**Online Supplementary Material**

**This document provides additional methodological detail in relation to the paper: Self-reported smoking, alcohol and drug use among adolescents and young adults with and without mild/moderate intellectual disability**

**Detailed Methodology**

Secondary analysis was conducted of data collected in Waves 1 to 7 of *Next Steps* (formerly known as the Longitudinal Study of Young People in England), an annual panel study that followed a cohort from early adolescence into adulthood. It has collected information about education, employment, economic circumstances, family life, physical and emotional health and wellbeing, social participation and attitudes. *Next Steps* data has been linked to the Department for Education’s National Pupil Database (NPD). *Next Steps* data files and documentation were obtained from the UK Data Service. Full details of the method and design of *Next Steps* are available in a series of user guides (Department for Education, 2011b). Key aspects are summarised below.

**Sampling**

Fieldwork commenced in 2004 when the sampled children were aged 13-14 (school year 9). The initial (Wave 1) sample was drawn from a sampling frame based on children attending maintained schools, independent schools and pupil referral units in England who in February 2004 were in Year 9 (or equivalent) and were born between 1 September 1989 and 31 August 1990. Schools in deprived areas and students from minority ethnic groups were oversampled. At Wave 1, 73% of selected schools participated leading to an issued sample of approximately 21,000 young people. The attained sample at W1 was 15,770 children (75% response rate). This cohort was followed-up every year until 2010 (age 19-20).

## Rate and Predictors of Sample Retention

Retention rates over time are presented in Table 1 for participants with/without intellectual disability.

|  |
| --- |
| Table 1: Retention rates for participants with/without intellectual disability  |
|  Wave | With intellectual disabilities | Without intellectual disabilities |
| N | Unweighted prevalence | % retention from W1 | % retention from previous wave | N  | % retention from W1 | % retention from previous wave |
| W1 | 527 | 3.5% |  |  | 14,687 |  |  |
| W2 | 415 | 3.2% | 79% | 79% | 12,654 | 86% | 86% |
| W3 | 354 | 2.9% | 67% | 85% | 11,649 | 79% | 92% |
| W4 | 314 | 2.8% | 60% | 89% | 10,721 | 73% | 92% |
| W5 | 256 | 2.6% | 49% | 82% | 9,551 | 65% | 89% |
| W6 | 241 | 2.6% | 46% | 94% | 8,944 | 61% | 94% |
| W7 | 206 | 2.5% | 39% | 85% | 7,941 | 54% | 89% |

Socio-demographic factors associated with sample attrition between Waves 1 and 7 were examined separately for participants with and without intellectual disability. Predictors of attrition were broadly similar for participants with and without intellectual disability, with male gender, membership of a minority ethnic group, household poverty (defined by FSM eligibility) and higher neighbourhood deprivation all being associated with higher rates of attrition (Groves, 2006; Groves & Couper, 1998). For the variables ‘males’ and ‘not White British’ the point estimate for attrition in the non-intellectual disability group lay (males 1.10 (1.07-1.15); not White British 1.17 (1.13-1.22)) within the 95% CI of the intellectual disability group (males 1.28 (1.09-1.51)); not White British 1.12 (0.97-1.29)). For the variables ‘FSM eligibility’ and ‘high neighbourhood deprivation’ the point estimate for non-participation in the non-intellectual disability group (FSM eligibility 1.37 (1.32-1.43); high neighbourhood deprivation 1.38 (1.30-1.40)) was greater than the upper 95% CI of the intellectual disability (FSM eligibility 1.15 (1.00-1.33); high neighbourhood deprivation 1.09 (0.94-1.26)).

**Identification of Participants with Mild/Moderate Intellectual Disability**

Data linkage with the 2004 and 2006 NPD was undertaken to identify participants with Special Educational Needs (SEN). Linkage was successful for 15,240 young people present at Wave 1 (97% of the *Next Steps* sample). Linkage included data on stage of assessment and primary/secondary category of SEN.

Following the example of previous studies (Emerson & Halpin, 2013; Naylor, Dawson, Emerson, & Tantam, 2011), we used the SEN category of Moderate Learning Difficulty (MLD), if the child was at the School Action Plus stage of assessment of SEN or had a formal Statement of SEN, as an indicator of mild/moderate intellectual disability. School Action Plus and Statements required the involvement of professionals external to the school in the categorisation of SEN. Current guidance defines MLD in relation to pupils having ‘attainments significantly below expected levels in most areas of the curriculum despite appropriate interventions [and having] ... much greater difficulty than their peers in acquiring basic literacy and numeracy skills and in understanding concepts’ (Department for Education, 2011a).

Of the children sampled, 527 (3.5% of the unweighted linked sample) were identified as having mild/moderate intellectual disabilities in either 2004 or 2006. Consistent with existing epidemiological research, the prevalence of intellectual disability was significantly higher among boys than girls (4.3% vs 2.5%; prevalence ratio (PR)=1.75 (95% CI 1.46-2.09)) and among children eligible for free school meals (FSM), an indicator of household poverty, (8.0% vs 1.9%; PR=4.10 (95% CI 3.14-5.35)) (Emerson, 2012; Maulik, Mascarenhas, Mathers, Dua, & Saxena, 2011; Roeleveld, Zielhuis, & Gabreels, 1997).

**Procedure**

Data in the first four waves was collected by face to face interviews using computer assisted personal interviewing with the young person themselves and their parents. Waves 5-7 used a mixed mode approach in which information, which was only collected from the young person, was collected by their choice of method (online, telephone or face to face).

**Measures**

**Smoking.** At Waves 1-3 participants were asked: *Do you ever smoke cigarettes at all*? If answering in the affirmative they were then asked: *Now read all the following statements carefully and type in the number next to the one which best describes you* *(response options: I have never smoked, I have only ever tried smoking once, I used to smoke sometimes but I never smoke a cigarette now, I sometimes smoke cigarettes now but I don't smoke as many as one a week, I usually smoke between one and six cigarettes a week, I usually smoke more than six cigarettes a week).* We created two binary variables from these data: ever smoked at any Wave; and has smoked more than six cigarettes a week at any Wave.

**Alcohol Use.** At Waves 1-4, 6 and 7 participants were asked: *Have you ever had a proper alcoholic drink?* *That is a whole drink, not just a sip. Please do not count drinks labelled low alcohol*. If answering in the affirmative they were then asked: *Thinking about the last 12 months, about how often did you usually have an alcoholic drink? Was it... (response options: most days, once or twice a week, 2 or 3 times a month, once a month, once every couple of months or; less often)?* We created three binary variables from these data: ever had an alcoholic drink under 18, regular drinker under 18 (response options *once or twice a week* or *most days*) at Wave 1-4, regular drinker age 18+ (response options *once or twice a week* or *most days*) at Wave 6-7.

At Waves 6 and 7 participants were also asked: *On those days when you did have an alcoholic drink, how often would you say you got drunk?* *Would you say it was… (*response options: *every time, most times, around half the time, less than half the time, rarely, never)*. We created two binary variables from these data: usually gets drunk (response options *around half the time* or more frequently); and regular drinker age 18+ who usually gets drunk (combining regular drinker age 18+ AND usually gets drunk).

**Drug Use.** At Waves 1-4 participants were asked: *the next question is about Cannabis, also called Marijuana, Dope, Pot, Blow, Hash, Skunk, Puff, Grass, Draw, Ganja, Spliff, Joints, Smoke, Weed. Remember your name is not on this questionnaire so no-one who knows you will find out your answer. Have you ever tried Cannabis even if only once?* We created a binary variable from these data: ever tried cannabis under 18 at Wave 1-4.

At Waves 6 and 7 participants were asked: *have you ever taken any of the following; 1. Cannabis; 2. other drugs like cocaine, LSD, ecstasy, heroin, crack, speed etc.?* We created two binary variables from these data: ever tried cannabis 18+ (at either Wave 6 or 7); and ever tried other drugs 18+ (at either Wave 6 or 7).

At Wave 6 participants were also asked: *in the last 4 weeks how often have you taken cannabis (response options not at all, once, twice, three or more times)?* We created a binary variable from these data: frequent cannabis user 18+ (frequent defined as response option *three or more times*).

At Wave 7 participants were also asked: *and have you taken cannabis in the last 12 months?* We created a binary variable from these data: had cannabis in last 12 months age 18+.

**Socio-Demographic Variables.**

***Family socio-economic position (SEP).*** Linkage to the 2004 (Wave 1) and 2006 (Wave 3) NPD included linkage to data on eligibility for FSM. Eligibility for FSMs is determined by data linkage to government records of receipt of at least one of a defined list of means-tested welfare benefits by the child’s parent(s). It should be noted that this indicator is of eligibility for, not uptake of, FSMs. We created a binary variable of FSM eligibility scored 1 if the child was eligible at Wave 1, Wave 3 or both Waves of *Next Steps* and scored 0if the child was not eligible at both Waves. FSM eligibility is a commonly used proxy indicator of low household socio-economic position (Kounali, Robinson, Goldstein, & Lauder, 2008).

We extracted data from *Next Steps* on the employment status of parental figures living in the household at Waves 1-4 inclusive. We created a binary variable of living in a workless household scored 1 if no resident parental figure was in employment or full time education at any of the four Waves and scored 0 if at least one resident parental figure was in employment or full time education in each of the four Waves.

***Young adult socio-economic position.*** We extracted data from *Next Steps* on the self-reported employment, education and training status of the young person at Waves 5-7. We created a binary variable of not in employment, education or training (NEET) scored 1 if the young person was NEET at any of the three Waves and scored 0 if they were in employment, education or training in each of the three Waves.

***Household composition.***We extracted data from *Next Steps* on household composition at Waves 1-4 inclusive. We created a binary variable of single parent household scored 1 if only one parental figure was resident at any of the four Waves and scored 0 if two parental figures were resident in each of the four Waves.

***Area deprivation.***Linkage to the 2004 (Wave 1) and 2006 (Wave 3) NPD also included linkage to data derived from the postal code of the child’s residence to the Income Deprivation Affecting Children Index (IDACI) (Noble et al., 2008). IDACI scores are the percentage of children in each Lower Level Super Output Area (LSOA) that live in families that are considered income deprived. Income deprivation is defined by receipt of means-tested welfare benefits. LSOAs are neighbourhoods with an average population of 1500 (range 1000-3000). IDACI scores were transformed into sample quintiles. We created a binary variable of High Neighbourhood Deprivation scored 1 if the child was living in the lowest IDACI quintile at Wave 1, Wave 3 or both Waves of *Next Steps* and scored 0if the child was not living in the lowest IDACI quintile at both Waves.

***Peer victimisation.*** We extracted data from *Next Steps* on child self-reported experience of peer victimisation (bullying) at Waves 1-3. At each wave, children were asked about exposure to five types of peer victimisation in the last 12 months:

*Have you ever been upset by being called hurtful names by other students, including getting text messages or emails from them?*

*Have you ever been excluded from a group of friends or from joining in activities?*

*Have other students at your school ever made you give them money or personal possessions?*

*Have other students ever THREATENED to hit you, kick you or use any other form of violence against you?*

*Have other students ever ACTUALLY hit you, kicked you or used any other form of violence against you?*

If the young participant selected a ‘yes’ option they were then asked about the frequency of exposure (response options: every day, a few times a week, once or twice a week, once every two weeks, once a month, less often than this, it varies). Preliminary analysis of responses indicated a strong association between threat of and actual violence, but weak associations between other forms of peer victimisation. As a result we combined self-report of threat of or actual violence at each of the three Waves. For each of the four types of peer victimisation (name calling, social exclusion, theft, violence) we created one binary variable; whether this had happened at all in any 12 month period in Waves 1-3 (contrasted with it never happening in any of the three Waves).

***Friendships.*** We extracted information on friendships from Waves 2, 6 and 7. At Wave 2 participants were asked: *When you have free time, do you mainly: (1) Go out somewhere with friends; (2) Go round to a friend's house (or friends come round to yours); (3) Spend time with brother(s)/sister(s); (4) Spend time with other members of your family or; (5) Spend time by yourself?* We created a binary variable, W2 spends time with friends, scored 1 if they selected option 1 or 2, scored 0 if they selected options 3-5.

At Waves 6 and 7 participants were asked: *How many close friends do you have – that is friends you could talk to if you were in some sort of trouble?* We created a binary variable, W6/7 few friends, scored 1 if they reported at either Wave they had no or only 1 close friend and scored 0 if they reported at any Wave they had two or more close friends.

**Approach to Analysis**

In the first stage of analysis we made simple bivariate comparisons between participants with and without intellectual disability with regard to available indicators of smoking, alcohol and drug use. In the second stage of analysis we investigated, for key indicators of smoking, alcohol and drug use, the strength of association between socio-demographic factors and outcomes separately for participants with and without intellectual disability. Missing data among socio-demographic variables was imputed using multiple imputation routines in SPSS 22 to create five parallel imputed data sets. The subsequent analysis used the following approach: (1) five blocks of variables were created (SEP, neighbourhood, family type, peer victimisation, friendships) and entered sequentially; (2) variables within blocks were entered in order of bivariate strength of association with the outcome of interest; (3) variables were only retained in the model if *at the point of entry* they were significantly related to the outcome of interest or had a PR of 1.50 or greater. Poisson regression with robust standard errors was used to estimate PRs uniquely associated with each variable in the model (Knol, Le Cessie, Algra, Vandenbroucke, & Groenwold, 2012; Zocchetti, Consonni, & Bertazzi, 1997).

In the final stage of analysis we estimated the strength of association between intellectual disability and smoking, alcohol and drug use while controlling for between group differences in exposure to socio-demographic variables that have been established as important social determinants of poorer health. Two separate approaches were used. First, we used Poisson regression with robust standard errors to estimate PRs associated with intellectual disability for each outcome with exposures included in the model as covariates. Second, we used Propensity Score Matching (PSM) routines in SPSS 22 to match each participant with intellectual disability with a participant without intellectual disability with a similar propensity score for intellectual disability based on exposure to the socio-demographic variables (Austin, 2011; Blackford, 2007; Oakes & Johnson, 2006). We used the lowest tolerance for matching (0.05) that allowed complete matching for all participants with intellectual disability.

**References**

Austin, P. C. (2011). An introduction to Propensity Score Methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*, 399-424. doi: 10.1080/00273171.2011.568786

Blackford, J. (2007). Statistical issues in developmental epidemiology and developmental disabilities research: Confounding variables, small sample size, and numerous outcome variables In R. Urbano & R. Hodapp (Eds.), *Developmental Epidemiology of Mental Retardation and Developmental Disabilities* (pp. 93-120). New York: Academic Press.

Department for Education. (2011a). Glossary of special educational needs (SEN) terminology. Available online at http://webarchive.nationalarchives.gov.uk/20130123124929/http://www.education.gov.uk/schools/pupilsupport/sen/schools/a0013104/glossary-of-special-educational-needs-sen-terminology (last accessed 12 Sept 2017): Department for Education.

Department for Education. (2011b). LYSPE Guide to the Datasets: Wave 1-Wave 7. London: Department for Education.

Emerson, E. (2012). Household deprivation, neighbourhood deprivation, ethnicity and the prevalence of intellectual and developmental disabilities *Journal of Epidemiology and Community Health, 66*, 218-224. doi: 10.1136/jech.2010.111773

Emerson, E., & Halpin, S. (2013). Anti‐social behaviour and police contact among 13‐ to 15‐year‐old English adolescents with and without mild/moderate intellectual disability. *Journal Of Applied Research In Intellectual Disabilities, 26*(5), 362-369. doi: 10.1111/jar.12041

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys *Public Opinion Quarterly, 70*, 646-675.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.

Knol, M. J., Le Cessie, S., Algra, A., Vandenbroucke, J. P., & Groenwold, R. H. H. (2012). Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives for logistic regression. *Canadian Medical Association Journal, 184*, 895-899. DOI:810.1503/cmaj.101715.

Kounali, D., Robinson, T., Goldstein, H., & Lauder, H. (2008). The probity of free school meals as a proxy measure for disadvantage. Bristol: University of Bristol.

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies *Research in Developmental Disabilities, 32*, 419-436. doi: 10.1016/j.ridd.2010.12.018

Naylor, P., Dawson, J., Emerson, E., & Tantam, D. (2011). Prevalence of bullying in secondary school by SEN type: Analysis of combined NPD and LSYPE data files. End of Award Report to ESRC. Swindon: ESRC.

Noble, M., McLennan, D., Wilkinson, K., Whitworth, A., Barnes, H., & Dibben, C. (2008). The English Indices of Deprivation 2007. London: Communities and Local Government.

Oakes, J. M., & Johnson, P. J. (2006). Propensity score matching for social epidemiology. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in Social Epidemiology*. San Francisco: Josey Bass.

Roeleveld, N., Zielhuis, G. A., & Gabreels, F. (1997). The prevalence of mental retardation: a critical review of recent literature. *Developmental Medicine & Child Neurology, 39*, 125-132.

Zocchetti, C., Consonni, D., & Bertazzi, P. (1997). Relationship between prevalence rate ratios and odds ratios in crosssectional studies. *International Journal of Epidemiology, 26*(1), 220-223.