-stage-profile-handbook>, Sep 2021.
- [2] Skipper Seabold and Josef Perktold. *statsmodels: Econometric and statistical modeling with python*. 2010.