SUPPLEMENTARY MATERIAL

1. Details on the multiple imputation procedure

Multiple imputation was implemented using the “multivariate imputation by chained equations” (MICE) approach (van Buuren, 2012). The wave-level measures for the adolescent risk factors (common mental disorder, disruptive behaviour, cannabis use and drinking behaviour), background factors (school location, parental education and parental divorce or separation) and the outcome, NEET status, measured at waves 7 and 8 were included in the imputation model. A small set of auxiliary variables was also included in the imputation model. These auxiliary variables were:

* wave-7 measures of common mental disorder, cannabis use and drinking behaviour
* a 6-level variable on participants’ self-rated academic performance compared with the performance of other students in their year level, measured at waves 2 to 6
* a 6-level variable on participant smoking status measured at waves 2 to 7 with categories: non-smoker, ex-smoker, not in the past week, smoking less than daily, daily with less than 10 cigarettes per day, and smoking more than 10 cigarettes each day
* a binary variable indicating if either parent smoked at any stage during participants’ adolescence

Self-rated academic performance was selected as an auxiliary variable as previous studies have reported associations between academic performance or cognitive ability and NEET status (Bynner and Parsons, 2002, Cornaglia *et al.*, 2012, Gladwell *et al.*, 2015, Moore *et al.*, 2015, Veldman *et al.*, 2015). Further, correlations between poor academic performance or educational underachievement and mental health (Fröjd *et al.*, 2008), cannabis use (Hall, 2015, Meier *et al.*, 2015) and drinking behaviour (Balsa *et al.*, 2011) have also been demonstrated in previous studies. We did not include self-rated academic performance as a potential confounder in the analysis model because we recognised the potential for bi-directional relationships between the adolescent risk factors and self-rated academic performance.

The selection of the parental and adolescent smoking variables as auxiliary variables was based on a previous VAHCS study that reported an association between parental and adolescent smoking and level of adolescent cannabis use (Swift *et al.*, 2012). The adolescent smoking behaviour variables have also been included as potential confounding factors in an analysis that examined the association between common mental disorder and alcohol use (McKenzie *et al.*, 2011).

A total of 41 variables were included in the imputation procedure. The full list of variables is given in Table A1. Gender and school location were the only complete variable included in the imputation procedure. Table A2 shows the twenty most frequent patterns of response for the variables that were included in the imputation model, including the auxiliary variables.

Table A1 also provides an overview of the univariate imputation models that were specified for each incomplete variable. We adopted the following approaches for the development of the univariate imputation models:

* The univariate imputation models for the NEET outcome at waves 7 and 8, and the family background variables were conditioned on the full set of variables included in the imputation procedure.
* Each univariate model for the imputation of the remaining variables included: the NEET outcome at waves 7 and 8; family background variables; the complete set of variables measured at wave 2; the complete set of variables measured at same wave as the incomplete variable being imputed; and the complete set of variables measured at the wave(s) adjacent to the relevant wave.

To respect any effect modification by gender, data were imputed separately for males and females.

The MICE procedure was implemented using the mi impute chained command in Stata 14 (StataCorp, 2015). The ordering of variables in the MICE algorithm was based on the amount of missing data in each variable, starting with the variable that had the least missing data and ending with the variable that had the most missing data. An examination of the convergence behaviour of the MICE algorithm indicated that a burn-in period of 20 cycles was appropriate (i.e. the algorithm cycled through the full sequence of univariate imputation models 20 times before drawing the imputations).

To increase our confidence in the reproducibility of the results obtained from the multiple imputation analysis, we generated 100 imputations. Data were imputed in the wide format (i.e. one row per participant). The summary adolescent variables were derived using wave-level data in each completed dataset after imputation, using the mi impute passive command in Stata. Data were then transformed to the long format with one row for each wave at which the outcome was measured (i.e. two rows per participant). The analyses were conducted using the mi impute estimate command in Stata. The full MI procedure was run separately for the two outcomes, NEET and NEEPT.

**Table A1.** Description of variables included in the imputation procedure and the imputation model specified for each incomplete variable

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Variable type** | **Imputation model** | **Wave** | **Missing****no. (%)** |
| NEET (outcome) | Binary | logistic regression | 7 | 344 (18) |
| 8 | 425 (22) |
| Common mental disorder (CIS-R) | Semi-continuousrange: 0-57 | two-part: logistic regression for zero/positive component and linear regression (conditional on positive component) | 2 | 159 (8.2) |
| 3 | 244 (12.6) |
| 4 | 312 (16.1) |
| 5 | 365 (18.8) |
| 6 | 410 (21.1) |
| 7 (A) | 339 (17.5) |
| Cannabis use | Binary | logistic regression | 2 | 227 (11.7) |
| 3 | 282 (14.6) |
| 4 | 343 (17.7) |
| 5 | 401 (20.7) |
| 6 | 434 (22.4) |
| 7 (A) | 350 (18.1) |
| Disruptive behaviour | Binary | logistic regression | 2 | 200 (10.3) |
| 3 | 266 (13.7) |
| 4 | 332 (17.1) |
| 5 | 389 (20.1) |
| 6 | 428 (22.1) |
| Drinking behaviour (number of alcohol units) | Semi-continuousrange: 0-50 | two-part: logistic regression for zero/positive component and linear regression (conditional on positive component) | 2 | 159 (8.2) |
| 3 | 244 (12.6) |
| 4 | 312 (16.1) |
| 5 | 365 (18.8) |
| 6 | 410 (21.2) |
| 7 (A) | 347 (17.9) |
| Self-rated academicPerformance (A) | Continuousrange: 1-6 | linear regression, no post-imputation rounding  | 2 | 183 (9.4) |
| 3 | 258 (13.3) |
| 4 | 371 (19.1) |
| 5 | 428 (22.1) |
| 6  | 540 (27.9) |
| Smoking (A) | 6-level ordinal | linear regression, no post-imputation rounding | 2 | 159 (8.2) |
| 3 | 244 (12.6) |
| 4 | 312 (16.1) |
| 5 | 365 (18.8) |
| 6 | 410 (21.2) |
| 7 | 371 (19.1) |
| Gender | Binary | NA | - | 0 (0) |
| Parental education | Ordinal | Ordinal logistic regression  | - | 100 (5.2) |
| School location | Binary | NA | - | 0 (0) |
| Parental divorce or separation | Binary | Logistic regression | - | 2 (0.1) |
| Parental smoking (A) | Binary | Logistic regression  | - | 97 (5.0) |

A = auxiliary variable for the imputation model.

**Table A2.** Twenty most frequent patterns of response on all variables included in the imputation model for the 1938 participants in the VAHCS analysis sample

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Family background factors | Response on all variables measured at wave‡(Numbers in columns represent the number of participants with data missing on all variables at the relevant wave) | Outcome | No. of participants | % of participants |
|  | parental divorce | parental education | parental smoking | wave 2 | wave 3 | wave 4 | wave 5 | wave 6 | wave 7\* | NEET wave 7 | NEET wave 8 |
| 1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 861 | 44.4 |
| 2 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 40 | ✓ | ✓ | ✓ | 77 | 4.0 |
| 3 | ✓ | ✓ | ✓ | 56 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 74 | 3.8 |
| 4 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | · | 73 | 3.8 |
| 5 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 50 | · | · | 50 | 2.6 |
| 6 | ✓ | ✓ | ✓ | ✓ | 46 | 42 | 45 | 44 | ✓ | ✓ | ✓ | 48 | 2.5 |
| 7 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 20 | 16 | ✓ | ✓ | ✓ | 40 | 2.1 |
| 8 | ✓ | ✓ | ✓ | ✓ | ✓ | 24 | ✓ | ✓ | ✓ | ✓ | ✓ | 37 | 1.9 |
| 9 | ✓ | ✓ | ✓ | ✓ | 23 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 36 | 1.9 |
| 10 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 33 | · | ✓ | 33 | 1.7 |
| 11 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 10 | ✓ | ✓ | ✓ | ✓ | 31 | 1.6 |
| 12 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 0 | ✓ | ✓ | 30 | 1.5 |
| 13 | ✓ | ✓ | ✓ | ✓ | ✓ | 20 | 24 | 16 | ✓ | ✓ | ✓ | 27 | 1.4 |
| 14 | ✓ | ✓ | ✓ | ✓ | 18 | 18 | 16 | 17 | ✓ | ✓ | · | 19 | 1.0 |
| 15 | ✓ | · | · | ✓ | ✓ | 13 | 19 | 19 | 19 | · | · | 19 | 1.0 |
| 16 | ✓ | ✓ | ✓ | ✓ | 16 | 14 | 15 | 14 | 17 | · | · | 17 | 0.9 |
| 17 | ✓ | · | · | ✓ | 13 | 15 | 15 | 15 | 15 | · | · | 15 | 0.8 |
| 18 | ✓ | ✓ | ✓ | ✓ | ✓ | 13 | 11 | ✓ | ✓ | ✓ | ✓ | 14 | 0.7 |
| 19 | ✓ | ✓ | ✓ | 2 | 11 | 8 | 13 | 14 | ✓ | ✓ | ✓ | 14 | 0.7 |
| 20 | ✓ | ✓ | ✓ | ✓ | ✓ | 10 | 12 | 10 | ✓ | ✓ | · | 13 | 0.7 |

✓ Indicates all data are available on the relevant variable or wave. ‡ For waves 2 to 7, non-response indicates data are missing on at least one variable measured at that wave. \*Response on the outcome variable (NEET status) was considered separately and so was not included in the response summary for wave 7. A total of 410 participants (21%) had a pattern of response other than one of those shown. Data on gender and school location were complete for all participants.

1. **Additional results for NEEPT status**

**Table B1.** Prevalence of NEEPT status in young adulthood, by adolescent risk and background factors

|  |  |  |
| --- | --- | --- |
| **Measures**  |  | **Percent NEEPT** |
|  | **Wave 7****(mean age 20.7)** |  | **Wave 8****(mean age 24.1)** |
| **n†** | **n(%)‡** | **(95% CI)** |  | **n(%)‡** | **(95% CI)** |
| **Adolescent risk factors** |  |  |  |  |  |  |
| Persistence of common mental disorder (CIS-R ≥12) |  |  |  |  |  |  |
|  | No waves | 1118 | 67 (6.1) | (4.4, 7.7) |  | 47 (4.2) | (2.8, 5.7) |
|  | 1 wave | 316 | 26 (8.3) | (4.3, 12.2) |  | 19 (6.2) | (2.8, 9.5) |
|  | 2+ waves | 504 | 48 (9.4) | (6.3, 12.4) |  | 40 (7.9) | (5.2, 10.7) |
| Cannabis use |  |  |  |  |  |  |
|  | None or infrequent use | 1638 | 95 (5.8) | (4.5, 7.2) |  | 78 (4.8) | (3.5, 6.1) |
|  | Frequent (at least weekly) use | 300 | 46 (15.2) | (10.0, 20.4) |  | 28 (9.5) | (4.9, 14.1) |
| Drinking behaviour |  |  |  |  |  |  |
|  | No drinking | 742 | 47 (6.3) | (4.1, 8.4) |  | 40 (5.5) | (3.4, 7.7) |
|  | Any drinking, below heavy binge levels | 818 | 57 (7.0) | (4.9, 9.2) |  | 41 (5.0) | (3.1, 7.0) |
|  | Any heavy binge drinking | 378 | 37 (9.8) | (6.0, 13.7) |  | 25 (6.6) | (3.4, 9.7) |
| Persistent disruptive behaviour |  |  |  |  |  |  |
|  | No | 1339 | 75 (5.6) | (4.2, 7.1) |  | 62 (4.7) | (3.3, 6.0) |
|  | Yes | 599 | 66 (11.0) | (8.0, 14.1) |  | 44 (7.5) | (4.9, 10.1) |
| **Background factors** |  |  |  |  |  |  |
| Gender |  |  |  |  |  |  |
|  | Male | 939 | 75 (8.0) | (5.8, 10.2) |  | 53 (5.8) | (3.9, 7.8) |
|  | Female | 999 | 66 (6.6) | (4.9, 8.4) |  | 52 (5.2) | (3.6, 6.8) |
| School location |  |  |  |  |  |  |
|  | Within metropolitan area | 1440 | 98 (6.8) | (5.2, 8.3) |  | 79 (5.5) | (4.0, 6.9) |
|  | Outside metropolitan area | 498 | 43 (8.8) | (5.9, 11.7) |  | 28 (5.7) | (3.2, 8.3) |
| Highest level of parental education |  |  |  |  |  |  |
|  | High school not completed | 652 | 56 (8.6) | (6.0, 11.2) |  | 36 (5.5) | (3.3, 7.8) |
|  | High school completed | 661 | 53 (8.0) | (5.7, 10.4) |  | 35 (5.4) | (3.2, 7.5) |
|  | University or college degree | 625 | 32 (5.1) | (3.1, 7.1) |  | 35 (5.7) | (3.6, 7.8) |
| Parental divorce or separation |  |  |  |  |  |  |
|  | No  | 1500 | 85 (5.7) | (4.3, 7.0) |  | 72 (4.9) | (3.5, 6.2) |
|  | Yes | 438 | 56 (12.9) | (9.1, 16.7) |  | 34 (7.8) | (4.6, 11.0) |

CI = confidence interval. CIS-R = Revised Clinical Interview Schedule. † ns were calculated using imputed percentage estimates and total number of participants. ‡% represents the number and estimated percentage of the n who are NEEPT (not in education, employment, parenting, or training), averaged over 100 completed datasets.

**Table B2.** Association between adolescent risk factors and NEEPT status in young adulthood (waves 7 and 8)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Measures** | **Model a****Risk factors fitted** **separately** |  | **Model b****Risk factors mutually adjusted** |  | **Model c****Adjusted for potential confounders\*** |
|  | **OR** | **(95% CI)** | **p-value** |  | **ORadj** | **(95% CI)** | **p-value** |  | **ORadj** | **(95% CI)** | **p-value** |
| **Adolescent risk factors** |  |  |  |  |  |  |  |  |  |  |  |
| Persistence of common mental disorder (CIS-R ≥12) |  |  | 0.01† |  |  |  | 0.11† |  |  |  | 0.15† |
|  | No waves | 1.00 |  |  |  | 1.00 |  |  |  | 1.00 |  |  |
|  | 1 wave | 1.42 | (0.85, 2.40) |  |  | 1.30 | (0.77, 2.19) |  |  | 1.33 | (0.78, 2.27) |  |
|  | 2+ waves | 1.74 | (1.20, 2.54) |  |  | 1.51 | (1.03, 2.22) |  |  | 1.53 | (0.99, 2.37) |  |
| Frequent (at least weekly) cannabis use | 2.51 | (1.64, 3.83) | <0.0001 |  | 2.17 | (1.29, 3.66) | 0.003 |  | 2.01 | (1.18, 3.40) | 0.01 |
| Drinking behaviour |  |  | 0.29† |  |  |  | 0.61† |  |  |  | 0.47† |
|  | No drinking | 1.00 |  |  |  | 1.00 |  |  |  | 1.00 |  |  |
|  | Any drinking, below heavy binge levels | 1.03 | (0.67, 1.57) |  |  | 0.83 | (0.54, 1.27) |  |  | 0.80 | (0.52, 1.22) |  |
|  | Any heavy binge drinking | 1.43 | (0.87, 2.34) |  |  | 0.78 | (0.44, 1.38) |  |  | 0.73 | (0.41, 1.29) |  |
| Persistent disruptive behaviour | 1.88 | (1.32, 2.68) | 0.001 |  | 1.49 | (0.98, 2.24) | 0.06 |  | 1.42 | (0.93, 2.19) | 0.11 |
| **Potential confounding factors** |  |  |  |  |  |  |  |  |  |  |
| School located outside metropolitan area |  |  |  |  |  |  |  |  | 1.22 | (0.56, 1.19) | 0.30 |
| Highest level of parental education |  |  |  |  |  |  |  |  |  |  | 0.30† |
|  | High school not completed |  |  |  |  |  |  |  |  | 1.00 |  |  |
|  | High school completed |  |  |  |  |  |  |  |  | 1.00 | (0.66, 1.50) |  |
|  | University or college degree |  |  |  |  |  |  |  |  | 0.74 | (0.48, 1.15) |  |
| Parental divorce or separation |  |  |  |  |  |  |  |  | 1.74 | (1.19, 2.54) | 0.005 |
| Female |  |  |  |  |  |  |  |  | 0.80 | (0.50, 1.30) | 0.39 |
| Outcome at wave 8 (*v*s wave 7) |  |  |  |  |  |  |  |  | 0.71 | (0.46, 1.09) | 0.05 |
| Female x wave 8 |  |  |  |  |  |  |  |  | 1.09 | (0.61, 1.94) | 0.76 |

Note: Odds ratios were obtained using generalized estimating equations assuming an exchangeable working correlation matrix with robust standard errors to allow for the repeated outcome measure.

OR = odds ratio. ORadj = adjusted odds ratio. CI = confidence interval. CIS-R = Revised Clinical Interview Schedule. NEEPT = not in employment, education, parenting, or training †p-value for joint test of category coefficients \*also adjusted for gender and wave

**References**

**Balsa, A. I., Giuliano, L. M. & French, M. T.** (2011). The effects of alcohol use on academic achievement in high school. *Economics of education review* **30**, 1-15.

**Bynner, J. & Parsons, S.** (2002). Social exclusion and the transition from school to work: The case of young people not in education, employment, or training (NEET). *Journal of Vocational Behavior* **60**, 289-309.

**Cornaglia, F., Crivellaro, E. & McNally, S.** (2012). Mental health and education decisions. London School of Economics: London.

**Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U. I., Marttunen, M. J., Koivisto, A.-M. & Kaltiala-Heino, R.** (2008). Depression and school performance in middle adolescent boys and girls. *Journal of Adolescence* **31**, 485-498.

**Gladwell, D., Popli, G. & Tsuchiya, A.** (2015). Estimating the impact of health on NEET status. In *Sheffield Ecomonic Research Paper Series*. Institute of ecomonic analyais of decision making: Sheffield.

**Hall, W.** (2015). What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? *Addiction* **110**, 19-35.

**McKenzie, M., Jorm, A. F., Romaniuk, H., Olsson, C. A. & Patton, G. C.** (2011). Association of adolescent symptoms of depression and anxiety with alcohol use disorders in young adulthood: findings from the Victorian Adolescent Health Cohort Study. *Medical Journal of Australia* **195**, 27.

**Meier, M. H., Hill, M. L., Small, P. J. & Luthar, S. S.** (2015). Associations of adolescent cannabis use with academic performance and mental health: A longitudinal study of upper middle class youth. *Drug and Alcohol Dependence* **156**, 207-212.

**Moore, S. E., Scott, J. G., Thomas, H. J., Sly, P. D., Whitehouse, A. J., Zubrick, S. R. & Norman, R. E.** (2015). Impact of adolescent peer aggression on later educational and employment outcomes in an Australian cohort. *Journal of Adolescence* **43**, 39-49.

**StataCorp** (2015). Stata Statistical Software: Release 14. StataCorp LP,: College Station, TX.

**Swift, W., Coffey, C., Degenhardt, L., Carlin, J. B., Romaniuk, H. & Patton, G. C.** (2012). Cannabis and progression to other substance use in young adults: findings from a 13-year prospective population-based study. *Journal of Epidemiology and Community Health* **66**, e26. 10.1136/jech.2010.129056.

**van Buuren, S.** (2012). *Flexible imputation of missing data*. CRC press: Boca Raton.

**Veldman, K., Reijneveld, S. A., Almansa Ortiz, J., Verhulst, F. C. & Bültmann, U.** (2015). Mental health trajectories from childhood to young adulthood affect the educational and employment status of young adults: results from the TRAILS study. *Journal of Epidemiology and Community Health* **69**, 588-593.