## Online appendix: Approach to multiple imputation

Missing data were handled using multiple imputation<sup>1</sup>. We imputed 20 complete datasets separately for males and females, under a multivariate normal model, incorporating all the analysis and auxiliary variables at each wave, except age and sexual risk-taking in adolescence, which was imputed as a binary summary measure. Wave 1 was omitted as it contained observations from only 46% of the cohort. Wave 1 responses were used to fill in wave 2 data, for any participant not seen at wave 2 as the same measures had been collected at the earlier wave. Waves 2 to 9 were imputed for all participants (N=1943) who completed the survey at least once in adolescence (waves 2-6). The imputation model contained 49 key variables used in the analysis and 8 auxiliary variables. Of 57 variables included in the imputation model, 18% of the variables had <10% missing values, 46% had  $\geq$ 10% to <20% missing, 28% had  $\geq$ 20% to <30% missing and only 5 variables (9%) had  $\geq$ 45% to <55% missing values. A maximum level of drinking variable was created at each wave, with three levels: no binge drinking, binge drinking and extreme binge drinking. These variables and alcohol consequences variables were log transformed before imputation. Smoking, cannabis use, antisocial behaviour, mental health and summary measures of sexual risk-taking were imputed as binary variables at each wave. Age was imputed as a normal variable. After imputation, transformed variables were converted back to their original scale and all were categorised for analysis, with adaptive rounding used for binary measures<sup>2</sup>. Nine deceased participants were excluded from the imputed datasets. Frequencies and odds ratios were obtained by averaging results across the imputed datasets; inferences under multiple imputation were made using Rubin's rules<sup>1</sup>.

Table 1a. Available case analysis: Prevalence of adolescent and young adult binge drinking by wave and sex

	Binge drinking in the past week											
	Males				Females				Total			
	Ν	n	%	(95% CI)	Ν	n	%	(95% CI)	Ν	n	%	(95% CI)
ADOLESCENT PHASE (WAVES 2-6)												
Binge drinking by wave												
Wave 2 (mean age 15.5 yr)	892	127	14.2	(11.9 - 16.5)	950	72	7.6	(5.90 - 9.30)	1842	199	10.8	(9.40 - 12.2)
Wave 3 (mean age 15.9 yr)	805	137	17.0	(14.4 - 19.6)	887	79	8.9	(7.00 - 10.8)	1692	216	12.8	(11.2 - 14.4)
Wave 4 (mean age 16.4 yr)	750	151	20.1	(17.3 - 23.0)	874	112	12.8	(10.6 - 15.0)	1624	263	16.2	(14.4 - 18.0)
Wave 5 (mean age 16.8 yr)	718	186	25.9	(22.7 - 29.1)	853	117	13.7	(11.4 - 16.0)	1571	303	19.3	(17.3 - 21.2)
Wave 6 (mean age 17.4 yr)	679	195	28.7	(25.3 - 32.1)	847	122	14.4	(12.0 - 16.8)	1526	317	20.8	(18.7 - 22.8)
YOUNG ADULT PHASE (WAVES 7-9)												
Binge drinking by wave												
Wave 7 (mean age 20.7 yr)	732	401	54.8	(51.2 - 58.4)	865	266	30.8	(27.7 - 33.8)	1597	667	41.8	(39.3 - 44.2)
Wave 8 (mean age 24.1 yr)	693	409	59.0	(55.4 - 62.7)	824	265	32.2	(29.0 - 35.4)	1517	674	44.4	(41.9 - 46.9)
Wave 9 (mean age 29.1 yr)	631	359	56.9	(53.0 - 60.8)	764	168	22.0	(19.1 - 24.9)	1395	527	37.8	(35.2 - 40.3)

Note: "Binge" drinking defined as 5 or more drinks in one day in the past week.

## Table 1b. Available case analysis: Prevalence of adolescent and young adult extreme binge drinking by wave by sex

	Extreme binge drinking in the past week											
			Males		Females				Total			
	Ν	n	%	(95% CI)	N	n	%	(95% CI)	N	n	%	(95% CI)
ADOLESCENT PHASE (WAVES 2-6)												
Binge drinking by wave												
Wave 2 (mean age 15.5 yr)	892	44	4.9	(3.50 - 6.40)	950	32	3.4	(2.20 - 4.50)	1842	76	4.1	(3.20 - 5.00)
Wave 3 (mean age 15.9 yr)	805	42	5.2	(3.70 - 6.80)	887	31	3.5	(2.30 - 4.70)	1692	73	4.3	(3.30 - 5.30)
Wave 4 (mean age 16.4 yr)	750	41	5.5	(3.80 - 7.10)	874	52	5.9	(4.40 - 7.50)	1624	93	5.7	(4.60 - 6.90)
Wave 5 (mean age 16.8 yr)	718	63	8.8	(6.70 - 10.8)	853	62	7.3	(5.50 - 9.00)	1571	125	8.0	(6.60 - 9.30)
Wave 6 (mean age 17.4 yr)	892	44	4.9	(3.50 - 6.40)	847	36	4.3	(2.90 - 5.60)	1526	85	5.6	(4.40 - 6.70)
YOUNG ADULT PHASE (WAVES 7-9)												
Binge drinking by wave												
Wave 7 (mean age 20.7 yr)	732	141	19.3	(16.4 - 22.1)	865	121	14.0	(11.7 - 16.3)	1597	262	16.4	(14.6 - 18.2)
Wave 8 (mean age 24.1 yr)	693	153	22.1	(19.0 - 25.2)	824	120	14.6	(12.2 - 17.0)	1517	273	18.0	(16.1 - 19.9)
Wave 9 (mean age 29.1 yr)	631	85	13.5	(10.8 - 16.1)	764	70	9.2	(7.10 - 11.2)	1395	155	11.1	(9.50 - 12.8)

Note: "*Extreme binge*" drinking defined as >11 drinks on a day in the past week for females, >20 for males.

## References

1. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons; 2004.

2. Bernaards CA, Belin TR, Schafer JL. Robustness of a multivariate normal approximation for imputation of incomplete binary data. Statistics in Medicine. 2007; 26: 1368-82.