Prior health-related behaviours in children (2014-2020) and association with a positive SARS-CoV-2 test during adolescence (2020-2021): a retrospective cohort study using survey data linked with routine health data in Wales, UK

Emily Marchant[1,2*], Emily Lowthian[1], Tom Crick[3], Lucy Griffiths[1], Richard Fry[1,2], Kevin Dadaczynski[4,5], Orkan Okan[6], Michaela James[1,2], Laura Cowley[2,7] Fatemeh Torabi[1], Jonathan Kennedy[1,2], Ashley Akbari[1], Ronan Lyons[1], Sinead Brophy[1,2]

[1] Population Data Science, Medical School, Swansea University, Swansea, SA2 8PP, United Kingdom

[2] National Centre for Population Health and Wellbeing Research, Swansea University, Swansea, SA2 8PP, United Kingdom

[3] Department of Education & Childhood Studies, Swansea University, Swansea, SA2 8PP, United Kingdom

[4] Department of Nursing and Health Science, Fulda University of Applied Sciences, Fulda, 36037, Germany

[5] Center for Applied Health Science, Leuphana University Lueneburg, Lueneburg, 21335, Germany

[6] Chair Health Literacy, Department of Sport and Health Sciences, Technical University Munich, Uptown München-Campus D, 4. OG, OfficeL412, Georg-Brauchle-Ring 60/62, 80992 München, Germany

[7] Research and Evaluation Division, Public Health Knowledge and Research Directorate, Public Health Wales, 2 Capital Quarter, Cardiff, CF10 4BZ, United Kingdom

[*] Corresponding author: E.K.Marchant@swansea.ac.uk

Word count: 3,809

ABSTRACT

Objectives

Examine if prior health-related behaviours during primary school are associated with being tested for SARS-CoV-2 and testing positive during adolescence.

Design

Retrospective cohort study using an online cohort survey (1 April 2014 to 28 February 2020) linked to routine PCR SARS-CoV-2 test results (1 March 2020 to 31 August 2021)

Setting

Children attending primary schools in Wales (2014-2020), UK who were part of the HAPPEN schools network.

Participants

Complete linked records of eligible participants were obtained for n=6,891 individuals. 43.2% (n=3,021) were tested (baseline age 12.3±2.0, 48% boys) and 11.2% (n=774) tested positive for SARS-CoV-2 (baseline age 12.8±2.1, 43.9% boys).

Main outcome measures

Logistic regression of health-related behaviours and sex, age, deprivation, clustered by school was used to determine Odds Ratios (OR) of factors associated with being tested for or testing positive for SARS-CoV-2.

Results

Sleeping 9+ hours (OR=1.15, 95% CI 1.01 to 1.29), participating in 3+ out of school clubs (OR=1.15, 95% CI 1.02 to 1.31), able to swim (OR=1.29, 95% CI 1.10 to 1.52) and ride a bike

(OR=1.16, 95% CI 0.98 to 1.37, p<0.1) were associated with being tested for SARS-CoV-2. Participating in 3+ out of school clubs (OR=1.12, 95% CI 1.02 to 1.56), able to ride a bike (OR=1.36, 95% CI 0.97 to 1.92, p<0.1), sex (girl; OR=1.25, 95% CI 1.06 to 1.47) and baseline age (OR=1.16, 95% CI 1.10 to 1.22) were associated with an increased likelihood of testing 1.06 to 1.47

Conclusions

Actions associated with a child being PCR-tested and identified as positive may be related to parental health literacy e.g. parents recognising symptoms, knowledge of testing services. Identification of adolescent positive cases may be highly skewed towards children whose parents have higher health literacy. As those not accessing testing services remain undetected true rates of COVID-19 are not known in adolescence.

STRENGTHS AND LIMITATIONS

- First study to investigate association of prior child health-related behaviour measures with subsequent SARS-CoV-2 testing and infection during adolescence.
- Reporting of multiple child health behaviours linked at an individual-level to routine records of SARS-CoV-2 testing data through the SAIL Databank.
- Child-reported health behaviours were measured before the COVID-19 pandemic and represent historical health behaviour which may not reflect behaviours during COVID-19.
- Health behaviours captured through the national-scale HAPPEN survey represent children attending schools that engaged with the HAPPEN Wales primary school network which may not be representative of the whole population of Wales.

- The period of study for PCR-testing for and testing positive for SARS-CoV-2 includes a time frame with varying prevalence rates, different variants, approaches to testing children, public health measures and restrictions which were not measured in this study.

BACKGROUND

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in widespread disruption to the lives of children across the world, defined by the *Convention on the Rights of the Child* as a person under the age of 18 years[1]. This has impacted on their education and health-related behaviours such as nutrition and physical activity[2]. While a growing body of literature suggests children display fewer clinical symptoms[3], the COVID-19 pandemic has contributed to widened inequalities in children's health, wellbeing and education[4,5]. Positive SARS-CoV-2 tests require periods of self-isolation, impacting children's physical health and wellbeing, limiting opportunities for children to engage in health-promoting behaviours such as regular physical activity[6].

Childhood is a critical developmental period during which healthy habits are formed which transcend into adolescence, recognised by the World Health Organization as those aged between 10 and 19 years (early; 10-14 years, middle; 15-17 years, late; 18+), and into adulthood[7]. It is important to minimise the risk of SARS-CoV-2 transmission in children and adolescents to prevent further exacerbation of pre-existing inequalities and safeguard their health, wellbeing and education, alongside reducing wider societal transmission.

Evidence has demonstrated the negative impact of the COVID-19 pandemic on children's health-related behaviours including reduced physical activity, increased sedentary behaviour and poorer nutrition[4,6]. However, it is unclear if this association is bidirectional, that is, whether these health behaviours may also be associated with risk of SARS-CoV-2 infection. Evidence suggests a plausible relationship between health risk behaviours such as

physical inactivity, poor nutrition and inadequate sleep with SARS-CoV-2 infection and severity of disease, attributed to immune system function and cardiometabolic health[8–11]. However, the focus of research to date has been adult populations, exploring single health behaviours or examining those with severe COVID-19 infection and hospitalisation[12,13].

Profiling research within the childhood population has generally centred on identifying the clinical characterisation of infection, with further attention to those with serious infection requiring hospital admission[14,15]. Whilst serious COVID-19 illness in children is relatively rare, mild or asymptomatic infection is common[16]. Furthermore, while there is a rollout of vaccination programmes throughout the adult population in the UK, the vaccination programme for 12–15-year-olds is currently in its early stages. Children below the age of 12 that are not clinically vulnerable are yet to be offered the first dose of vaccination in the UK as of January 2022.

Identifying the prior health-related behavioural characteristics of children and adolescents subsequently requiring a SARS-CoV-2 test or testing positive for SARS-CoV-2 infection could yield insight into the clustering of health behaviours during childhood and adolescence and subsequent infection risk during the current COVID-19 pandemic and future pandemic/endemic scenarios. This can also allow targeted intervention to minimise transmission risk that complements national public health measures and guidelines, and importantly, mitigate the disruption to children's lives. In Wales, one of the four nations of the UK, approaches to performing PCR tests on children include the presence of COVID-19 symptoms, if identified as a close contact to a positive case (e.g. household contacts), or following a positive lateral flow test (e.g. showing symptoms and having a positive lateral flow test performed in the home).

This study investigates the association of prior health-related behaviours self-reported by children aged 8-11 years during primary school before the COVID-19 pandemic between 1 April 2014 and 28 February 2020, with the odds of having a test and testing positive for SARS-CoV-2 during adolescence (aged 10-19). We aim to examine whether these self-reported markers of health-related behaviours during primary school are associated with likelihood of; i) being tested for SARS-CoV-2 (e.g. presence of symptoms) and ii) testing positive for SARS-CoV-2 during the adolescent period (aged 10-19 years), between 1 March 2020 and 31 August 2021.

METHODS

Study design

This study was conducted through the HAPPEN primary school network (*Health and Attainment of Pupils in a Primary Education Network*)[17]. HAPPEN was established in Wales, UK in 2014, following research with headteachers who advocated for increased collaboration to prioritise pupils' health and wellbeing[18,19]. The network brings together primary schools with research and runs up to the current date. School participation in HAPPEN is voluntary and is either once, annually or bi-annually (e.g. to evaluate school-based interventions). Through HAPPEN, children aged 8-11 (years 4 to 6) complete the HAPPEN survey, an online cohort survey that captures a range of validated self-reported health behaviours including physical activity, nutrition and sleep[20]. Retrospective health-related behaviour data were obtained from responses from the HAPPEN survey between 1 April 2014 and 28 February 2020. These retrospective survey responses were linked with polymerase chain reaction (PCR) SARS-CoV-2 test results obtained from the Pathology COVID-19 Daily (PATD) routine dataset

between 1 March 2020 and 31 August 2021. Linkage was performed using the SAIL (*Secure Anonymised Information Linkage*) Databank[21–23]. Data were linked at the individual level using an Anonymous Linkage Field (ALF) to identify participants and link SARS-CoV-2 test results (figure 1). The RECORD checklist[24] for this study is presented in online supplemental appendix 1.

Ethics

Ethical approval was granted by the Swansea University Medical School Research Ethics Committee (2017-0033H). Electronic data (survey responses) were stored in password-protected files only accessible to the research team. The routine data used in this study are available in the SAIL Databank and are subject to review by an independent Information Governance Review Panel (IGRP), to ensure proper and appropriate use of SAIL data. Before any data can be accessed, approval must be received from the IGRP. When access has been approved, it is accessed through a privacy-protecting safe haven and remote access system referred to as the SAIL Gateway. SAIL has established an application process to be followed by anyone who would like to access data. This study has been approved by the SAIL IGRP (project reference: 0911).

The HAPPEN survey and linked SAIL data

Primary schools in Wales, UK were invited to participate in the HAPPEN survey between 1 April 2014 and 28 February 2020 via a number of methods including email, social media promotion and through stakeholders in health and education (including local authority health and wellbeing teams, regional education consortia). Schools were invited to share

details of the survey with parents/guardians (including information sheets). To participate in the HAPPEN survey and link data to routine records, child assent was required in addition to parental consent (between 2014 to 2018) and opt-out parental consent (2019 onwards).

The HAPPEN survey is completed by children aged 8-11 as a self-guided activity within the school setting as a classroom activity with supervision from a teacher/teaching assistant. The survey takes approximately 30 minutes to complete and includes validated self-report measures of typical health behaviours including physical activity, screen time, nutrition, sleep and wellbeing[20]. A full copy of the survey can be found in online supplemental appendix 2 and items, response categories and the coding framework included within analyses in online supplemental appendix 3.

The process of data coding involved two researchers. The first (MJ) cleaned the raw data (including checking for duplicate entries), removed identifiable information and generated a unique participant ID number to protect participants' anonymity. The second (EM) researcher coded the anonymised raw dataset using STATA (version 16) to produce a dataset for analyses. This HAPPEN dataset was uploaded to the SAIL Databank, a trusted research environment (TRE) containing individual-level anonymised population-scale data sources about the population of Wales that enables secure data linkage and analysis for research, to be linked with SARS-CoV-2 testing data from the PATD dataset. To link the data, the person-based identifiable data are separated from the survey data and sent to a trusted third party, Digital Health and Care Wales (the national organisation that designs and builds digital services for health and social care in Wales). The survey data is sent to SAIL using a secure file upload. A unique Anonymous Linking Field (ALF) is assigned to the person-based record before it is joined to clinical data via a system linking field. The ALF was used to link records at the individual level between the HAPPEN dataset and PATD dataset containing PCR

testing data. This dataset was accessible to authors listed from the Population Data Science group, Swansea University.

Quantitative analysis

The primary outcomes were i) whether the child was PCR tested for the SARS-CoV-2 virus and ii) whether the child had a positive SARS-CoV-2 test between 1 March 2020 and 31 August 2021. Eligibility criteria (see cohort flow diagram, Figure 1) within final analyses models were any unique participant with complete linked survey and routine records, aged at least 10 years on 1 March 2020 (start of period of interest) (n=6,891). Inclusion dates of survey responses for analyses were between 1 April 2014 and 28 February 2020. Logistic regression analyses, adjusting for confounding variables (sex, age on 1 March 2020, area-level deprivation using the Welsh Index of Multiple Deprivation (WIMD)[25] (version 2019) and clustered by school (to account for differences between schools), determined Odds Ratios (OR) for i) being PCR tested for SARS-CoV-2 virus (1; tested at least once for SARS-CoV-2 between 1 March 2020 and 31 August 2021, 0; no evidence of SARS-CoV-2 test during period of interest) and ii) having a positive PCR SARS-CoV-2 test (1; testing positive for SARS-CoV-2 between 1 March 2020 and 31 August 2021, 0; testing negative for SARS-CoV-2, 0; no evidence of PCR test).

Independent variables as measures of typical health-related behaviours included within analyses were obtained retrospectively from the HAPPEN survey, completed between 1 April 2014 and 28 February 2020 (online supplemental appendix 3). Health-related behaviour measures included in multivariable analyses related to the behaviours from the previous day (ate breakfast, travel actively to and/or from school, ate at least five portions of fruit/vegetables, twice-daily toothbrushing, slept at least nine hours), behaviours every day

the previous seven days (physically active at least 60 minutes, sedentary/screen time at least two hours, felt tired, ate a sugary snack), participate in at least three out of school clubs, can ride a bike and can swim 25 metres[20]. For the purpose of analyses, survey items with multiple category responses (e.g. active travel to school response categories: walked, on bike, ran/jogged, scooter, skateboard/rollerblade assigned as binary active travel) or continuous numerical values (e.g. out of school clubs assigned as binary value indicating participation in at least three clubs) were assigned binary values. A list of variables included in analyses, coding response categories and a coding framework is presented in online supplemental appendix 3. Independent variables were first entered concurrently and examined for association with outcomes. Then a process of backward stepwise selection was manually followed to build the final regression models. This involved the inclusion of all variables within the initial model, followed by the individual removal of the least significant variables until no nonsignificant variables at the 10% level remained within the model.

Figure 1: Cohort flow diagram

RESULTS

Table 1: Descriptive statistics of study sample by PCR-tested for SARS-CoV-2.

		Tested for SARS-CoV-2 n (%)	**Not tested for SARS-CoV-2 n (%)**
Sample		43.8% (3,021)	56.2% (3,870)
Age at the time HAPPEN survey		10.5 ± 0.6	10.5 ± 0.6
Age on 1 March 2020 (start of period of interest)		12.3 ± 2.0	12.2 ± 1.9
Age on 1 March 2020 (start of period of interest)		12.3 ± 2.0	12.2 ± 1.9
	Early adolescence (10-14 years)	85.1% (2,570)	86.7% (3,357)
	Mid adolescence (15-17 years)	14.9% (451)	13.3% (513)
	Late adolescence (18+ years)	0	0
Sex	**Boy**	48.0% (1,449)	48.4% (1,871)
	Girl	50.9% (1,537)	50.9% (1,970)
	Missing	*1.2% (35)*	*0.8% (29)*
WIMD version 2019 quintiles	**1** (most deprived)	22.3% (675)	23.8% (919)
	2	16.2% (488)	15.9% (614)
	3	16.1% (487)	17.5% (678)
	4	12.5% (378)	11.6% (448)
	5 (least deprived)	21.7% (655)	19.9% (770)
	Missing	*11.2% (338)*	*11.4% (441)*
	Previous day		
Ate breakfast	**Yes**	91.0% (2,750)	90.1% (3,520)
	No	7.5% (227)	7.6% (292)
	Missing	*1.5% (44)*	*1.5% (58)*
Active travel to school	**Yes**	39.5% (1,194)	41.8% (1,617)
	No	58.9% (1,779)	56.7% (2,194)
	Missing	*1.6% (48)*	*1.5% (59)*
Active travel from school	**Yes**	44.7% (1,349)	45.8% (1,771)
	No	53.8% (1,626)	52.7% (2,040)
	Missing	*1.5% (46)*	*1.5% (59)*

Toothbrush ≥ two times	**Yes**	79.3% (2,396)	78.2% (3,025)
	No	19.1% (577)	20.2% (782)
	Missing	*1.6% (48)*	*1.6% (63)*
Consumed ≥ five fruit & veg	**Yes**	23.7% (717)	24.0% (930)
	No	74.8% (2,261)	74.4% (1.6%)
	Missing	*1.4% (43)*	*1.6% (61)*
Sleep ≥ nine hours	**Yes**	78.0% (2,357)	76.6% (2,963)
	No	19.8% (599)	21.4% (829)
	Missing	*2.2% (65)*	*2.0% (78)*
Previous 7 days			
Physically active ≥60 mins every day	**Yes**	21.2% (640)	20.8% (805)
	No	77.3% (2,335)	77.7% (3,008)
	Missing	*1.5% (46)*	*1.5% (57)*
Sedentary/screen time ≥2 hours every day	**Yes**	30.5% (922)	32.4% (1,252)
	No	68.0% (2,053)	66.2% (2,561)
	Missing	*1.5% (46)*	*1.5% (57)*
Tired every day	**Yes**	15.2% (460)	15.1% (584)
	No	83.3% (2,515)	83.4% (3,229)
	Missing	*1.5% (46)*	*1.5% (57)*
Sugary snack every day	**Yes**	18.0% (545)	18.4% (710)
	No	80.4% (2,430)	80.2% (3,103)
	Missing	*1.5% (46)*	*1.5% (57)*
General			
Participate in ≥ three out of school clubs per week	**Yes**	35.9% (1,083)	32.0% (1,239)
	No	57.2% (1,729)	61.6% (2,382)
	Missing	*6.9% (209)*	*6.4% (249)*
Can ride a bike	**Yes**	89.7% (2,711)	87.7% (3,394)
	No	8.5% (257)	10.6% (409)
	Missing	*1.8% (53)*	*1.7% (67)*
Can swim 25m	**Yes**	81.0% (2,446)	76.4% (2,955)
	No	17.3% (523)	22.0% (853)
	Missing	*1.7% (52)*	*1.6% (62)*

See online supplemental appendix 3 for variable codebook.

Table 2: Descriptive statistics of study sample by PCR-tested positive for SARS-CoV-2.

		Tested positive for SARS-CoV-2 n (%)	**No evidence of positive SARS-CoV-2 test n (%)**
Sample		11.2% (774)	88.8% (6,117)
Age at the time HAPPEN survey		10.5 ± 0.7	10.5 ± 0.6
Age on 1 March 2020 (start of period of interest)		12.8 ± 2.1	12.2 ± 1.9
Sex	**Boy**	43.9% (340)	48.7% (2,980)
	Girl	55.6% (430)	50.3% (3,077)
	Missing	*0.5% (<5)*	*1.0% (60)*
WIMD version 2019 quintiles	**1** (most deprived)	25.7% (199)	22.8% (1,395)
	2	13.7% (106)	16.3% (996)
	3	18.7% (145)	16.7% (1,020)
	4	9.8% (76)	12.3% (750)
	5 (least deprived)	22.0% (170)	20.5% (1,255)
	Missing	*10.1% (78)*	*11.5% (701)*
	Previous day		
Ate breakfast	**Yes**	91.5% (708)	90.9% (5,562)
	No	6.7% (52)	7.6% (467)
	Missing	*1.8% (14)*	*1.4% (88)*
Active travel to school	**Yes**	37.9% (293)	41.2% (2,518)
	No	60.2% (466)	57.3% (3,07)
	Missing	*1.9% (15)*	*1.5% (92)*
Active travel from school	**Yes**	43.7% (338)	45.5% (2,782)
	No	54.3% (420)	53.1% (3,246)
	Missing	*2.1% (16)*	*1.5% (89)*
Toothbrush ≥ two times	**Yes**	80.5% (623)	78.4% (4,798)
	No	17.6% (136)	20.0% (1,223)
	Missing	*1.9% (15)*	*1.6% (96)*
Consumed ≥ five fruit & veg	**Yes**	26.2% (203)	23.6% (1,444)
	No	72.0% (557)	74.9% (4,583)
	Missing	*1.8% (14)*	*1.5% (90)*
Sleep ≥ nine hours	**Yes**	78.4% (607)	77.1% (4,713)
	No	18.6% (144)	21.0 (1,284)
	Missing	*3.0% (23)*	*2.0% (120)*
	Previous 7 days		
Physically active ≥60 mins every day	**Yes**	21.8% (169)	20.9% (1,276)
	No	76.1% (589)	77.7% (4,754)
	Missing	*2.1% (16)*	*1.4% (87)*
	Yes	30.5% (236)	31.7% (1,938)

Sedentary/screen time ≥2 hours every day	**No**	67.4% (522)	66.9% (4,092)
	Missing	*2.1% (16)*	*1.4% (87)*
Tired every day	**Yes**	12.4% (96)	15.5% (948)
	No	85.5% (662)	83.1% (5,082)
	Missing	*2.1% (16)*	*1.4% (87)*
Sugary snack every day	**Yes**	17.3% (134)	18.3% (1,121)
	No	80.6% (624)	80.3% (4,909)
	Missing	*2.1% (16)*	*1.4% (87)*
	General		
Participate in ≥ three out of school clubs per week	**Yes**	39.0% (302)	33.0% (2,020)
	No	53.1% (411)	60.5% (3,700)
	Missing	*7.9% (61)*	*6.5% (397)*
Can ride a bike	**Yes**	90.8% (703)	88.3% (5,402)
	No	6.9% (53)	10.0% (613)
	Missing	*2.3% (18)*	*1.7% (102)*
Can swim 25m	**Yes**	82.2% (636)	77.9% (4,765)
	No	15.5% (120)	20.5% (1,256)
	Missing	*2.3% (18)*	*1.6% (96)*

See online supplemental appendix 3 for variable codebook.

Survey responses were obtained from n=11,339 participants (figure 1), of which n=1,101 were duplicate participants (occasions of multiple survey entries). In the case of duplicates, the first instance of survey participation was used and later responses excluded to create a dataset consisting of unique participants (n=10,238). Inclusion dates of survey responses for analyses were between 1 April 2014 and 28 February 2020. Participants with survey responses outside the period of interest were excluded (n=248 excluded, March 2020). Age criteria within final analyses was aged at least 10 years on 1 March 2020 (n=3,099 excluded, aged <10 years). Complete linked unique records of participants meeting eligibility criteria were obtained for n=6,891 individuals. Of the total sample, 43.2% (n=3,021) were PCR-tested for SARS-CoV-2 (table 1) and 11.2% (n=774) tested positive for SARS-CoV-2 (table 2). The mean age on 1 March 2020 (start of period of interest) was 12.3 (±2.0) for those PCR-tested (table 1) and 12.8 (±2.1) for those who tested positive for SARS-CoV-2 (table 2). As the adolescence period spans a range of ages, we have presented a breakdown of distribution of

developmental stage on 1 March 2020 for testing for SARS-CoV-2 and testing positive for SARS-CoV-2 in table 1 and table 2 respectively. Of the total sample, 86% (n=5,927) were aged between 10-14 (early adolescence) 14% aged 15-17 (middle adolescence) and 0% in late adolescence. Complete case analyses are presented below.

Table 3: Backward stepwise logistic regression model of significant (p<0.1) health behaviour markers and probability of being PCR-tested for SARS-CoV-2 accounting for baseline age, sex and deprivation, and clustered by school.

PCR-tested for SARS-CoV-2 (0; no, 1; yes) ***Number of obs=5,581*** ***R^2=0.006***	**OR**	**P**	**95% CI**
Sleep ≥ nine hours	1.15	0.028	1.01 to 1.29
Participate in ≥ three out of school clubs per week	1.15	0.023	1.02 to 1.31
Can ride a bike	1.16	0.095	0.98 to 1.37
Can swim 25m	1.29	0.002	1.10 to 1.52
Age on 1 March 2020 (start of period of interest)	1.02	0.328	0.98 to 1.05
Sex (girl)	0.97	0.591	0.86 to 1.09
WIMD quintile 2	1.10	0.303	0.92 to 1.31
3	0.96	0.626	0.80 to 1.14
4	1.10	0.394	0.89 to 1.36
5 (least deprived)	1.04	0.577	0.90 to 1.22

OR: Odds Ratio; 95% CI: 95% confidence intervals. See online supplemental appendix 3 for variable codebook. Low correlation between variables (coefficients -0.10 to 0.21).

Children reporting to sleep at least nine hours (OR=1.15, 95% CI 1.01 to 1.29), participate in at least three out of school clubs (OR=1.15, 95% CI 1.02 to 1.31), able to ride a bike (10% significance level) and able to swim 25 metres (OR=1.29, 95% CI 1.10 to 1.52) showed an increased likelihood of being tested for SARS-CoV-2 (table 3). The model showed a low goodness-of-fit (R^2=0.006). See online supplemental appendix 4 for multivariable logistic regression model of health behaviour markers and probability of being PCR-testing for SARS-CoV-2.

Table 4: Backward stepwise logistic regression model of significant (p<0.1) health behaviour markers and probability PCR-testing positive for SARS-CoV-2 accounting for baseline age, sex and deprivation, and clustered by school.

PCR-test positive for SARS-CoV-2 (0; no, 1; yes) ***Number of obs=5,616*** ***R^2=0.02***	**OR**	**P**	**95% CI**
Tired every day previous week	0.78	0.032	0.61 to 0.98
Participate in ≥ three out of school clubs per week	1.26	0.031	1.02 to 1.56
Can ride a bike	1.36	0.079	0.97 to 1.92
Age on 1 March 2020 (start of period of interest)	1.16	0.000	1.10 to 1.22
Sex (girl)	1.25	0.005	1.06 to 1.47
WIMD quintile 2	0.80	0.084	0.62 to 1.03
3	1.05	0.686	0.82 to 1.34
4	0.78	0.093	0.58 to 1.04
5 (least deprived)	0.86	0.193	0.68 to 1.08

OR: Odds Ratio; 95% CI: 95% confidence intervals. See online supplemental appendix 3 for variable codebook. Low correlation between variables (coefficients -0.08 to 0.15).

Children who participated in at least three out of school clubs (OR=1.12, 95% CI 1.02 to 1.56) and being able to ride a bike (OR=1.36, 95% CI 0.97 to 1.92, p<0.1) were associated with an increased likelihood of testing positive for SARS-CoV-2, whilst reporting to feel tired (OR=0.78, 95% CI 0.61 to 0.98) was associated with a reduced likelihood (table 4). Age on 1 March 2020 (OR=1.16, 95% CI 1.10 to 1.22) and sex (girl; OR=1.25, 95% CI 1.06 to 1.47) was associated with an increased likelihood of testing positive (table 4). Those in WIMD quintile 2 (OR=0.80, 95% CI 0.62 to 1.03) and quintile 4 (OR=0.78, 95% C 0.58 to 1.04) were less likely (p<0.1) to test positive compared to quintile 1 (most deprived) (table 4). The model showed a low goodness-of-fit (R^2=0.02). There was very low correlation between independent variables in backward stepwise regression models (table 3 and table 4) with coefficient values ranging from -0.10 to 0.21 and -0.08 to 0.15 respectively. See online supplemental appendix 5 for

multivariable logistic regression model of health behaviour markers and probability of PCR-testing positive for SARS-CoV-2.

DISCUSSION

This study aims to examine whether markers of health-related behaviours reported by children during primary school between 1 April 2014 and 28 February 2020 is associated with the likelihood of being PCR-tested for SARS-CoV-2 (e.g. presence of symptoms) and testing positive between 1 March 2020 and 31 August 2021 during adolescence (10-19 years). This study did not find evidence that reporting positive health-related behaviours is associated with a reduced odds of being tested or testing positive for SARS-CoV-2. Findings suggest that reporting the recommended level of sleep (at least nine hours), participating in at least three out of school clubs, being able to ride a bike and being able to swim 25 metres were associated with an increased likelihood of being tested for SARS-CoV-2. Participating in at least three out of school clubs and being able to ride a were associated with an increased likelihood of testing positive for SARS-CoV-2, whilst reporting to feel tired every day was associated with a reduced likelihood of testing positive for SARS-CoV-2. Girls and older age were was associated with increased likelihood of testing positive for SARS-CoV-2. Those living in WIMD quintiles 2 and 4 were less likely to test positive compared to those living in the most deprived quintile (quintile 1) (10% significance level).

Detecting positive SARS-CoV-2 cases through testing and adhering to self-isolation is an important strategy in reducing community transmission[26]. The majority of children in this study (86%) were in the early adolescence stage (10-14) at the baseline date, with the remaining 14% in middle adolescence (15-17 years). The detection of child positive cases using routine PCR testing data in this study requires a parent/guardian to take the child for

testing and thus relies on parental/caregiver influence and involvement. We find associations between child-reported health-related behaviours with both PCR-testing for SARS-CoV-2 and testing positive for SARS-CoV-2. We theorise that parents who have higher levels of health literacy are more likely to take their child for a SARS-CoV-2 test and are then more likely to be detected as positive. Parenting is an important contributor to promoting positive health behaviours in children, and is represented by a constellation of attitudes, behaviours and values for the child. Indeed, monitoring behaviours occur and our study suggests that these actions associated with a parent taking a child for a SARS-CoV-2 test represent parental health literacy, for example through ensuring the child has a sleep routine[27]. The clustering of physically active behaviours represented by the association of being able to swim and ride a bike may represent underlying parental involvement and modelling behaviour, including involvement in leisure time activities, providing financial and transport provision to attend organised activities such as access to swimming lessons and the provision of equipment[28]. This may also have a socioeconomic component, building on the ideas of Bourdieu in terms of social capital, and access to classes and health enhancing material items[29].

The detection of positive child cases also relies on parents recognising symptoms, knowledge of how to access testing services, ability to access services (e.g. transport) and willingness to provide personal information for test and trace services. This is likely influenced by parental health literacy, recognised as the ability to access, understand, interpret and apply medical information and make informed decisions regarding medical advice, issues or guidelines[30]. Parental health literacy impacts the decision a parent makes relating to their child[31] and is correlated with a number of health indicators including knowledge of health and health services, and the parent and child engaging in health-promoting behaviours[30,32]. Therefore, these findings suggest that the tracking and reporting SARS-

CoV-2 in children and adolescent may be highly skewed towards children whose parents who are health literate, and those not accessing testing services remain undetected.

In addition, it will also be important to consider the mechanism of parental involvement and health literacy in the context of child vaccination. The COVID-19 vaccine has been approved to children aged 12 and above in the UK, with trials currently underway to examine the vaccine response within children aged 5-11 (approved in for example the USA and Israel)[33]. High population-level vaccine uptake is a primary strategy for many governments globally. Thus, the findings in this study suggest future investment should be made to address parental health literacy and its influence on child COVID-19 vaccine uptake. This research has implications for informing public health practice and emerging policy by integrating the views of parents, children and young people to the design of testing services and future vaccination programmes.

Previous research examining transmission in school-aged children found SARS-CoV-2 infections within the household to be the strongest predictor for a subsequent positive SARS-CoV-2 test[34]. Findings in this study that girls are more likely to test positive for SARS-CoV-2 may suggest sex differences between household contact patterns including more repeated, extended contact with household members by girls. This is supported by a study in the United States that finds sex differences between patterns of social interaction in the home, with adolescent females (15 to 19 years) having higher mean number of household member contacts and mean total contact duration[35]. Previous research also demonstrates assortative mixing patterns by age, with the highest frequency of contacts by those aged 10-19 with individuals of the same age group, though this is not stratified by sex[36]. Further research is required in the context of COVID-19 to examine sex differences of adolescent contact patterns in areas of high frequency such as the school setting.

Regarding the association of increasing age with likelihood of testing positive for SARS-CoV-2 in this study, a systematic review and meta-analysis by Viner and colleagues determined that whilst children younger than 10 to 14 years have lower susceptibility to SARS-CoV-2 transmission, susceptibility by adolescents may be similar to that of adults[3]. Our findings to do not show an area-level social gradient. Those in WIMD quintile 2 and 4 were less likely (10% significance level) to test positive for SARS-CoV-2 compared to the most deprived quintile. Whilst it is possible that children mixing in the school setting are in contact with children residing in different area-level quintiles, this finding may reflect community prevalence which was not captured in the current study.

STRENGTHS AND LIMITATIONS

Strengths of this study include the reporting of multiple child health behaviours, and the use of individual-level linked routine records of SARS-CoV-2 testing data through the SAIL Databank. This is the first study to investigate association of prior child health-related behaviour measures with subsequent SARS-CoV-2 testing and infection during adolescence in Wales and highlights the importance of targeting and improving parental health literacy to increase uptake of appropriate testing and vaccinations among children. The child-reported health behaviours captured in this study were measured before the COVID-19 pandemic and represent historical health behaviour which may not reflect the child's lifestyle behaviours during COVID-19. It is possible that these behaviours were disrupted due to public health measures in place to reduce transmission. Health behaviours captured through the national-scale HAPPEN survey represent children attending schools that engaged with the HAPPEN Wales primary school network. This may not be representative of the whole population of

Wales. The period of study for PCR-testing for and testing positive for SARS-CoV-2 includes a time frame with varying approaches to testing children, public health measures and restrictions such as prolonged school closures, schools reopening and other measures to minimise contacts which were not measured in this study. As a result, contacts with infectious people and risk of transmission are likely to have varied. Rates of SARS-CoV-2 transmission also varied during this study, and this study does not differentiate between variants with higher or lower transmissibility, or vaccination status.

CONCLUSION

This study did not find evidence that reporting health-promoting behaviours (e.g. fruit and vegetable consumption, regular physical activity or meeting sleep guidelines) prior to the COVID-19 pandemic reduced the likelihood of having symptoms of COVID-19 (measured as being PCR-tested for SARS-CoV-2) or having a positive test. Instead, this study suggests that actions associated with a child being tested for SARS-CoV-2 and being identified as positive related to health-promoting behaviours may be a proxy of parental health literacy and monitoring behaviours. Further research is required to examine parental health literacy and monitoring behaviours in the context of testing for SARS-CoV-2. Adhering to public health guidance, social distancing, reducing number of contacts and having the vaccine remain the primary means of minimising infection risk. The first vaccine doses are currently being offered to children aged 12-15. Children below the age of 12 are not currently offered first doses. In order to minimise the widespread disruption to children's lives through COVID-19 infection, expanding the vaccination programme to primary school-aged children, community testing, accompanied by following public health guidance reflective of current community

transmission rates is important. Based on the proposed theory in this study of the influence parental health literacy on uptake of SARS-CoV-2 testing and the detection of positive cases, it is also important to consider this in the context of vaccinating children. This study suggests that we do not know the true rates of COVID-19 in children and adolescents as they are dependent on their parent taking them for testing and this may be determined by the parent's health literacy and understanding.

Acknowledgements

The authors would like to thank all participating primary schools and pupils that took part in this study. This work was supported by the National Centre for Population Health and Wellbeing Research through the HAPPEN network. This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who make anonymised data available for research. We would also like to thank Dr Annemarie Docherty and Dr Olivia Swann from The University of Edinburgh for providing informal peer review input to the final draft.

Availability for data and materials

The routine data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK. All proposals to use SAIL data are subject to review by an IGRP. Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been approved, it is gained through a privacy-protecting safe haven and remote access system referred to as the SAIL Gateway. SAIL has established an application process to be followed

by anyone who would like to access data via SAIL https://www.saildatabank.com/application-process. This study has been approved by the IGRP as project 0911.

Funding

The Economic and Social Research Council (ESRC) funded the development of the HAPPEN network (grant number: ES/J500197/1) which this research was conducted through. The National Centre for Population Health and Wellbeing Research (NCPHWR) funded by Health and Care Research Wales provided infrastructural support for this work. This work was supported by the Con-COV team funded by the Medical Research Council (grant number: MR/V028367/1). This work was supported by Health Data Research UK, which receives its funding from HDR UK Ltd (HDR-9006) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and the Wellcome Trust. This work was a collaboration with the ADR Wales programme of work. ADR Wales is part of the Economic and Social Research Council (part of UK Research and Innovation) funded ADR UK (grant ES/S007393/1). This work was supported by the Wales COVID-19 Evidence Centre, funded by Health and Care Research Wales and by the COVID-19 Longitudinal Health and Wellbeing National Core Study funded by the Medical Research Council (MC_PC_20030).

Transparency statement

The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned have been explained.

Competing interests

The authors declare that they have no competing interests.

Patient and public involvement

The SAIL Databank has a Consumer Panel that provides the public's perspective on data linkage research. The Panel members are involved in all elements of the SAIL Databank process, from developing ideas, advising on bids through approval processes (via the independent Information Governance Review Panel), to disseminating research findings. For more information visit https://saildatabank.com/about-us/public-engagement/.

Contributorship statement

EM and SB conceptualised the study design. EM and JK curated the data. EM performed the statistical analysis, undertook the initial interpretation of the data and was responsible for the original draft. EL and SB contributed to the writing of the manuscript. EL, LC, JK, RL and SB provided critical interpretation of the data. The manuscript was critically reviewed and edited by all authors. EM is the guarantor. AD, RL and OS critically reviewed the final manuscript..

Figure 1: Cohort flow diagram

References

1 UNICEF. The United Nations Convention on the Rights of the Child. 2004. doi:10.1093/tropej/fmp024

2 OECD. Combatting COVID-19's effect on children - OECD. 2020.https://www.oecd.org/coronavirus/policy-responses/combatting-covid-19-s-effect-on-children-2e1f3b2f/ (accessed 1 Oct 2021).

3 Viner RM, Mytton OT, Bonell C, *et al.* Susceptibility to SARS-CoV-2 Infection among Children and Adolescents Compared with Adults: A Systematic Review and Meta-analysis. JAMA Pediatr. 2020. doi:10.1001/jamapediatrics.2020.4573

4 James M, Marchant E, Defeyter MA, *et al.* Impact of school closures on the health and well-being of primary school children in Wales UK: a routine data linkage study using the HAPPEN Survey (2018-2020). *BMJ Open* 2021;**11**. doi:10.1136/BMJOPEN-2021-051574

5 Marchant E, Todd C, James M, *et al.* Primary school staff perspectives of school closures due to COVID-19, experiences of schools reopening and recommendations for the future: A qualitative survey in Wales. *PLoS One* 2021;**16**:e0260396. doi:10.1371/JOURNAL.PONE.0260396

6 Moore SA, Faulkner G, Rhodes RE, *et al.* Impact of the COVID-19 virus outbreak on movement and play behaviours of Canadian children and youth: A national survey. *Int J Behav Nutr Phys Act* 2020;**17**. doi:10.1186/s12966-020-00987-8

7 Craigie AM, Lake AA, Kelly SA, *et al.* Tracking of obesity-related behaviours from childhood to adulthood: A systematic review. *Maturitas* 2011;**70**:266–84.

doi:10.1016/j.maturitas.2011.08.005

8 Chowdhury R, Van Daalen KR, Franco OH. Cardiometabolic Health: Key in Reducing Adverse COVID-19 Outcomes. *Glob Heart* 2020;**15**:58. doi:10.5334/gh.879

9 Koch R, Jordan S, Starker A, *et al.* Health behaviour and COVID-19: Initial findings on the pandemic Journal of Health Monitoring 2. *J Heal Monit* 2020;**2020**:1–14. doi:10.25646/7055

10 Childs CE, Calder PC, Miles EA. Diet and Immune Function. *Nutrients* 2019;**11**. doi:10.3390/nu11081933

11 Lange KW, Nakamura Y. Lifestyle factors in the prevention of COVID-19. *Glob Heal J* 2020;**4**:146–52. doi:10.1016/J.GLOHJ.2020.11.002

12 Hamer M, Kivimäki M, Gale C, *et al.* Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: A community-based cohort study of 387,109 adults in UK. *Brain Behav Immun* 2020;**87**:184–7. doi:10.1016/J.BBI.2020.05.059

13 Lee SW, Lee J, Moon SY, *et al.* Physical activity and the risk of SARS-CoV-2 infection, severe COVID-19 illness and COVID-19 related mortality in South Korea: a nationwide cohort study. *Br J Sports Med* 2021;**0**:18. doi:10.1136/BJSPORTS-2021-104203

14 Zhou B, Yuan Y, Wang S, *et al.* Risk profiles of severe illness in children with COVID-19: a meta-analysis of individual patients. *Pediatr Res 2021 902* 2021;**90**:347–52. doi:10.1038/s41390-021-01429-2

15 Shekerdemian LS, Mahmood NR, Wolfe KK, *et al.* Characteristics and Outcomes of Children With Coronavirus Disease 2019 (COVID-19) Infection Admitted to US and Canadian Pediatric Intensive Care Units. *JAMA Pediatr* 2020;**174**:868–73. doi:10.1001/JAMAPEDIATRICS.2020.1948

16 Castagnoli R, Votto M, Licari A, *et al.* Severe Acute Respiratory Syndrome Coronavirus

2 (SARS-CoV-2) Infection in Children and Adolescents: A Systematic Review. *JAMA Pediatr* 2020;**174**:882–9. doi:10.1001/JAMAPEDIATRICS.2020.1467

17 Todd, C., Christian, D., Tyler, R., Stratton, G., Brophy S. Developing HAPPEN (Health and Attainment of Pupils involved in a Primary Education Network): working in partnership to improve child health and education. *Perspect Public Health* 2016;**136**:115–6.

18 Christian D, Todd C, Davies H, *et al.* Community led active schools programme (CLASP) exploring the implementation of health interventions in primary schools: headteachers' perspectives. *BMC Public Health* 2015;**15**:238. doi:10.1186/s12889-015-1557-0

19 Todd C, Christian D, Davies H, *et al.* Headteachers' prior beliefs on child health and their engagement in school based health interventions: a qualitative study. *BMC Res Notes* 2015;**18**:161. doi:10.1186/s13104-015-1091-2

20 Everson B, Mackintosh KA, McNarry MA, *et al.* Can Wearable Cameras be Used to Validate School-Aged Children's Lifestyle Behaviours? *Children* 2019;**6**:20. doi:10.3390/children6020020

21 Ford D V., Jones KH, Verplancke JP, *et al.* The SAIL Databank: Building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;**9**. doi:10.1186/1472-6963-9-157

22 Jones KH, Ford DV, Thompson S, *et al.* A Profile of the SAIL Databank on the UK Secure Research Platform. *Int J Popul Data Sci* 2019;**4**. doi:10.23889/ijpds.v4i2.1134

23 Lyons J, Akbari A, Torabi F, *et al.* Understanding and responding to COVID-19 in Wales: protocol for a privacy-protecting data platform for enhanced epidemiology and evaluation of interventions. *BMJ Open* 2020;**10**:43010. doi:10.1136/bmjopen-2020-043010

24 Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Med* 2015;**12**:e1001885. doi:10.1371/JOURNAL.PMED.1001885

25 Welsh Government. Welsh Index of Multiple Deprivation. 2019.https://gov.wales/welsh-index-multiple-deprivation-full-index-update-ranks-2019

26 Mercer TR, Salit M. Testing at scale during the COVID-19 pandemic. *Nat Rev Genet 2021 227* 2021;**22**:415–26. doi:10.1038/s41576-021-00360-w

27 Pyper E, Harrington D, Manson H. Do parents' support behaviours predict whether or not their children get sufficient sleep? A cross-sectional study. *BMC Public Health* 2017;**17**. doi:10.1186/s12889-017-4334-4

28 Edwardson CL, Gorely T. Parental influences on different types and intensities of physical activity in youth: A systematic review. *Psychol Sport Exerc* 2010;**11**:522–35. doi:10.1016/j.psychsport.2010.05.001

29 Bourdieu P. The Forms of Capital. In: *Handbook of Theory and Research for the Sociology of Education*. New York: : Greenwood 1986. 241–58.

30 Sørensen K, Van den Broucke S, Fullam J, *et al.* Health literacy and public health: A systematic review and integration of definitions and models. Published Online First: 2012. doi:10.1186/1471-2458-12-80

31 Morrison AK, Glick A, Yin HS. Health Literacy: Implications for Child Health. *Pediatr Rev* 2019;**40**:263–77. doi:10.1542/PIR.2018-0027

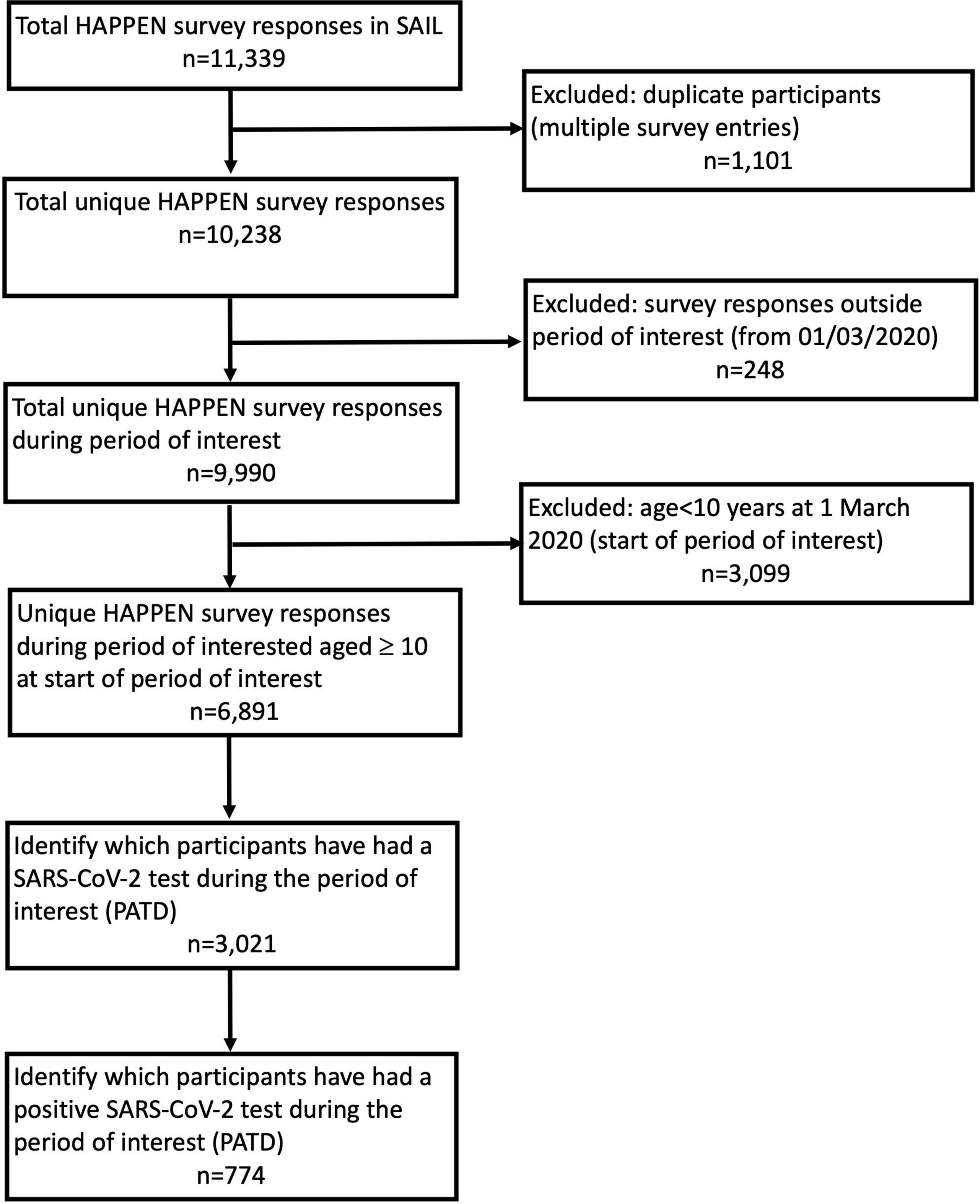
32 Pawellek M, Kopf FM, Egger N, *et al.* Pathways linking parental health literacy with health behaviours directed at the child: a scoping review. *Health Promot Int* 2021;**2021**:1–29. doi:10.1093/HEAPRO/DAAB154

33 Pfizer. Pfizer and BioNTech Announce Positive Topline Results From Pivotal Trial of COVID-19 Vaccine in Children 5 to 11 Years | Pfizer. 2021.https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-positive-topline-results (accessed 10 Dec 2021).

34 Thompson DA, Abbasizanjani H, Fry R, *et al.* Staff-pupil SARS-CoV-2 infection pathways in schools in Wales: A population-level linked data approach. *BMJ Paediatr Open* 2021;**5**:1049. doi:10.1136/bmjpo-2021-001049

35 Dorélien AM, Ramen A, Swanson I, *et al.* Analyzing the demographic, spatial, and temporal factors influencing social contact patterns in U.S. and implications for infectious disease spread. *BMC Infect Dis* 2021;**21**:1–19. doi:10.1186/S12879-021-06610-W/FIGURES/10

36 Mossong J, Hens N, Jit M, *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* Published Online First: 2008. doi:10.1371/journal.pmed.0050074


Total HAPPEN survey responses in SAIL
n=11,339
Excluded: duplicate participants (multiple survey entries)
n=1,101
Total unique HAPPEN survey responses
n=10,238
Excluded: survey responses outside period of interest (from 01/03/2020)
n=248
Total unique HAPPEN survey responses during period of interest
n=9,990
Excluded: age<10 years at 1 March 2020 (start of period of interest)
n=3,099
Unique HAPPEN survey responses during period of interested aged ≥ 10 at start of period of interest
n=6,891
Identify which participants have had a SARS-CoV-2 test during the period of interest (PATD)
n=3,021
Identify which participants have had a positive SARS-CoV-2 test during the period of interest (PATD)
n=774