**QUEEN'S
UNIVERSITY
BELFAST**

Early illicit drug use and the age of onset of homelessness

McVicar, D., Moschion, J., & van Ours, J. C. (2018). Early illicit drug use and the age of onset of homelessness. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. <https://doi.org/10.1111/rssa.12411>

Published in:

Journal of the Royal Statistical Society: Series A (Statistics in Society)

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2018 Wiley. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Early illicit drug use and the age of onset of homelessness

Duncan McVicar¹

Julie Moschion²

Jan C. van Ours³

June 2018

Abstract

We investigate the effect of taking up daily cannabis use on the onset of homelessness using Australian data. We use a bivariate simultaneous mixed proportional hazard model to address potential biases due to common unobservable factors and reverse causality. We find that taking up daily cannabis use substantially increases the probability of transition into homelessness for young men but not young women. In contrast, homelessness onset increases the probability of taking up daily cannabis use for young women but not for young men. In a trivariate extension we find that the use of other illicit drugs at least weekly has no additional effect on transitions into homelessness for either gender but there is a large if imprecisely estimated impact of homelessness onset on taking up weekly use of such drugs for young women.

Keywords: homelessness, cannabis, drug use, simultaneous duration model, bivariate duration model, trivariate duration model, timing of events

JEL-codes: C41, D12, I19, I39, J13

¹ Corresponding Author. Queen's Management School, Queen's University Belfast, Riddel Hall, 185 Stranmillis Road, Belfast BT9 5EE, United Kingdom. Tel: +44 (0)28 9097 4200, and IZA (Bonn); d.mcvicar@qub.ac.uk.

² Melbourne Institute of Applied Economic and Social Research, University of Melbourne, University of Melbourne, Vic 3010, Australia, and IZA (Bonn). Tel: +61(0)3 8344 2100. moschion@unimelb.edu.au.

³ Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands; Department of Economics, University of Melbourne, Australia; Tinbergen Institute (Amsterdam/Rotterdam); CEPR (London) and IZA (Bonn); email: vanours@ese.eur.nl.

1 Introduction

There is a growing quantitative social science literature on the effects of using illicit drugs such as cannabis or cocaine at a young age on a variety of health and social outcomes. One strand of this literature examines drug use impacts on educational attainment, with many studies suggesting negative impacts (e.g. Bray et al., 2000; Register et al., 2001; Roebuck et al., 2004; Chatterji, 2006; Duarte et al., 2006; Van Ours and Williams, 2009) and some suggesting no impact (e.g. McCaffrey et al., 2010; Cobb-Clark et al., 2015). Another strand of the literature concerns drug use impacts on wages and other labour market outcomes, where estimated effects range from the positive (e.g. Kaestner 1991 on wages), through zero (e.g. Kaestner 1994 and Zarkin et al. 1998 on hours worked; McDonald and Pudney 2000 on occupational status), to the negative (e.g. DeSimone 2002 on employment; McDonald and Pudney 2000 on (avoiding) unemployment). Adverse impacts have also been found on mental and physical health (e.g. Van Ours and Williams, 2011; Van Ours and Williams, 2012; Van Ours et al., 2013). All of the studies cited above attempt to deal with the likely endogeneity of drug use, albeit with mixed success, using a variety of approaches (most commonly IV). In general, this literature suggests the younger the age of onset of drug use the bigger the detrimental impact. For recent reviews see Hall (2015) and Van Ours and Williams (2015).

In this paper we examine another key social outcome – the onset of homelessness – that has been widely mooted, but not yet convincingly established, as a possible consequence of early illicit drug use. Several studies have demonstrated a strong positive association between youth homelessness and substance abuse (e.g. Greene et al., 1997; Mallett et al., 2005; Shelton et al., 2009). Others have gone further by specifically examining associations between early substance use and onset of homelessness including in the context of multivariate regression models, with mixed results (e.g. Johnson and Fendrich, 2007; Johnson and Chamberlain, 2008; van den Bree et al., 2009). No existing study, however, has satisfactorily addressed the likely endogeneity of drug use in this case, whether stemming from selection on unobservables (i.e. that early illicit drug users are likely to be different from non-users in ways that are associated with homelessness onset but that we do not observe) or simultaneity (homelessness may impact on drug use as well as vice versa). This is also the case in the wider homelessness-substance abuse literature with the partial exception of McVicar et al. (2015) which uses longitudinal data from the Australian Journeys Home (JH) study to estimate the impact of changes in substance use on changes in homelessness status using individual fixed effects models. Conditional on these fixed effects they find no impact of cannabis use or other illicit drug use on homelessness,

although they do find an effect of heavy alcohol use. In other words, once you've been homeless – and almost all of the sample studied by McVicar et al. (2015) have prior experience of homelessness – changes in drug use behaviour have little impact on changes in homelessness status. Neither McVicar et al. (2015) nor any other existing study, however, presents convincing quantitative evidence on whether the initial *take-up* of drug use behaviours impacts on the probability of becoming homeless *for the first time*, i.e. homelessness onset.

To address this important gap in the homelessness and substance use literatures we use information from the JH study on the ages at which individuals first engage in regular use of illicit drugs and first become homeless to estimate bivariate and trivariate simultaneous duration models allowing for common unobserved factors to influence the onset of both homelessness and substance use behaviours. Specifically, we examine whether daily cannabis use impacts on the hazard rate for becoming homeless for the first time, and vice versa, conditional on these unobserved factors and other observables. In an extension we examine whether the use of other illicit drugs at least weekly additionally impacts on the hazard rate for becoming homeless and vice versa. In the absence of valid instruments or natural experiments for substance use and homelessness, exploiting information about the *timing of transitions* into drug use and homelessness allows us to make causal inferences about how one impacts the other under plausible assumptions.

2 Data - Journeys Home

Whereas earlier studies of homelessness are based on specific groups of homeless we focus on a broader population. Rather than providing information about individuals who sleep rough or stay in emergency accommodation – a group Curtis et al. (2013) refer to as the acute homeless – in particular geographic areas, JH is a longitudinal dataset on disadvantaged Australians who are not necessarily homeless but are facing or have faced some form of housing instability during their life. Specifically, the JH sample comprises recipients of any income support (i.e. welfare) payment in Australia who are either homeless or at-risk of homelessness (Wooden et al., 2012).¹ We use the first three waves collected between September 2011 and November

¹ The latter category includes people flagged by Centrelink – the agency that administers welfare benefits in Australia – as at risk of homelessness and people identified by the research team as vulnerable to homelessness, i.e. persons that have not been flagged but nevertheless have characteristics similar to those that have been (using a propensity score matching approach). Local Centrelink office staff are required to flag customers they determine to be either 'homeless' or 'at risk of homelessness'. Individuals flagged as 'homeless' are without conventional accommodation (e.g. sleeping rough, squatting, or living in a car); or living in/moving frequently between temporary accommodation arrangements (e.g. with friends or extended family, emergency accommodation, or youth refuges). A person flagged 'at risk' of homelessness lives medium to long term in a boarding house, caravan

2012 focusing on the balanced panel, i.e. the 1347 respondents who were interviewed in all three waves of the survey and for whom, crucially, we have retrospective information on the onset of homelessness (collected in wave 1) and the onset of illicit drug use (collected in wave 3).² The fact that the JH survey targets a disadvantaged population makes it particularly well suited to analyse behaviours such as intense drug use and events such as homelessness onset which are seldom observed in the general population. The trade-off, of course, is that relationships observed in the JH sample may not generalise to the wider population.

Homelessness can be defined in different ways and with different thresholds. JH adopts a broad conceptualization of homelessness very similar to that under the US 2009 Homeless Emergency Assistance and Rapid Transition to Housing Act, and similar to that used by Link et al. (1994) and Curtis et al. (2013) in the US or Johnson and Chamberlain (2008) in Australia. Under this definition, homelessness is defined as sleeping rough or squatting in abandoned buildings; staying with relatives or friends temporarily with no alternative (doubling-up); or staying in a caravan, boarding house, hotel or crisis accommodation.

We construct for each respondent the age at which (s)he became homeless for the first time – using this definition but also exploring sensitivity to how homelessness is defined. If the respondent has been homeless prior to JH wave 1 we use the retrospective information collected at wave 1 on “How old [she was] the first time that [she was] without a place to live”.³ If the respondent was not homeless before JH we use the information collected at every wave on the type of accommodation the respondent lived in between waves to construct her homelessness status between those waves and from this we derive the age of first onset. Note that although we cannot rule out recall error in the reported age at which the individual first experienced

park or hotel, where accommodation is not covered by a lease; lives in accommodation which falls below the general community standards which surround health and wellbeing, such as access to personal amenities, security against threat, privacy and autonomy; is facing eviction; lives in accommodation not of an appropriate standard which may be detrimental to their physical and mental well-being, or where they have no sense of belonging or connection (e.g., Indigenous Australians living in crowded conditions or disconnected from their land, family / kin, spiritual and cultural beliefs and practices). For further details see Wooden et al. (2012).

² The JH sample contains 1,682 respondents in wave 1. Differences between the characteristics of the 3-wave balanced panel and wave 1 respondents are minor (see Melbourne Institute, 2013), and we find no statistically significant differences in sample means of observable controls between the two samples.

³ The relevant question asks ‘...have you ever stayed in any of the following places because you did not have a place to live?’ with the following options: stayed with relatives temporarily (because you did not have a place to live); stayed at a friend’s house temporarily (because you did not have a place to live); stayed in a caravan, mobile home, cabin, houseboat (because you did not have a place to live); stayed at a boarding house or hostel (because you did not have a place to live); stayed in a hotel or motel (because you did not have a place to live); stayed in crisis accommodation or a refuge; squatted in an abandoned building; slept rough (such as sleeping in cars, tents, trains or anywhere else outdoors) (because you did not have a place to live); other (because you did not have a place to live) (specify).

homelessness, systematic under-reporting of prior homelessness seems unlikely here given the nature of the survey. Only 7% of wave 1 respondents report never having been homeless.

Similarly, we use retrospective information collected in wave 3 about the age at which respondents first used cannabis daily or, in the extension, first used illegal/street drugs other than cannabis (at least weekly). Again we cannot rule out recall error in the reported age at which the individual first engaged in these behaviours. Neither can we entirely rule out systematic under-reporting of drug use of the kind discussed by Brown et al. (2018), although only 20% (52%) of the wave 1 sample report never having used cannabis (illegal drugs other than cannabis).

JH also collects information on a large number of potential childhood risk factors for illicit drug use and for homelessness. We use these as controls in our empirical model for homelessness onset (see Appendix 1). Sample means for these variables are reported in Table A1, separately by gender, and give a good indication of the disadvantaged nature of the JH sample. For example note the high proportions of respondents who report having experienced physical violence, emotional abuse, or sexual violence as a child (the latter particularly but not only for women). Also note high rates of substance abuse, long-term unemployment, hospitalisation for mental health problems and incarceration among caregivers.

Finally, note that although we use all sample members in our analysis, because we are primarily interested in relationships between early illicit drug use and youth homelessness we censor onsets which occur at ages older than 30 years, although we later explore sensitivity to this restriction.

3 The onset of homelessness and illicit drug use

Figure 1 presents hazard rates for first using cannabis daily and first episode of homelessness, i.e. the probability to start using regularly or to become homeless conditional on not having done so already. Hazard rates for using cannabis daily are especially high at younger ages, with spikes more pronounced for males than females. For instance, at age 16 years, 13% of males who have not already started using cannabis daily report starting daily cannabis use. The take-up of heavy cannabis use also peaks at 16 years old for females, with 8% reporting the onset of daily use. The onset of homelessness also peaks between the ages of 15 and 18 years (women) and 16 and 19 years (men), at above 10% in each year.

Figure 1: Transition rates into homelessness and daily cannabis use by age 30

Figure 2 shows the corresponding cumulative probability distributions for the uptake of daily cannabis use and the onset of homelessness, separately for males and females. It shows that the uptake of daily cannabis use increases sharply between age 13 and age 18 years for both men and women. It then levels off with few respondents starting after the age of 20.⁴

⁴ Uptake of *any* cannabis use (i.e. not restricted to daily use) shows an even steeper rise to around 70% for men

Figure 2: Cumulative starting probabilities for the uptake of daily cannabis use and the onset of homelessness by age 30

and 60% for women by age 18 years followed by a similar levelling off. Van Ours and Williams (2015) presents comparable figures drawn from survey data representative of all 25-50 year olds in Australia, showing slower and later take up – reaching around 50% rather than 75% by age 30 years – and with a less pronounced spike which occurs two years later than for the JH sample. McVicar et al. (2015) shows that the prevalence of other forms of substance use is also much higher among the JH sample than among the wider Australian population.

Note that males are much more likely than females to use cannabis daily. This gender imbalance in intensity of cannabis use is also found elsewhere (e.g. Van Ours et al., 2013). The onset of homelessness also increases sharply until around 18 years old for both genders. It then continues to increase at a slower pace up to age 30 years. In contrast to the take up of regular cannabis use both the proportions and the timing of the onset of homelessness are very similar for males and females.

Table 1 gives summary statistics for the prevalence rates and the age of onset of daily cannabis use and homelessness up to age 30 years for our sample. Almost all the JH sample experienced homelessness at some stage in their life – 97 percent of women and 98 percent of men – and 78 percent of women and 77 percent of men did so by age 30. On average, respondents who became homeless by 30 years old did so at an average of about 18 years old. Daily cannabis use is also highly prevalent in our sample with 33 percent of females and 56 percent of males reporting having done so by the age of 30.

Table 1: Prevalence and onset of homelessness and daily cannabis use by age 30

	Ever (%)		Mean age of onset	
	Men	Women	Men	Women
Homelessness	77.4	78.4	17.7	17.5
Cannabis daily	56.4	33.2	16.2	16.3
N	708	639	708	639

If the association between daily cannabis use and homelessness represents a causal relationship from drug use to homelessness, then the take-up of drug use should, on average, precede the onset of homelessness. If anything, for women, Figures 1 and 2 seem to suggest the opposite; the proportion experiencing homelessness onset appears to lead the proportions taking up regular cannabis use, and transitions into homelessness begin to peak prior to the peak in drug use. For men, Figures 1 and 2 are ambiguous in this respect. At the same time the average age of onset for cannabis use precedes that of homelessness (Table 1).

Table 2 takes a closer look at this question by tabulating the probability associated with the possible combinations of timing of events with respect to the onset of homelessness and daily cannabis use uptake in the raw data, separately by gender. It shows that 12% of females and 13% of males became homeless prior to using cannabis daily, that 6% of females and 6% of males became homeless in the same year as they started using cannabis daily, and that 11% of females and 28% of males first became homeless after they started using cannabis daily. In other words, daily cannabis use can occur before, coincident with or after the onset of

homelessness, with no discernible tendency for one to precede the other among young women, although there is a tendency for daily cannabis use to precede the onset of homelessness among young men.

Table 2: Association between the timing of homelessness and substance use by age 30

	Men	Women
Drugs before	28.3%	11.1%
Same age	6.2%	6.3%
Homelessness before	12.7%	12.4%
Not homeless, drugs	9.2%	3.4%
No drugs, homeless	30.2%	48.7%
Not homeless, no drugs	13.4%	18.2%
N	708	639

If drug use leads to homelessness for at least some of our sample then we might expect some respondents to report drug use as a factor in them becoming homeless when asked the following (wave 1) question: ‘What led to you being without a place to live the first time?’, for which ‘problematic drug or substance use’ is one of 13 specified options.⁵ For the wave 1 cross-section sample, Scutella et al. (2012) find that problem drug use was the fourth most common response, with 10% of those who had been homeless at some stage reporting this as a factor in their homelessness onset.⁶ Restricting to the 3-wave balanced panel and splitting the sample into those who first experienced homelessness prior or post age 30 years (see Table 3) shows that problem drug use is more frequently cited among the former group than the latter, particularly among women. Other commonly stated factors that differ markedly by age include relationship breakdown and domestic violence or abuse (more commonly cited among those first becoming homeless aged under 30 years) and financial difficulties and mental and other health issues (more commonly cited among those first becoming homeless over 30 years). These patterns are broadly in line with those reported for the smaller-sample study of Smith et al. (2008).

⁵ Multiple responses were permitted for this question.

⁶ The most common response was ‘relationship/family breakdown or conflict’ (62%), followed by ‘domestic and family violence or abuse’ (19%) and ‘financial difficulties’ (16%).

Table 3: Stated factors contributing to first homelessness

	Balanced panel	Homelessness before/at 30	Homelessness after 30
<u>Women</u>			
Financial difficulties	16%	12%	34%
Relationship breakdown or conflict	67%	72%	47%
Domestic and family violence or abuse	25%	26%	22%
Non-family violence	2%	2%	1%
Employment problems/unemployment	4%	3%	9%
Mental health issues	7%	6%	10%
Other health/medical issues	4%	3%	9%
Problematic drug or substance use	7%	8%	3%
Problematic gambling	1%	1%	0%
Transition from State Care	1%	2%	0%
Was evicted/asked to leave by landlord	7%	6%	13%
Natural disaster or fire	1%	1%	3%
End of lease	1%	1%	3%
Other	9%	9%	7%
N	584	468	116
<u>Men</u>			
Financial difficulties	15%	12%	28%
Relationship breakdown or conflict	62%	67%	44%
Domestic and family violence or abuse	15%	19%	3%
Non-family violence	1%	1%	1%
Employment problems/unemployment	8%	7%	15%
Mental health issues	7%	5%	16%
Other health/medical issues	4%	2%	11%
Problematic drug or substance use	13%	13%	12%
Problematic gambling	2%	2%	4%
Transition from State Care	2%	2%	0%
Was evicted/asked to leave by landlord	5%	3%	12%
Natural disaster or fire	1%	1%	1%
End of lease	1%	0%	1%
Other	10%	9%	12%
N	654	516	138

Note: this table includes 584 women and 654 men (i.e. less than the 639 women and 708 men in our sample) because 55 women and 54 men did not respond to this specific question.

4 Approach to Estimation: A Bivariate Simultaneous Durations Model

Our outcome variables are durations until transition from one state to another (the onset of homelessness and the onset of daily cannabis use) and the causal effects of interest relate to the realization of one transition on the transition rate of the other. Given the structure of our model these are established by the timing of events conditional on observable and unobservable characteristics. According to Abbring and Heckman (2007) using bivariate duration models to identify treatment effects of this kind goes back to Freund (1961), where the model was applied to engine failures in twin-engine planes. Abbring and Heckman (2007) discuss the additional complications when attempting to model *human* behaviour in this way. Whereas the failure rates of machines may be constant over time, *persons* may respond to the duration of the process in question. For example, an unemployed worker may lower his standards for accepting a job as unemployment duration increases. If so, the rate of transition from unemployment to a job may increase, and duration models need to account for this duration dependence. Furthermore, whereas machines may be very similar, unemployed persons may not be, i.e. they may differ in (observed or unobserved) characteristics that may have differential effects on their exit rate from unemployment. For example, more motivated unemployed workers are more likely to find a job quickly. The econometric framework of Abbring and Heckman (2007) accounts for this heterogeneity by including observable characteristics x as well as unobserved characteristics v , which are assumed to be temporally invariant, in a mixed proportional hazard (MPH) model set up (see Van den Berg, 2001). Finally, whereas the failure of one machine is a random event, individuals may anticipate events. For example, an unemployed worker who knows exactly when a training program starts may change his job search behaviour in anticipation of the start of the program. Therefore, in this setting the identification of treatment effects relies on a ‘no-anticipation’ assumption, which is set out formally, and discussed at length, in Abbring and van den Berg (2003) and again in Abbring and Heckman (2007).

In this paper we specify a bivariate simultaneous duration model for homelessness onset and onset of daily cannabis use. In this model, each transition may causally affect the other transition rate, i.e. the realization of one duration can be considered as a treatment that causally affects the other duration through its transition rate. A major advantage of using this kind of approach is that, as shown by Abbring and Van den Berg (2003), identification of the treatment effect does not rely on a standard conditional independence assumption –

we condition on both observables and jointly-distributed unobservables – and it is not necessary to have a valid instrument. Rather, identification comes from the timing of events, i.e. the order in which initiation into drug use and homelessness occurs. In this context the causal interpretation relies on the no-anticipation assumption and the estimation of unobserved heterogeneity.⁷ Although Abbring and Van den Berg (2003) demonstrate this for a bivariate duration model where one transition is (pre-)defined as the treatment and the other the outcome, Abbring and Heckman (2007) show that this naturally extends to the simultaneous case where each transition can impact the other. In all respects below we adopt the framework of Abbring and Heckman (2007).

Let $C(h)$ be the duration until the onset of daily cannabis use given the time of initiation into homelessness h , and let $H(c)$ denote the duration until the onset of homelessness given the time of initiation into daily cannabis use c . Assume ex ante heterogeneity across agents is fully captured by observed characteristics x and unobserved characteristics v , which are assumed to be external and temporally invariant. Initiation into homelessness may causally affect the duration $C(h)$ through its transition (i.e. hazard) rate. Similarly, initiation into daily cannabis use may causally affect the duration $H(c)$ through its hazard rate. Denote the hazard rate into daily cannabis use at time t for an individual with characteristics (x, v_c) as $\theta_c(t|x, h, v_c)$. Similarly the hazard rate into homelessness at time t for an individual with characteristics (x, v_h) is $\theta_h(t|x, c, v_h)$. The unobservables v_c and v_h are from an unrestricted joint distribution (v_c, v_h) in which the unobservables may have elements in common. Both observables x and unobservables v_c and v_h capture ex ante heterogeneity, i.e. they are given at the start of the processes. The inclusion of correlated unobserved heterogeneity is an important feature of bivariate duration models that allows them to account for time-invariant characteristics that, in our application, may lead individuals to be simultaneously more prone to substance use and to homelessness.⁸

The no-anticipation condition excludes anticipation of future outcomes, i.e. there can be no anticipation of outcome C on the hazard of H and there can be no anticipation of

⁷ Because of this advantage bivariate duration modelling has become an increasingly common approach in parts of the social policy literature, e.g. on the impact of benefit sanctions on welfare exit and job entry (see McVicar 2014 for a review). The bivariate duration approach has also been used in several studies of drug use impacts, most commonly to investigate various impacts of cannabis use (see Van Ours and Williams 2015 for a review). For applications to other questions see references in Abbring and Heckman (2007).

⁸ There might also be ex post shocks represented by exponential errors ε_c and ε_h . These shocks represent the randomness in the transition processes after conditioning on x , v and survival. Abbring and Heckman (2007) assume that $\varepsilon_c \perp \varepsilon_h$, so that the potential outcomes are only dependent through the observed and unobserved characteristics (x, v) .

outcome H on the hazard of C (Abbring and Heckman, 2007). Conditional on observed characteristics and the distribution of unobserved heterogeneity, current hazards therefore depend only on past events and the transition processes evolve recursively. Formally, for all $t \in R_+$:

$$\theta_c(t|x, h, v_c) = \theta_c(t|x, h', v_c) \text{ and } \theta_h(t|x, c, v_h) = \theta_h(t|x, c', v_h)$$

For all $h, h', c, c' \in [t, \infty)$. Under this assumption Abbring and Heckman (2007) show that such models are non-parametrically identified from single-spell duration data under the conditions for the identification of competing risks models based on multivariate MPH models.

The no-anticipation assumption implies that although individuals may have an *expectation* of an outcome occurring, because they cannot foresee the exact time it will occur they do not change their (other outcome-relevant) behaviour in anticipation of it. In our specific case, it requires that, conditional on x and v , an individual does not alter her behaviour relevant to cannabis consumption because she knows that she will become homeless in a particular future year, and vice versa. We have already shown in Table 3 that homelessness onset is reportedly driven by a wide variety of factors, the most common of which involve actions by others (e.g. family members and carers, landlords, employers) which are not perfectly foreseeable by individuals and therefore not likely to change their behaviour in relation to cannabis use in advance. Similarly, the timing of initiation into regular drug use is likely to depend on imperfectly predictable factors such as the degree to which occasional drug use leads to addiction, which seem equally unlikely to change behaviour with respect to homelessness in advance.

More generally, the violation of the no-anticipation assumption entails that individuals: (i) have information regarding their future transition into the treatment (respectively homelessness and daily cannabis use); (ii) know precisely when this transition will occur; (iii) alter their behaviour with respect to the other outcome of interest in anticipation (respectively daily cannabis use and homelessness) and (iv) that what led them to anticipate their transitions is not accounted for in the model (via observed or unobserved characteristics). It is in this sense that the no-anticipation assumption is arguably weaker than a standard conditional independence assumption (for a more detailed discussion see Lalive et al., 2008).

The hazard rate of initiation into daily cannabis use at duration t conditional on observed and unobserved characteristics and whether or not an individual has been homeless before t , is given by:

$$\begin{aligned}
\theta_c(t|x, h, v_c) &= \lambda_c(t) \exp(x' \beta_c + v_c) \quad \text{for } t \leq h \\
\theta_c(t|x, h, v_c) &= \lambda_c(t) \exp(x' \beta_c + \delta_h + v_c) \quad \text{for } t > h
\end{aligned} \tag{1}$$

The hazard rate of initiation into homelessness at duration t conditional on observed and unobserved characteristics and whether or not an individual has started using cannabis daily before t , is given by:

$$\begin{aligned}
\theta_h(t|x, c, v_h) &= \lambda_h(t) \exp(x' \beta_h + v_h) \quad \text{for } t \leq c \\
\theta_h(t|x, c, v_h) &= \lambda_h(t) \exp(x' \beta_h + \delta_c + v_h) \quad \text{for } t > c
\end{aligned} \tag{2}$$

The effect of previous cannabis use on the onset of homelessness is measured by δ_c . This is the key parameter of interest as it informs us as to whether previous cannabis use increases the risk of homelessness ($\delta_c > 0$), reduces the risk of homelessness ($\delta_c < 0$), or has no direct effect on the likelihood of experiencing homelessness ($\delta_c = 0$). The other key parameter is the impact of previous homelessness onset on initiation into daily cannabis use, given by δ_h . The baseline hazards $\lambda_c(t)$ and $\lambda_h(t)$ capture duration dependence of individual initiation rates. Furthermore, β_h and β_c are vectors of parameters capturing the effects of observable characteristics on the homeless initiation rate and the initiation rate into cannabis.

In modelling the uptake of daily cannabis use, we assume that potential exposure to drugs occurs from the age of 10. We model transitions up to and including age 30 to capture early onset of drug use and early onset of homelessness. Individuals who have not started using cannabis daily by age 25 are very unlikely to do so later in their life (Van Ours, 2006a) and, as discussed in the previous section, homelessness onsets in later life appear to be driven by a different set of factors than earlier onsets.

We model duration dependence in the cannabis initiation hazard in a flexible way using a step function $\lambda_c(t) = \exp(\sum_k \lambda_{c,k} I_k(t))$, where $k = (1, \dots, 10)$ is a subscript for age categories and $I_k(t)$ are time-varying dummy variables that are one in the relevant category. We specify ten age dummies, seven of which are for individual ages (age 12, . . . , 18), while the first one is for ages less than 12, the second last one for ages between 19 to 21 and the last interval is for ages from 22 years onward up to 30 years. Because we also estimate a constant term, we normalize $\lambda_{c,1} = 0$.⁹

The conditional density function for the completed durations until the uptake of daily cannabis use can be written as:

⁹ As we only know the age at which each event first occurs and not the actual date, we are unable to determine whether homelessness occurred first if both the onset of homelessness and substance use occurred at the same age. It is for this reason that we allow homelessness to impact on substance uptake if and only if it occurred in a previous period.

$$f_c(t|x, h, v_c) = \theta_c(t|x, h, v_c) \exp(-\int_0^t \theta_c(s|x, h, v_c) ds) \quad (3)$$

Individuals who have not used cannabis by the last age they are observed in the survey are assumed to have a right-censored duration of non-use.

In the initiation to homelessness hazard, $\lambda_h(t)$ represents individual duration dependence which is modelled using a step function which is specified in the same way as for substance use. The conditional density function for the completed duration until first homelessness can be written as:

$$f_h(t|x, c, v_h) = \theta_h(t|x, c, v_h) \exp(-\int_0^t \theta_h(\sigma|x, c, v_h) d\sigma) \quad (4)$$

Individuals who have not experienced homelessness by the age at which they are last observed in the data are assumed to have a right-censored duration until the onset of homelessness.

The potential correlation between the unobserved components in the hazard rates for substance uptake and homelessness is considered by specifying the joint density function for both durations of time until daily cannabis use c and the duration of time until homelessness h conditional on x as:

$$f(c, h|x) = \int_{v_h} \int_{v_c} f_c(t|x, h, v_c) \cdot f_h(t|x, c, v_h) dG(v_h, v_c) \quad (5)$$

As is standard in recent applications of bivariate duration models, $G(v_h, v_c)$ is assumed to be a flexible discrete distribution with an unknown number of points of support. We start by assuming that for every transition process its unobserved heterogeneity can be specified by a discrete distribution of with two points of support. In combination this leads to four points of support: $(v_{h1}, v_{c1}), (v_{h1}, v_{c2}), (v_{h2}, v_{c1}), (v_{h2}, v_{c2})$, reflecting two types of individuals in the hazard rates for cannabis use (high susceptibility and low susceptibility) and two types in the hazard rate for homelessness (high susceptibility, low susceptibility). The four mass points imply that conditional on observed characteristics there are four types of individuals. The associated probabilities are denoted as follows:

$$\begin{aligned} \Pr(v_h = v_{h1}, v_c = v_{c1}) &= p_1, \Pr(v_h = v_{h1}, v_c = v_{c2}) = p_2, \\ \Pr(v_h = v_{h2}, v_c = v_{c1}) &= p_3, \Pr(v_h = v_{h2}, v_c = v_{c2}) = p_4 \end{aligned} \quad (6)$$

with $0 \leq p_j \leq 1$ for $j = 1, \dots, 4$. These probabilities are modelled using a multinomial logit specification, i.e. $p_j = \exp(\alpha_j) / (\sum_j \exp(\alpha_j))$ and we normalize $\alpha_4 = 0$. Furthermore, because we also estimate constants we normalize $v_{h1} = v_{c1} = 0$.

The parameter estimates are obtained using the method of maximum likelihood considering that our duration information relates to intervals rather than to exact durations.

For example, an individual who indicated to have become homeless at age 16 may have become homeless on his 16th birthday or on the day before his 17th birthday. For this individual, we model that he did not yet start at age 15, but started before turning 17.

A graphical explanation of this model is presented in the Appendix (Figures A1 and A2).

The control variables included in our model are listed in Appendix 1 and described in Table A1. They include three dummy variables for whether the respondent was not living with his parents at age 14 because they were separated, because they were dead, or because of conflict; three dummy variables for whether the respondent experienced emotional, physical or sexual abuse as a child; five dummy variables characterising the behaviour of the main male and female caregivers of the respondent while growing up (substance abuse, incarceration, mental health problems, long-term unemployment and gambling issues). So, for example, if the male caregiver ever spent time in jail, this variable is coded 1, 0 if he did not spend time in jail, if the respondent had no male caregiver or if the information on the caregivers' jail time was missing. Because some of this background information is missing for a significant portion of our sample (in some cases more than 10 percent), and because this is unlikely to be random, we also include dummy variables for missing information on control variables: one dummy for missingness on any of the reasons for not living with parents at age 14; one dummy for missingness on any of the childhood violence variables; one dummy for missingness on any of the male caregiver's information and one for the female caregiver's information.¹⁰ The missing dummies for the caregiver characteristics are coded 1 if the respondent had missing information on the presence of a caregiver or if any of the caregivers' characteristic is missing, 0 if none is missing. As a result, the caregivers' variables capture the effect of having a caregiver with a certain negative characteristic (jail time, substance use, long-term unemployment, mental health issues, gambling issues) relative to caregivers with no known such issues or no caregiver.¹¹

¹⁰ Alternatives, including dropping observations for which there are missing data, are explored in a sensitivity analysis discussed in Section 5.2.

¹¹ Note that including specific dummy variables for respondents with no male (resp. female) caregiver to distinguish between those two reference categories does not alter the results. Indeed, the percentage of the reference category pertaining to "no caregiver" is relatively small (depending on the behaviour analysed, between 2.5% and 4.5% of the reference category do not have a female caregiver and between 14% and 19% do not have a male caregiver). And having "no caregiver" is very correlated with not living with one's parents at age 14 which we control for and appears to be the major divide in our sample: 46% of respondents are not living with their parents at age 14 but among those, only 3% cannot identify a female caregiver and 19% cannot identify a male caregiver.

5 Results and Discussion

5.1 Bivariate model for daily cannabis use and first homelessness onset

Table 4 presents estimates of the bivariate model for initiation into daily cannabis use and homelessness onset, allowing for causality in both directions. First we find evidence of the presence of unobserved heterogeneity. We are able to identify four mass points of which the distribution is given at the bottom of the table. Thus conditional on observed characteristics and age we can distinguish four types of individuals with different predispositions to daily cannabis use and homelessness. Among men the largest group of 60 percent has a high predisposition to daily cannabis use and homelessness while 11 percent has two low predispositions. This implies that for 71 percent of the men there is a positive correlation in unobserved characteristics affecting daily cannabis use and homelessness. Failing to take this into account would lead to a spurious effect of cannabis use on homelessness or vice versa. The distribution of unobserved heterogeneity is different for women for whom the largest group of 41 percent combines a low predisposition to daily cannabis use with a low predisposition to homelessness. However since conditional on observed characteristics and age 28 percent of the women combine positive predispositions to homelessness and daily cannabis, the majority of women (69%) also have a positive correlation in unobserved characteristics.

Now consider impacts on homelessness onset for young men. Taking up daily cannabis use is associated with a higher hazard rate for homelessness onset than that of otherwise equivalent men who do not use cannabis daily by a factor of 1.7, and is statistically significant at the 1% level.¹² Under the assumptions set out in the previous section this can be plausibly interpreted as a causal effect. Possible causal mechanisms for such an effect include breakdown of family relationships and financial strain resulting from drug use (qualitative evidence for both is presented by Johnson and Chamberlain, 2008). These estimates provide the strongest quantitative evidence to date on the impact of cannabis use on the onset of homelessness. They also add evidence of a further negative social outcome to the literature on the causal impacts of cannabis use.

The magnitude of this effect is similar to that of not living with one's parent because they are separated, having experienced emotional abuse as a child, or having a female caregiver that had spent time in hospital because of mental health problems (all have similar signs and

¹² The proportional impact on the hazard rate – the hazard ratio – is given by $\exp(\beta)$, i.e. $\exp(0.53)$. Also note that measurement error in the age of onset of daily cannabis use, whether stemming from systematic under-reporting or random recall error, may impart a downwards bias on this estimate.

magnitudes and are statistically significant at 1%). The strongest association with homelessness onset among this sample comes from not living with one's parents at age 14 because of conflict, which almost quadruples the hazard rate for homelessness onset. This is consistent with the high proportion of young men who cite this as a reason for first becoming homeless as presented in Table 3. Other factors with smaller magnitude but statistically significant impacts are being orphaned by age 14, having a long-term unemployed male caregiver, and having a female caregiver with a substance abuse problem. These covariate parameter estimates reflect findings in the existing homelessness literature concerning the importance of adverse childhood experiences as predictors of homelessness (e.g. see van den Bree et al., 2009; Shelton et al., 2009, and references therein).

We find little evidence of reverse impacts – from homelessness onset to initiating daily cannabis use – for young men. The point estimate is actually negative, relatively small and only marginally statistically significant (at 10% but not 5%). (We show below that in most robustness tests it is negative, similar or smaller in magnitude, and statistically insignificant.) The strongest association with the take-up of daily cannabis use is again conflict with parents (which triples the hazard rate), followed by experiencing physical violence as a child (which doubles the hazard rate). Other statistically significant factors include a male caregiver with substance abuse problems or who has spent time in prison. Data that allow a quantitative analysis of the impacts of adverse childhood experiences such as these on cannabis use are rare, but existing studies have found a positive association between parental substance use and own drug use (e.g. DeSimone 2002; Abelson et al., 2006).

The results for women are almost the polar opposite of those for men. We see no evidence of an impact of daily cannabis use on homelessness onset, but clear evidence of a reverse effect from homelessness to daily cannabis use (the corresponding hazard ratio is 1.55). This gender contrast is to some extent visible in the raw data presented in Table 2, where homelessness onset more often precedes drug use uptake for women whereas the opposite is the case for men. Conflict with parents at age 14 is again the variable with the strongest association with both outcomes. As for males, not living with one's parents at age 14 because they are separated, emotional abuse, physical violence, male caregiver substance abuse and a female caregiver with prison time or hospital time for mental health problems are also associated with higher hazards for homelessness onset and/or take-up of daily cannabis use. Note that reporting a male caregiver with a gambling problem again takes a negative sign. Experiencing sexual violence as a child is an additional factor that increases the hazard for both homelessness onset and daily cannabis use for young women but not young men, as is missing data about the male caregiver.

Table 4: Bivariate simultaneous mixed proportional hazards model, daily cannabis use and homelessness by age 30, coefficients (absolute t-statistics)

	Men		Women	
	Homelessness	Cannabis Daily	Homelessness	Cannabis Daily
Effect cannabis daily on	0.53 (3.7)***	-	-0.13 (0.7)	-
Effect homelessness on	-	-0.29 (1.7)*	-	0.44 (2.2)**
<i>Childhood</i>				
Parents separated	0.60 (4.3)***	0.32 (2.3)**	0.68 (4.6)***	1.03 (4.3)***
Parents dead	0.44 (1.8)*	0.02 (0.1)	0.46 (1.6)*	-0.25 (0.6)
Conflict parents	1.36 (6.1)***	1.10 (4.0)***	1.52 (7.7)***	2.26 (6.9)***
Emotional abuse	0.59 (3.3)***	0.10 (0.5)	0.79 (3.4)***	0.38 (1.0)
Physical violence	0.19 (1.0)	0.79 (3.9)***	0.30 (1.3)	0.87 (2.2)**
Sexual violence	0.13 (0.7)	0.13 (0.7)	0.25 (1.7)*	0.56 (2.1)**
<i>Male caregiver</i>				
Substance abuse	-0.11 (0.7)	0.36 (2.4)**	-0.14 (0.8)	0.56 (2.2)**
Jail	-0.01 (0.0)	0.37 (1.8)*	0.24 (1.1)	0.18 (0.4)
Hospital	-0.49 (1.5)	-0.09 (0.3)	0.02 (0.1)	0.07 (0.2)
Unemployed	0.37 (2.1)**	0.19 (1.0)	0.25 (1.2)	-0.22 (0.8)
Gambling	0.12 (0.5)	-0.50 (2.2)**	-0.30 (1.3)	-1.18 (3.1)***
<i>Female caregiver</i>				
Substance abuse	0.41 (2.1)**	0.25 (1.3)	0.11 (0.5)	0.14 (0.5)
Jail	-0.34 (0.6)	0.56 (1.4)	0.72 (2.1)**	1.89 (3.4)***
Hospital	0.70 (3.1)***	0.23 (1.1)	0.72 (3.3)***	0.93 (3.3)***
Unemployed	0.19 (1.5)	0.19 (1.3)	0.24 (1.6)	-0.21 (0.9)
Gambling	-0.09 (0.4)	-0.06 (0.2)	-0.17 (0.7)	0.73 (2.0)**
<i>Missing info</i>				
Reason	0.61 (1.2)	-0.01 (0.0)	1.47 (2.5)**	0.04 (0.0)
Violence	0.33 (1.5)	0.19 (0.9)	0.32 (1.5)	0.09 (0.2)
Male caregiver	0.11 (0.5)	0.32 (1.5)	0.76 (3.4)***	0.93 (2.5)**
Female caregiver	0.12 (0.6)	-0.19 (0.8)	0.38 (1.8)*	-0.28 (0.6)
Constant	-6.57 (15.1)***	-5.40 (18.2)***	-7.57 (14.8)***	-6.47 (10.8)***
Age 12	-0.20 (0.6)	0.65 (1.7)*	0.87 (2.6)***	1.03 (1.9)*
Age 13	0.25 (0.8)	1.61 (5.2)***	1.30 (4.0)***	2.03 (4.2)***
Age 14	1.04 (4.1)***	2.01 (6.7)***	2.08 (7.2)***	2.88 (6.1)***
Age 15	1.61 (6.8)***	2.31 (7.7)***	2.87 (10)***	3.09 (6.3)***
Age 16	2.08 (8.9)***	2.84 (9.6)***	3.18 (11)***	3.96 (7.9)***
Age 17	2.35 (9.6)***	2.78 (9.0)***	3.38 (11)***	3.70 (6.9)***
Age 18	2.22 (8.2)***	2.68 (7.8)***	3.48 (10)***	3.95 (7.0)***
Age 19-21	1.95 (7.0)***	2.02 (5.9)***	2.75 (7.6)***	2.82 (4.5)***
Age 21+	1.79 (5.7)***	1.31 (3.4)***	2.99 (7.8)***	3.01 (4.8)***
ν_2	1.97 (7.3)***	$-\infty$	1.93 (7.3)***	-4.34 (11.4)***
α_1	1.24 (1.6)		-2.31 (3.1)***	
α_2	0.56 (0.7)		0.01 (0.1)	
α_3	2.22 (3.1)***		0.37 (1.3)	
-Loglikelihood	3206.3		2449.7	

Distribution of unobserved heterogeneity (%):

	Cannabis daily				Cannabis daily		
	High	Low	Total		High	Low	Total
Homeless - Low	22	11	33	Homeless Low	3	41	44
Homeless - High	60	7	67	Homeless High	28	28	56
Total	82	18	100	Total	31	69	100

Note: absolute t statistics in parentheses; ***, **, * indicates significance at a 1%, 5% and 10% level.
Sample size: 639 women and 708 men.

Ours is not the first study to find evidence of stronger impacts of cannabis use for men than for women on social outcomes. Both DeSimone (2002) and Van Ours (2006b) find a weaker

association for women than for men between cannabis or cocaine use and subsequent employment outcomes, which the latter speculates may reflect gender-specific unobservables such that women who were inclined to explore the drug scene in the past are more ambitious to find a job in the present.

Other studies have also flagged up potential gender differences specifically in the nature of the association between drug use and homelessness which are consistent with our results. For example, Smith et al. (2008) find that fewer women than men cite substance use as a reason for becoming homeless. This is consistent with the evidence from JH presented in Table 3. Instead women cite domestic violence much more frequently, both in the JH study and elsewhere (e.g. Lehmann et al., 2007). Others have suggested that the tendency for women to have stronger social networks than men may act as a protective factor (Susser et al., 1993; Bassuk et al., 1997). Other potential protective factors that may play a bigger role for women than for men might include having caring responsibilities (and associated cash and other welfare benefits) for children (Bassuk et al., 1997).

Possible causal mechanisms for the reverse effect from homelessness to cannabis use include adaptation to a subculture of substance use among the homeless and/or using substances as a coping mechanism (Johnson and Chamberlain, 2008). Kidd (2007) discusses whether young homeless women face greater adversity or perceive greater social stigma than young homeless men on average, which would be consistent with young homeless women disproportionately turning to drug use via one or both the adaptation and coping mechanisms. Evidence suggestive of an impact of youth homelessness on drug use, both qualitative and quantitative, has been presented elsewhere in the homelessness literature (e.g. Martijn and Sharpe, 2006; Johnson and Fendrich 2007; Johnson and Chamberlain, 2008). None of the existing quantitative studies, however, satisfactorily deals with selection on unobservables. Further, although these studies exploit data on both men and women, they do not present evidence on gender differences in the nature of adaptation to homelessness.

5.2 Extensions and sensitivity analysis

Table 5 presents the key estimates from several variations of the model, separately for men and women. We consider each in turn below.

5.2.1 Censoring at age 25 or age 35 years

Row 1a in Table 5 shows that censoring at age 25 years – given earlier evidence that drug use initiation generally occurs prior to this age (Van Ours 2006a) – has little impact on our estimates. This also holds for censoring at age 35 years (row 1b of Table 5), at least for men. For women, the impact of daily cannabis use on homelessness onset is larger (hazard ratio 1.36) and just statistically significant at 5% when individuals are censored at age 35 years rather than age 30 years (the reverse effect is also stronger). This may signal a stronger relationship between cannabis use and homelessness onset among women in their early 30s than among younger women. Nevertheless, the majority of our estimates suggest this effect is zero, and where it is not zero it is always smaller in magnitude than the corresponding effect for men.

5.2.2 Setting the reverse causality effect to zero

Row 2a of Table 5 shows the estimated effect of daily cannabis use when the effect of homelessness on cannabis use is assumed to be zero, i.e. when we remove the possibility of reverse causality. In this case, the effect of cannabis use on homelessness is slightly overestimated for both gender and this model is rejected for women (LR-test = 3.4, 1 df), consistently with a very significant effect of homelessness on cannabis use for women. Row 2b then imposes the effect of cannabis use on homelessness to be zero. In this case, the effect of homelessness on daily cannabis use is again overestimated for both gender, especially for men for whom this model is rejected (LR-test = 12.8, 1df), consistently with a very significant effect of cannabis use on homelessness in the baseline (0.53***). This evidence suggests that accounting for reverse causality improves estimates in our setting where homelessness and daily cannabis use are correlated but where for each gender, the effect mostly goes in one direction.

5.2.3 Single equation duration models

Row 3 of Table 5 shows that the estimated association between daily cannabis use and homelessness onset for men is stronger in a single equation model than when we account for correlated unobservables as in the baseline case. Similarly the estimated association between homelessness onset and daily cannabis use for women is stronger. There is also now a large and statistically significant estimated effect of cannabis use on homelessness onset for women. The estimated effect of homelessness onset on cannabis use for men is again zero. In other words, accounting for correlated unobservables as in the baseline case both quantitatively and qualitatively affects our conclusions, and we substantially over-estimate the magnitude of the causal relationships between homelessness and cannabis use when we ignore these common factors.

5.2.4 A Trivariate model for daily cannabis use, weekly street drug use and homelessness

The use of cannabis, and cannabis daily, is strongly positively associated with the use of other illicit drugs. Figure A3 in the Appendix illustrates this for our sample. Only 15.8 percent of our male sample and 26.1 percent of our female sample never engaged in any of these behaviours by age 30. Second, very few of those who report having used illicit/street drugs, and none of those who report having used illicit/street drugs at least weekly, fail to report having also used cannabis. In contrast, there are many cannabis users, including daily cannabis users, who do not report having used other illicit/street drugs¹³. These interrelationships suggest that modelling only daily cannabis use risks capturing some aggregated effect of multiple substance use behaviours. We therefore extend the analysis by estimating a *trivariate* duration modelling framework for both drug use behaviours and homelessness together. In other words we model all three transitions as a fully simultaneous system in which the unobserved heterogeneity terms entering the transition rates are allowed to be correlated. To be specific, uptake of both prior drug use behaviours now enters the hazard for transitions into homelessness and vice versa. For the unobserved heterogeneity, we continue to assume a discrete distribution with two points of support for each transition process, which in combination leads to a maximum of eight points of support (types). The key parameter estimates are reported in row 4 of Table 5.

For both men and women the extension to the trivariate model has little effect on the estimated impact of daily cannabis use on homelessness, which remains positive and of similar magnitude and statistical significance for men and zero for women, or on the estimated impact of homelessness on daily cannabis use, which remains zero for men but positive with increased magnitude and statistical significance for women. Take-up of weekly other illicit drug use has no additional effect on homelessness onset for either young men or young women; as shown in Table 2 it is relatively rare for uptake of other illicit drugs to precede homelessness onset in our sample. Neither is there any reverse effect from homelessness to other illicit drug use for men. But, like for daily cannabis use, there is evidence of a reverse effect from homelessness to weekly illicit street drug use for women. Although it is imprecisely estimated, the estimated magnitude of this effect is large (it more than doubles the hazard rate), again suggesting adaptation to a subculture of substance use among the homeless and/or substance use as a coping mechanism among young homeless women but not young homeless men.

¹³ Appendix 3 also provides some descriptive statistics on the use of illegal/street drugs in our sample.

Table 5: Extensions and sensitivity analysis, coefficients (absolute t-statistics)

	Effect of daily cannabis [weekly street drug] use on homelessness	Effect of homelessness onset on daily cannabis [weekly street drug] use	-Loglikelihood
<i>Men</i>			
1. Baseline	0.53 (3.7)***	-0.29 (1.7)*	3206.3
1a. Age ≤ 25	0.61 (4.2)***	-0.28 (1.6)*	2983.7
1b. Age ≤ 35	0.68 (6.0)***	-0.10 (0.6)	3293.3
2a. No effect of homelessness	0.63 (4.4)***	0 (-)	3207.4
2b. No effect of cannabis	0 (-)	-0.51 (3.3)***	3212.7
3. Single equation estimates	0.75 (5.9)***	-0.06 (0.4)	3208.5
4. Trivariate cannabis	0.52 (4.9)***	-0.09 (0.5)	
Trivariate weekly street drug	0.08 (0.4)	-0.24 (0.5)	3690.8
5a. Acute homelessness	0.80 (5.0)***	-0.10 (0.6)	2883.9
5b. Acute unsheltered homelessness	0.88 (5.3)***	-0.03 (0.2)	2789.9
<i>Women</i>			
1. Baseline	-0.13 (0.7)	0.44 (2.2)**	2449.7
1a. Age ≤ 25	0.05 (0.2)	0.55 (2.5)***	2243.0
1b. Age ≤ 35	0.31 (2.0)**	0.90 (4.6)***	2538.7
2a. No effect of homelessness	-0.22 (1.2)	0 (-)	2451.4
2b. No effect of cannabis	0 (-)	0.65 (3.1)***	2449.7
3. Single equation estimates	1.16 (5.5)***	0.66 (3.7)***	2456.1
4. Trivariate cannabis	0.07 (0.3)	0.70 (3.2)***	
Trivariate weekly street drug	0.03 (0.1)	0.86 (1.8)*	2766.3
5a. Acute homelessness	-0.31 (1.3)	0.40 (1.9)*	2035.3
5b. Acute unsheltered homelessness	0.17 (0.6)	0.66 (2.7)***	1771.7

Note: absolute t statistics in parentheses; ***, **, * indicates significance at a 1%, 5% and 10% level.

Sample size: 639 women and 708 men (except for 5a and 5b: 638 women and 707 men).

5.2.5 Sensitivity to the definition of homelessness

The definition of homelessness used here is quite broad – for example including those ‘doubling up’ with friends or family – for two reasons. First, as discussed in Section 2 this definition has been widely adopted in the literature and in policy. Second, the JH data only record the age of onset for this broad definition of homelessness, not for particular types of homelessness. Nevertheless in Table 5 row 5a we report estimates from the bivariate duration model where we restrict those we treat as having experienced homelessness onset to those who report having been ‘acute’ homeless at some stage in their life (sleeping rough or squatting in abandoned buildings, or in crisis accommodation). This is not the age of onset of acute homelessness per se – this is not recorded in the data – but it does wash out those who never experienced this kind

of homelessness. Finally row 5b in Table 5 shows the relevant parameter estimates if we focus on acute unsheltered homelessness (sleeping rough or squatting in abandoned buildings) in the same way. Our conclusions remain qualitatively unchanged, although the magnitude of the effect of daily cannabis use on homelessness onset for men increases in both cases.

5.2.6 Sensitivity to the treatment of missing information

Around 26% of the sample has missing information for at least one of the variables used as controls. This is a significant proportion of our sample and given the information contained in these variables (caregivers' substance abuse, incarceration, violence as a child...), this missing information is unlikely to be random and is possibly correlated with the respondents' substance use and homelessness onsets. Therefore, our main estimates presented in Tables 4 and 5 are based on models retaining the full sample and including 'missing' categories for these variables as additional control variables. Because we cannot rule out that this approach to missing data may impart bias to our estimates (Greenland and Finkle, 1995), and because other approaches are possible (although, short of collecting better data, none are ideal), a selection of simple alternative approaches to missing data is presented in the Appendix 4. These alternatives drop observations with missing information for certain variables and/or drop variables from the model.

Overall, the alternatives considered support our results. In all cases the positive and statistically significant impact of cannabis use on homelessness onset for young men remains. In most cases, the positive effect of homelessness on cannabis for women is also retained, and the magnitude of our main estimate (0.44) is in the middle range of all alternatives considered..

5.2.7 Magnitude of the effects in the baseline bivariate model

To illustrate the magnitude of the effects of daily cannabis use on the onset of homelessness for men, the results from a simulation exercise are presented in Table 6. For these simulations, we use the parameter estimates of the first column of Table 4. The first column of Table 6 shows the cumulative probability to have become homeless at various ages up to age 30, for a reference person in panel a and a respondent who was not living with his parents at age 14 because of conflict in panel b. For the reference person all the explanatory variables are set at zero, so only the constant, the age dependence parameter and the unobserved heterogeneity distribution is used in the simulations. The reference man has not used cannabis, did not report a substantially troubled childhood, and did not have caregivers with unfavourable behaviour. The reference person had a probability of 7 percent to have been homeless by age 15. By age 20 this is 29

percent, by age 25 it is 41 percent and by age 30 it is 50 percent. The second column gives the simulation results if the person with the reference characteristics started using cannabis daily at age 20. Then, by age 30 this person had a 58 percent probability to have been homeless at least once. If the reference person started using cannabis daily at age 15 the corresponding probability is 64 percent (column 3).

Table 6 Simulations of the cumulative probability (%) of homelessness by age 30 (men)

a. Reference person

	No cannabis	Cannabis from age 20	Cannabis from age 15
Age 15	7.3 (1.2)	7.3 (1.2)	7.3 (1.2)
Age 20	29.4 (3.1)	29.4 (3.1)	40.7 (4.3)
Age 25	41.2 (3.6)	46.8 (3.6)	54.8 (4.4)
Age 30	49.8 (3.9)	58.3 (3.9)	63.6 (4.2)

b. Conflict with parents

	No cannabis	Cannabis from age 20	Cannabis from age 15
Age 15	24.8 (5.7)	24.8 (5.7)	24.8 (5.7)
Age 20	66.0 (7.0)	66.0 (7.0)	75.6 (5.9)
Age 25	75.9 (5.3)	79.3 (4.6)	83.4 (4.3)
Age 30	80.9 (4.3)	85.1 (3.8)	87.6 (3.7)

Note: Simulations based on the parameter estimates of Table 4 first column. Reference individuals: all observed characteristics have the reference value. Standard errors in parentheses.

Panel b shows similar simulations for a respondent who experienced conflict with parents during childhood. The first column shows that for this man the cumulative probability to have experienced at least one spell of homelessness by age 30 is 81 percent. In other words, the effect of a conflict with parents during childhood is larger than the effect of starting daily cannabis use at age 15. However, starting to use cannabis daily at age 15 has a substantial additional effect on the probability to experience homelessness. By age 30 this is 88 percent. Clearly, daily cannabis use has a substantial effect on the onset of homelessness but it is just one among the many potential determinants some of which have an even bigger estimated effect.

6 Concluding remarks

A recent point-in-time estimate for the United States suggested almost 600,000 individuals (0.2% of the US population), a third of whom were aged under 25 years, were sleeping rough on the streets or in shelters on a particular night in January 2014 (Department of Housing and Urban Development 2014a). This estimate more than doubles to 1.42 million if the count is of people housed in shelters (not including unsheltered homeless) at any time during the year to

September 2013 (Department of Housing and Urban Development, 2014b). Were such estimates to cover a longer period or to include those ‘doubling up’ with family or friends or in other forms of insecure housing – in line with the definition of homelessness used for the 2009 Homeless Emergency Assistance and Rapid Transition to Housing Act and the broader definition used here in this paper – they would be higher still. For example Link et al. (1994) estimated that 14% of the US population had been homeless at some point during their lives using such a definition.

Homelessness deprives young people of a basic human need (Curtis et al., 2013) and may put them at greater risk of a range of negative health and social outcomes (e.g. Greene et al., 1997). This motivates policy interventions with the potential to reduce youth homelessness. This in turn motivates research to improve our understanding of the factors that increase the risk of youth homelessness. The hurdle for such research to overcome is distinguishing what are likely to be causal relationships with the potential for policy intervention from non-causal associations. The vast majority of the quantitative homelessness literature arguably falls short of this hurdle. Here we make a significant contribution to this literature by showing there exists a strong adverse effect of regular cannabis use on the probability of young men becoming homeless, conditional on a wide range of observables, other drug use behaviours, and common unobservables. We also make a significant contribution to the ‘drug use impacts’ literature by providing credible evidence that cannabis use impacts adversely on male housing outcomes alongside earlier studies showing evidence of adverse effects on health, educational and (in some cases) labour market outcomes. We find no substantial drug use impact for young women, although we do find evidence of a reverse effect running from homelessness to drug use (both cannabis and other illicit street drugs). Although we are unable to tease out the particular causal mechanism(s) for this effect, the gender contrast suggests that the homelessness impacts of cannabis use work through social rather than chemical changes.

With the usual caveat about the extent to which these conclusions will generalise across contexts, the main policy implication is that interventions to reduce or delay the onset of cannabis use among young men can have positive impacts not only on physical and mental health and educational attainment, but can also reduce or delay the onset of youth homelessness. Further, interventions that reduce or delay entry to homelessness among young women can help to reduce or delay the onset of illicit drug use among young women. Finally, there is a tentative indication here that targeting both types of interventions – to reduce/delay drug use uptake and to reduce/delay homelessness onset – by gender may be something to consider if the gender pattern of results we find here can be replicated using other methods and other data.

Acknowledgements

This paper uses unit record data from Journeys Home: Longitudinal Study of Factors Affecting Housing Stability (Journeys Home). The study was initiated and is funded by the Australian Government Department of Social Services (DSS). The Department of Employment has provided information for use in Journeys Home and it is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the authors and should not be attributed to DSS, the Department of Employment or the Melbourne Institute. The authors would also like to thank Alexandra de Gendre, Rafael Lalive, Dan O’Flaherty, David Ribar, Dominique Meurs and participants at the Microeconomics Workshop at the University of Lausanne, at the Economics Seminar of the University Paris Dauphine, at the Workshop on Poverty, Fundamental Needs and Public policy at the University of Nanterre, at the European Society for Population Economics (2015) and at the Irish Economic Association Annual Conference 2016 for useful comments on earlier drafts.

References

- Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation, Chapter 72, *Handbook of Econometrics*. Elsevier, pp. 5145-5303.
- Abbring, JH. and van den Berg, G. (2003), 'The nonparametric identification of treatment effects in duration models', *Econometrica* 71, 1491–1517.
- Abelson, J., Treloar, C., Crawford, J., Kippax, S., van Beek, I. and Howard, J. (2006). 'Some characteristics of early-onset injection drug users prior to and at the time of their first injection.' *Addiction* 101, 548-555.
- Bassuk, EL., Buckner, JC., Weinreb, LF., Browne, A., Bassuk, SS., Dawson, R. and Perloff, JN. (1997). Homelessness in female-headed families: childhood and adult risk and protective factors. *American Journal of Public Health* 87, 241-248.
- Bray JW., Zarkin GA., Ringwalt C. and Qi J. (2000). The relationship between marijuana initiation and dropping out of high school. *Health Economics* 9: 9-18.
- Brown S., Harris MN., Srivastava P. and Zhang X. (2018). Modelling illegal drug participation. *Journal of the Royal Statistical Society A* 181: 133-154.
- Chatterji P. (2006). Illicit drug use and educational attainment. *Health Economics* 15: 489-511.
- Cobb-Clark D., Kassenboehmer, S., Le T., McVicar, D. and Zhang, R. (2015). "'High"-school: the relationship between early marijuana use and educational outcomes.' *Economic Record* 91: 247-266.
- Curtis, M. A., Corman, H., Noonan, K. and Reichman, N. E. (2013). Life shocks and homelessness. *Demography*, 50, 2227–2253.
- Department of Housing and Urban Development (2014a). *The 2014 Annual Homeless Assessment Report (AHAR) to Congress Part 1: Point-in-Time Estimates of Homelessness*. The U.S. Department of Housing and Urban Development.
- Department of Housing and Urban Development (2014b). *The 2013 Annual Homeless Assessment Report (AHAR) to Congress Part 2: Estimates of Homelessness in the United States*. The U.S. Department of Housing and Urban Development.
- DeSimone J. (2002). Illegal drug use and employment. *Journal of Labor Economics* 20: 952-977.
- Duarte R., Escario JJ. and Molina JA. (2006). Marijuana consumption and school failure among Spanish students. *Economics of Education Review* 25: 472–481.
- Freund, J. (1961). A bivariate extension of the exponential distribution. *Journal of the American Statistical Association* 56, 971–977.
- Greene, JM., Ennett, ST. and Ringwalt, CL. (1997). Substance use among runaway and homeless youth in three national samples. *American Journal of Public Health*, 87, 229-235.
- Greenland, S. and Finkle, WD. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analysis. *American Journal of Epidemiology* 142: 1255-1264.

- Hall, W. (2015). What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? *Addiction* doi:10.1111/add.12703
- Johnson, G. and Chamberlain, C. (2008). Homelessness and substance abuse: which comes first? *Australian Social Work*, 61(4), 342-356.
- Johnson, TP. and Fendrich, M. (2007). Homelessness and drug use: evidence from a community sample. *American Journal of Preventive Medicine* 32, S211–S218.
- Kaestner R. (1991). The effect of illicit drug use on the wages of young adults. *Journal of Labor Economics* 9: 381-412.
- Kaestner R. (1994). The effect of illicit drug use on the labor supply of young adults. *Journal of Human Resources* 29: 126-155.
- Kidd, SA. (2007). Youth homelessness and social stigma. *Journal of Youth and Adolescence* 36, 291–299.
- Lalive, R., van Ours JC., Zweimüller, J. (2008) The impact of active labour market programmes on the duration of unemployment in Switzerland, *Economic Journal* 118: 235-257.
- Lehmann, ER., Kass, PH., Drake, CM and Nichols, SB. (2007). Risk factors for first-time homelessness in low-income women. *American Journal of Orthopsychiatry* 77(1), 20–28.
- Link, B. G., Susser, E., Stueve, A., Phelan, J., Moore, R. E., and Struening, E. (1994). Lifetime and five-year prevalence of homelessness in the United States. *American Journal of Public Health*, 84, 1907–1912.
- Mallett, S., Rosenthal, D. and Keys, D. (2005). Young people, drug use and family conflict: pathways into homelessness. *Journal of Adolescence* 28, 185–199.
- Martijn, C. and Sharpe, L. (2006). Pathways to youth homelessness. *Social Science & Medicine* 62: 1–12.
- McCaffrey DF., Pacula RL., Han B. and Ellickson P. (2010). Marijuana use and high school dropout: the influence of unobservables. *Health Economics* 19:1281-1299.
- McDonald Z. and Pudney S. (2000). Illicit drug use, unemployment, and occupational attainment. *Journal of Health Economics* 19: 1089–1115.
- McVicar, D. (2014). The impact of monitoring and sanctioning on unemployment exit and job-finding rates. *IZA World of Labor*.
- McVicar, D., Moschion, J. and van Ours, JC. (2015). From Substance Use to Homelessness or Vice Versa? *Social Science & Medicine* 136, 89-98.
- Melbourne Institute (2013). *Journeys Home Wave 3 Technical Report – Fieldwork, Response and Weighting*. Report prepared for the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs.
- Register CA., Williams DR. and Grimes PW. (2001). Adolescent drug use and educational attainment. *Education Economics* 9: 1-18.
- Roebuck, MC., French, MT and Dennis, ML. (2004). Adolescent marijuana use and school attendance. *Economics of Education Review* 23: 133–141.

- Scutella, R., Johnson, G., Moschion, J., Tseng, Y. and Wooden M. (2012). *Journeys Home Research Report 1*. Report prepared for the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs.
- Shelton, KH., Taylor, PJ., Bonner, A. and van den Bree, M. (2009). Risk factors for homelessness: evidence from a population-based study. *Psychiatric Services* 60(4), 465-472.
- Smith, J., Bushnaq, H., Campbell, A., Hassan, L., Pal, S. and Akpadio, S. (2008). *Valuable lives: capabilities and resilience amongst single homeless people*. London Metropolitan University.
- Susser, E., Moore, R. and Link, B. (1993). Risk factors for homelessness. *American Journal of Epidemiology* 15(2), 546-556.
- Van den Berg, G. J. (2001). Duration models: specification, identification and multiple durations. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter 55, 3381–3460. Amsterdam: Elsevier.
- Van den Bree, M., Shelton, K., Bonner, A., Moss, S., Thomas, H. and Taylor, PJ. (2009). A longitudinal population-based study of factors in adolescence predicting homelessness in young adulthood. *Journal of Adolescent Health* 45, 571–578.
- Van Ours, JC. (2006a). Dynamics in the use of drugs. *Health Economics* 15, 1283–1294.
- Van Ours, JC. (2006b). Cannabis, cocaine and jobs. *Journal of Applied Econometrics* 21, 897-917.
- Van Ours, JC. and Williams, J. (2009). Why parents worry: Initiation into cannabis use by youth and their educational attainment. *Journal of Health Economics* 28, 132-142.
- Van Ours, JC. and Williams, J. (2011). Cannabis use and mental health problems. *Journal of Applied Econometrics* 26, 1137-1156.
- Van Ours, JC. and J. Williams (2012). The effects of cannabis use on physical and mental health. *Journal of Health Economics* 31, 564-577.
- Van Ours, JC., Williams, J., Fergusson, D. and Horwood, L. (2013). Cannabis use and suicidal ideation. *Journal of Health Economics* 32 (3), 524-537.
- Van Ours, JC. and Williams, J. (2015). Cannabis use and its effects on health, education and labor market success. *Journal of Economic Surveys* DOI: 10.1111/joes.12070.
- Wooden, M., Bevitt, A., Chigavazira, A., Greer, N., Johnson, G., Killackey, E., Moschion, J. (2012). Introducing ‘Journeys Home’. *Australian Economic Review* 45(3), 368-378.
- Zarkin GA., Mroz TA., Bray JW. & French MT (1998). The relationship between drug use and labor supply for young men. *Labour Economics* 5: 385–409.

Appendix

1. Data details

Definition of explanatory variables

1. Parents separated: Respondent not living with parents at age 14 because they were divorced/separated
2. Parents dead: Respondent not living with parents at age 14 because they were dead
3. Conflict with parents: Respondent not living with parents at age 14 because of conflict with parents
4. Emotional abuse: Emotional abuse/neglect as a child
5. Physical violence: Physical violence as a child
6. Sexual violence: Sexual violence as a child
7. Male caregiver substance abuse: Male caregiver had an alcohol or drug problem
8. Male caregiver jail: Male caregiver spent time in jail
9. Male caregiver hospital: Male caregiver spent time in hospital overnight because had mental health problems
10. Male caregiver unemployed: Male caregiver was unemployed more than 6 months
11. Male caregiver gambling: Male caregiver had a gambling problem
12. Female caregiver substance abuse: Female caregiver had an alcohol or drug problem
13. Female caregiver jail: Female caregiver spent time in jail
14. Female caregiver hospital: Female caregiver spent time in hospital overnight because had mental health problems
15. Female caregiver unemployed: Female caregiver was unemployed more than 6 months
16. Female caregiver gambling: Female caregiver had a gambling problem
17. Missing info reason: Missing information on reason not living with parents at age 14
18. Missing info violence: Missing information on violence and abuse during childhood
19. Missing info father: Missing information on the male caregiver's variables
20. Missing info mother: Missing information on the female caregiver's variables

Table A1: Sample means of control variables, balanced panel, by gender

	Women	Men
Parents separated	36.5	31.6
Parents dead	6.7	6.4
Conflict with parents	8.3	6.1
Emotional abuse	57.6	58.8
Physical violence	57.3	63.0
Sexual violence	38.7	16.0
Male caregiver substance abuse	28.5	30.4
Male caregiver jail	10.3	10.6
Male caregiver hospital	5.5	4.8
Male caregiver unemployed	16.4	16.0
Male caregiver gambling	8.8	8.5
Female caregiver substance abuse	18.5	15.8
Female caregiver jail	2.7	2.0
Female caregiver hospital	13.2	10.0
Female caregiver unemployed	41.3	36.2
Female caregiver gambling	9.4	5.1
Missing info reason	1.6	1.7
Missing info violence	12.8	10.3
Missing info male caregiver	11.0	10.3
Missing info female caregiver	11.9	11.9
Total	639	708

2. A graphical summary of the bivariate simultaneous duration model

The two durations $H(c)$ and $C(h)$ can be thought of as potential outcomes. $H(c)$ is the potential outcome duration for the onset of homelessness and $C(h)$ is the potential outcome duration for the onset of daily cannabis use. ‘Treatment’ in this set up corresponds to the realization of one duration (i.e. initiation into daily cannabis use or onset of homelessness), which can causally affect the outcome duration of the other through its transition rate (the treatment effect). Once an individual is treated at a particular age, the transition rate for experiencing the potential outcome changes by the estimated treatment effect δ . Note that in our analysis duration is equivalent to age minus 10, i.e. we rule out transitions occurring before age 10. This can be illustrated as follows:

Figure A1: Treatment effect on the hazard rate for the outcome duration

Note: the figure represents a positive treatment effect occurring at age 20.

Figure A2 shows the cumulative starting probabilities for the potential outcome duration depending on whether and when the treatment starts, given observed characteristics and the distribution of unobserved heterogeneity (x, v_c, v_h) . The solid line gives the estimated cumulative

probability for an untreated individual in our sample to have experienced the outcome by the age of 10, 11, ..., 30. The dashed line “treatment onset at 25 years old” gives the probability to have experienced the potential outcome by the age of 10, 11, ..., 30 if instead the individual was treated at 25. Under the no-anticipation assumption these two lines coincide until the treatment at 25. At each subsequent age, the difference between the two potential outcomes represents the estimated effect of the treatment (transformed into a cumulative probability) since the time of the treatment (here 25). The difference between the solid line and the dashed line therefore depends on δ , individual characteristics (x , v_c , v_h) and age dependance (Table 6 provides some simulations of the cumulative probability of homelessness for men). Potential outcome trajectories can be drawn assuming that the treatment starts at any age between 10 and 30 (four examples are represented below).

Figure A2: Treatment effect on the cumulative probability of the potential outcome

3. Illicit/street drug use

Figure A3: Venn diagram of substance use by age 30, men (left) and women (right)

In our sample, like for cannabis daily, prevalence rates are higher than in the general population but smaller than for cannabis daily: 15 percent of females and 24 percent of males report having used other illicit street drugs at least weekly before the age of 30. Figure A4 presents hazard rates for first using illicit/street drugs weekly and shows that the take-up happens slightly later than for cannabis daily with a peak at 18. On average take-up happens at 19 for women and 20 for men. And for both genders there is a greater tendency for homelessness to precede drug use.

Figure A4: Transition rates into illicit/street drug use by age 30

4. Alternative treatments of missing data

We conduct four sensitivity analyses to the treatment of missing data in control variables underlying the estimates in the main text. Because our model is already complex, each is a simple strategy either dropping observations with missing information, dropping variables with missing information, or some combination of the two. These approaches may impart different biases to the estimates due to dropping subsamples for whom the effect may be heterogeneous or dropping variables that are correlated with homelessness and substance use onsets. Nevertheless, robustness of the key conclusions to these variations in treatment of missing information would suggest that missing information does not play a substantial confounding role in this case. The key estimates from these four alternative treatments of missing data, along with the baseline estimates from Table 4, are presented in Table A2.

Table A2: Sensitivity to alternative treatment of missing data

	Men			Women		
	Cannabis daily on Homelessness	Homelessness on Cannabis Daily	N	Cannabis daily on Homelessness	Homelessness on Cannabis Daily	N
Baseline	0.53 (3.7)***	-0.29 (1.7)*	708	-0.13 (0.7)	0.44 (2.2)**	639
Test 1	0.69 (4.4)***	-0.37 (1.8)*	533	-0.19 (0.9)	0.16 (0.7)	459
Test 2	0.77 (6.2)***	0.21 (1.3)	708	0.41 (2.3)**	1.05 (5.2)***	639
Test 3	0.57 (3.8)***	-0.23 (1.3)	630	-0.00 (0.0)	0.38 (1.7)*	551
Test 4	0.67 (4.6)***	-0.58 (2.8)***	581	0.39 (1.7)*	0.92 (4.1)***	523

Removing all observations for which any information is missing (Test 1) reduces the sample size by over 25% but hardly affects the parameter estimates for men. For women, however, the effect of homelessness on cannabis daily is now insignificantly different from zero. Note that this is the only specification for which this effect is insignificant and there are reasons to believe that this specification may not be superior to the baseline. Indeed, this alternative specification could be biased if the positive effect from homelessness to cannabis daily comes from the women in the sample who have missing information, for example if homelessness leads to cannabis daily for women with possibly harder childhoods. Removing all variables with missing information from the model specification retains the full sample (Test 2). Unsurprisingly, because homelessness and cannabis daily are positively correlated with each other and with the control variables in the model, all estimates in this case are significantly more positive. Removing observations with missing information on reason for separation from parents and violence/abuse experienced as a child, and removing variables on caregivers (Test 3), loses fewer observations than Test 1 and retains more control variables than Test 2. In this case the parameter estimates for neither men nor women are much affected. Finally, removing

observations with missing information on caregivers and removing variables with missing information on reason for separation from parents and violence/abuse (Test 4) again has little impact on the parameter estimates for men but the effect of cannabis daily on homelessness becomes marginally statistically significant for women.