CRIME AND JUSTICE
Bulletin

NSW Bureau of Crime
Statistics and Research

Contemporary Issues in Crime and Justice *Number 193*

This bulletin has been independently peer reviewed.

July 2016

Does the Custody-based Intensive Treatment (CUBIT) program for sex offenders reduce re-offending?

Imogen Halstead

***Aim**: To investigate whether completing the Custody-based Intensive Treatment (CUBIT) program for moderate to high risk/needs sex offenders reduces re-offending.*

***Method**: The study sample includes 386 male offenders identified by custodial staff as suitable for participation in CUBIT who were released from NSW custody during the period 2000 to 2010. An Instrumental Variables (IV) approach is adopted to control for potential unobservable selection bias in treatment outcomes (a key concern in the treatment effects literature), with the CUBIT completion rate amongst commencements around the time of an offender's own potential participation in CUBIT employed as the preferred IV. This method is statistically inefficient, and has limited power to assess the impact of treatment on sex re-offending (which is relatively rare) specifically. As such, three separate two-stage least squares models (including a range of control variables) are used to estimate the impact of completing CUBIT on whether or not an offender will re-offend with a sex, violent, or general offence, respectively, within five years of free time following their release from custody.*

***Results**: Within five years free time following release from custody 12 per cent of offenders who completed CUBIT committed a proven sex offence, 27 per cent re-offended with a new violent offence and 41 per cent committed a proven offence of any type. Multivariate models controlling for a range of important observable risk factors found that CUBIT completers had, on average, a 5-year general recidivism risk that was 13 percentage points lower than a similar cohort of offenders suitable for CUBIT but who did not participate. No significant differences between the treated and untreated groups were found for violent or sex re-offending.*

***Conclusion**: There is some evidence to suggest that completing CUBIT results in a considerable reduction in general recidivism risk. No evidence is found to suggest that CUBIT completion reduces sexual or violent re-offending. However, it is difficult to draw any conclusion from the null results, since the power of the statistical methods employed to detect a treatment impact in this study is limited given the relatively small sample size.*

***Keywords**: sex offender treatment, re-offending, instrumental variables, impact evaluation*

INTRODUCTION

Corrective Services NSW provides a range of custody-based treatment programs for sex offenders of varying risk, each with an overarching aim to "reduce the likelihood that treated individuals will continue with sexual offending behaviour upon their return to the community" (Corrective Services NSW, 2015b, p. 1). Naturally, evidence around the extent to which these programs reduce recidivism is in high demand, but evaluating the causal impact of sex offender programs is notoriously challenging. In particular, there is generally no natural control group for program graduates and so it is difficult to estimate their counterfactual outcome. Offenders deemed to be suitable for sex offender treatment face strong institutional incentives to enlist and progress through the program (for example, in order to attain parole), but ultimately participation is voluntary. There is every possibility that observable characteristics—such as age, criminal history, and so on—are insufficient to adequately control for differences in the innate recidivism risk of sex offenders who end up completing treatment prior to release from custody, and sex offenders who are ultimately released untreated. Most standard methods which could conceivably be used to estimate the impact of treatment on recidivism risk would yield biased results in this context, reflecting the presence of unobservable selection effects, rather than isolating the impact of treatment itself.

Suggested citation: Halstead, I. (2016). *Does the Custody-based Intensive Treatment (CUBIT) program for sex offenders reduce re-offending?* (Crime and Justice Bulletin No. 193). Sydney: NSW Bureau of Crime Statistics and Research.

This study aims to evaluate the impact of the NSW Custody-based Intensive Treatment (CUBIT) program for sex offenders on re-offending using Instrumental Variables (IV) methods, which are arguably more robust than standard regression or matching methods in the presence of unobservable selection effects. IV methods work by exploiting some factor that affects the likelihood with which an inmate will complete treatment, but which is otherwise unrelated to the inmate's re-offending risk (after conditioning on available observable characteristics). In this paper, I argue that the CUBIT completion rate around the time of an offender's own potential participation in CUBIT can be exploited to identify the causal impact of treatment—that it is a valid IV. I posit that this variable captures time-varying exogenous institutional factors which affect both the marginal opportunity, and marginal willingness, of offenders to participate in CUBIT and ultimately to complete treatment.

THE CUBIT PROGRAM[1]

While the overarching aim of NSW custody-based sex offender treatment programs is to reduce sexual recidivism, the suite of associated treatment goals is arguably more general. The Corrective Services NSW (2015b) manual on *Institutional Programs for Individuals who have committed sexual offences* states that these programs aim to help offenders develop the skills to:

- lead a more fulfilled and pro-social life in general
- experience less interpersonal conflict with others
- reduce the likelihood of future trouble with the law
- have an increased feeling of being more in control of their life and their future
- be able to manage problems more effectively
- have an increased sense of emotional closeness with the people around them
- have more enjoyable relationships
- have a greater sense of confidence in their own abilities to achieve the things that they want in their lives
- have a less-stressed approach to life
- experience a more positive and realistic view of oneself, and
- experience a more satisfying and pro-social sex life (p. 6).

CUBIT, which has been in operation since 1999, is the most intensive sex offender treatment program on offer in NSW prisons. Any male inmate (serving a sufficiently long custodial sentence) who has committed (and not persistently denied) a proven sex offence or an offence with a sexual element to it is potentially eligible for the program. However, it is targeted towards offenders who present a moderate to high risk of recidivism, and/or with moderate to high treatment needs. Alternative (less intensive) sex offender treatment programs are available for offenders with a lower risk of recidivism and/or more modest treatment needs.

CUBIT treatment is delivered over the course of 6-12 months, during which time participants live in a self-contained therapeutic community (see Ware, Frost, & Hoy, 2010). Formal therapeutic sessions are delivered in groups led by one or two psychologists, with up to ten inmates per group. Program staff at Corrective Services NSW reportedly exert considerable time and effort in attempting to ensure that all potentially suitable inmates are referred to treatment, and institutional incentives to participate include mechanisms associated with awarding parole and assigning security classification levels (Corrective Services NSW, 2015a; NSW State Parole Authority, 2012). However, referral and treatment are ultimately voluntary.

There have been a number of changes to CUBIT operations since the program's inception, which I posit are reflected in the IV proposed in this study—the variation in CUBIT completion rates over time. For example, CUBIT commenced in 1999 as a 20-bed program, grew to support 40-beds in 2001, and then was restructured in 2005 to expand capacity further (potentially by around 50 per cent) (Ware & Bright, 2008). In theory, when there is excess capacity relative to demand, commencement rates are likely to be fairly high (all else being equal). Program staff might also be better motivated and equipped with the resources required to support each individual's progress through to treatment completion. These factors would tend to lift the aggregate CUBIT completion rate amongst an inmate's 'peer' cohort, and increase the likelihood with which that inmate himself will undertake treatment. Therapist characteristics might also be an important institutional influence on CUBIT enrolments and attrition (Ware & Bright, 2008). Not only do therapist characteristics obviously vary across individuals, and individual therapists vary over time, but institutionally an effort was made to engender and apply positive therapist characteristics following CUBIT's restructure in 2005.

RELATED LITERATURE

There is one existing paper assessing the impact of NSW custody-based treatment programs for sex offenders on re-offending. Woodrow and Bright (2011) conduct a riskband analysis for 117 inmates who completed CUBIT and CUBIT-Outreach (a variation aimed at low-moderate risk/needs offenders), comparing realised recidivism outcomes to recidivism rates predicted by a pro-forma actuarial risk assessment tool—the Static-99 (Hanson & Thornton, 2000). The authors calculate a sexual recidivism rate amongst the treated sample of 8.5 per cent, which compares to 26 per cent predicted by the risk-assessment tool; the realised violent re-offending rate amongst the treated sample was 12.8 per cent, well-below the 36 per cent prediction. These comparisons are stark, and the authors characterise the findings as "consistent with the general consensus that well-implemented cognitive-behavioural treatment can have a positive effect on offending behaviour" (Woodrow & Bright, 2011, p. 1). However, perhaps due to data limitations, the authors do not report re-offending rates for any comparable group of untreated offenders in NSW.

Unfortunately, there are many reasons why the actuarial tool might not accurately reflect underlying recidivism risk amongst the NSW study sample; the Static-99 was initially benchmarked against four populations of sex offenders with completely different characteristics (some of whom underwent some form of treatment), in completely different contexts in Canada and the UK (Hanson & Thornton, 2000). It is difficult to confidently conclude that NSW custody-based sex offender treatment programs reduce recidivism from this study.

There is quite a large body of literature which finds similar results suggesting sex offender treatment programs reduce re-offending. This has been extensively summarised and reviewed elsewhere; for example, see meta-analyses by Hanson et al. (2002) and Lösel and Schmucker (2005), and for reviews focusing on the Australian setting, see Lievore (2004), Gelb (2007), and Macgregor (2008). Harkins and Beech (2007) document the debate prevalent in the literature around the benefits and limitations of alternative approaches to evaluating the impact of sex offender treatment. In general, the studies that employ more rigorous methods in an attempt to control for unobservable selection effects fail to identify any significant impact of treatment (though small sample sizes and low incidence rates limit the power of many of these analyses).

Hanson et al.'s (2002) seminal meta-analysis of 43 studies of psychological treatment for sex offenders concludes that both sexual and general recidivism rates are lower for offenders who participate in treatment than for comparison groups (on average, 12.3 versus 16.8 per cent for sexual re-offending, and 27.9 versus 39.2 per cent for general re-offending). However, this meta-analysis is dominated by studies likely to suffer biases due to selection effects, including studies which compare outcomes for participants and people who refused treatment, completers versus dropouts, and outcomes for participants who received varied dosage. Rice and Harris (2003) critique the quality of inference in this review, and cite particular concern about the extent to which control groups are adequately designed to mitigate selection bias. They conclude that "the effectiveness of psychological treatment for sex offenders remains to be demonstrated" (p. 428). In a separate meta-analyses of 80 sex offender treatment outcome evaluations, Schmucker and Lösel (2008) come to a related conclusion, that "methodological study characteristics explained the largest proportion of effect size variance" (p. 10).

Given the potential for unobservable selection effects in treatment participation, researchers Marques, Wiederanders, Day, Nelson, and Van Ommeren (2005) arguably provide some of the more robust evidence around the impact of sex offender treatment on recidivism. Their work is based on a randomised control trial (RCT) that was administered in the 1980s in California, in the United States. During the trial, 484 inmates who volunteered for treatment were matched on several characteristics (relating to their age, characteristics of their sex offending, prior criminal history) and parties to a match were randomly assigned to either a treatment group or control group. The researchers monitored subjects' sexual and violent re-offending for up to 14 years (with an average of 8 years) following their release from custody, and they found no statistically significant differences in re-offending rates across the treated and untreated groups. For comparative purposes, the researchers also constructed a matched group of non-volunteers from the wider population of inmates identified as suitable for treatment (N=220, from a pool of more than 16,000 sex offenders). Sexual and violent offending rates were also no different for this group, suggesting that any unobservable selection effects inherent in volunteering for treatment might be insignificant in this context.[2] Of course the sample size in this study is limited, so small treatment effects would have been difficult to detect, and there remains an open question around the extent to which results from this study are generalisable to sex offender treatment programs run elsewhere.

THE PRESENT STUDY

This paper aims to contribute to that branch of the literature that attempts to estimate the causal impact of sex offender treatment on re-offending in the presence of unobservable selection effects. In the context of CUBIT, the direction of potential selection bias is not immediately obvious. General scepticism around the dominant results in the existing literature (which imply treatment is effective) suggests that observers are primarily concerned with the possibility that sex offenders who are innately less risky, might be more likely to complete treatment. This is certainly possible in the case of voluntary participation, as applies to CUBIT. Consider, for example, an offender's motivation to change—a characteristic largely unobservable to the statistician. A motivated offender might be less likely to re-offend regardless of treatment, but also more likely to volunteer for treatment in the first place. At the same time, due to a host of institutional incentives for higher risk offenders to participate in CUBIT, the opposite bias also arise. For example, psychologists might prioritise treatment offers so that therapy is more likely to be available for offenders who they know (above and beyond the indicators from observable pro forma quantitative risk-assessment tools) to be of higher risk of re-offending. Offenders whose crimes were of such a nature that they would be unlikely to gain parole in the absence of treatment might also be particularly eager to attend. In this case, standard estimates of the treatment effect would be upwardly biased, potentially making treatment seem ineffective, even if in actual fact it reduces recidivism. The IV approach adopted in this paper aims to circumvent these potential issues.

The paper proceeds as follows. Immediately following this introduction, the next section provides detail on the study method: the data sources utilised, the study scope (the outcomes of interest, the study sample, and the treatment status variable), and detail on the IV approach, including the proposed IVs, the statistical methods applied, and the range of potential observable control variables. The third section presents the results: descriptive statistics relating to prima farce recidivism

rates, the sample characteristics, and the proposed IVs, and results from the models predicting re-offending with a sex, violent and general offence, respectively. The fourth section concludes with a discussion.

METHOD

DATA SOURCES

Data for this study were sourced from three separate databases:

1. Corrective Services NSW sex-offender treatment program (SOTP) database. This database is primarily used for recording administrative data pertaining to SOTP operations. It contains records dating back to 1998. The database records offenders' progress through the SOTP referral system: whether or not an inmate consents to referral; whether a referred inmate is found suitable or unsuitable for treatment; whether a suitable inmate is offered a place in a treatment program; whether that offer is accepted, and treatment commenced; and whether or not an individual completes a treatment program, or is discharged prior to completion. This database also records some individual characteristics, including data relating to the nature of crimes committed, and outcomes of official risk assessments.
2. The centralised Offender Integrated Management System (OIMS) database maintained by Corrective Services NSW. This database provides administrative data relating to inmate characteristics and custodial episodes (such as security classification levels).
3. BOCSAR's Re-Offending Database (ROD), which links all finalised NSW criminal court appearances and all movements in and out of NSW custody for a given individual from January 1994 to the present (Hua & Fitzgerald, 2006). ROD data used in the current study relate to all court appearances finalised up to 30 June 2015.

Offenders appearing in the SOTP database were linked to OIMS records by the OIMS offender identification number (the Master Index Number), with the relevant custodial episode identified from dates entered in the SOTP database (such as the date of referral). In turn, these data were linked to ROD records using each offender's name, date of birth and the OIMS database offender identification number; the relevant custodial episode was identified using the episode start date. There were 2,993 unique individuals with records in the SOTP database, eight of whom could not be matched to ROD. There were also nine custodial episodes pertaining to individuals who were matched to ROD, that could not themselves be matched to ROD custodial episode records. Fifty of the remaining 3,124 matched custodial episodes were not matched on the episode start date, but were instead matched on the episode end date. Closer inspection of the matched ROD data cast some doubt on a subset of the successful matches; these were flagged and ultimately excluded from the re-offending analyses, reducing the size of the final sample size by around 7 per cent. A small number of observations were also excluded due to other issues.[3] (Ultimately only a very small fraction of the matched SOTP database records were useful in this study, since it focuses on the CUBIT program only, and requires offenders are released from custody, with at least a five-year follow-up window; the study sample is detailed further below.)

STUDY SCOPE

Outcome of interest

Recidivism outcomes are measured by a binary variable recording whether or not an individual commits a proven offence within five years of 'free' time (that is, excluding time spent in prison) following their release from the index custodial episode. The five-year window is informed by existing literature on the actuarial prediction of sex offending, which demonstrates that a longer time-frame is helpful given the relatively low incidence and/or rate of sexual reconviction (Quinsey, Rice, & Harris, 1995).

Regardless, the low incidence of sex re-offending limits the power of the statistical analyses to detect any impact of treatment on sexual recidivism, particularly given the relatively small sample size (see section following). As such, this study examines three nested classes of re-offence:

1. A sex offence. Any offence falling within Australian and New Zealand Standard Offence Classification (ANZSOC; Australian Bureau of Statistics, 2011a) category 03 Sexual assault and related offences.
2. A violent offence (including sex offences). Any offence falling within ANZSOC categories: 011 Murder, 012 Attempted murder, 02 Acts intended to cause injury, 03 Sexual assault and related offences, 05 Abduction, harassment and other offences against the person, 06 Robbery, extortion and related offences, and 1211 Property damage by fire or explosion, and offences with law part codes 1207, 62079, 65020, capturing breaches of Apprehended Domestic Violence Orders (ADVOs). This definition attempts to replicate that employed in the benchmark recidivism study relating to CUBIT by Woodrow and Bright (2011); in that study, violent offences include sex offences and other violent offences "as defined by the Static-99 guidelines" (p. 48)—that is, according to Harris, Phenix, Hanson, and Thornton (2003, p. 23).
3. A general offence. Any new proven offence with the exception of breach offences unrelated to ADVOs (ANZSOC categories 151, 152 or 153 where law part codes are not 1207, 62079, or 65020).

Study sample

The study sample is drawn from the set of individuals with records in the SOTP database. The database includes administrative data spanning multiple custodial episodes for some individuals, and only the first custodial episode during which a record was entered in the SOTP database for any

particular individual is considered for inclusion in the sample. The sample also only includes those individuals who were specifically identified as suitable for CUBIT during this particular episode; it excludes individuals who were found unsuitable for treatment, who were found suitable for alternative treatment (such as the CUBIT-Outreach program, targeted at low-to-moderate risk offenders), and individuals whose suitability was not assessed. This reduces the sample size considerably. The final study sample also includes only those individuals for whom we can observe re-offending outcomes over a five-year follow-up window after their release from the index custodial episode—at a minimum, individuals released onto parole, or released due to their sentence expiring (not through being deported, for example) before 30 June 2010 (five years prior to the ROD cut-off date). This reduces the sample size further.

The final study sample comprises 386 individuals found suitable for CUBIT, of whom 229 (59.3 per cent) completed treatment prior to their release from custody. Of the remaining 157 offenders who did not complete treatment, 65 commenced but failed to complete CUBIT, while 92 failed to commence treatment prior to release from custody.

Smaller subsamples are used in the re-offending analyses, since individuals may return to prison without having recorded any new offence (for example through breaching parole conditions), or without having recorded a new offence within the category of interest in the model (for example, after committing robbery, rather than a sexual offence). In these cases, outcome data relating to an individual's propensity to re-offend are effectively censored (individuals' opportunity to re-offend in custody is limited). To address this issue in the re-offending models, only 'free time' spent outside of custody is counted when quantifying the follow-up window and defining the analytical samples. That is, an offender is included in the sample for sex offending if the time he has spent outside of custody following release from the index episode exceeds five years; or if the sum of the time he spent outside of custody prior to re-offending with a sexual offence, plus the window of time from that first sexual re-offence up to 30 June 2015 (the ROD cut-off date), exceeds 5 years. Analogous formulas are used for a re-offence with a violent offence, and a re-offence with a general offence.

The final analytical samples include:

1. 347 individuals for analysing re-offending with a sex offence (59.4 per cent of whom completed CUBIT);
2. 369 individuals for analysing re-offending with a violent offence (59.3 per cent of whom completed CUBIT); and
3. 379 individuals for analysing general re-offending (59.4 per cent of whom completed CUBIT).

Note that these final study samples are not quite perfectly representative of the general population of offenders found suitable for CUBIT. By definition, the samples exclude people who entered custody relatively recently and the later experiences of offenders with multiple custodial episodes (in instances where the SOTP database includes a record relating to a previous custodial episode for that offender), and it tends to exclude offenders with longer sentences (who have shorter follow-up periods, all else being equal).

In addition, the re-offending sub-samples are more likely to exclude offenders who, following their index custodial episode, returned to prison for reasons unrelated to a new re-offence of interest. For example, the re-offending samples are more likely to exclude individuals who returned to custody for a breach of parole conditions, which does not in and of itself constitute a new offence. In addition, the sample used to assess sex (and analogously violent) offending, is biased through excluding those individuals who are more likely to commit non-sex related offences (since these individuals are less likely to have sufficient follow-up free time over which to assess their sex re-offending). It is encouraging to note that this censoring appears unlikely to bias the results regarding the impact of treatment on re-offending. Similar shares of treated and untreated offenders are censored from the sex re-offending analyses (10.0 and 10.2 per cent, respectively), the violent re-offending analyses (4.4 and 4.5 per cent, respectively), and the general re-offending analyses (1.8 and 1.9 per cent, respectively).

TREATMENT STATUS

For the purposes of this study, individuals were flagged as having been 'treated' if they were recorded as having completed CUBIT prior to their release from custody. Untreated individuals constitute the remainder of the sample—individuals who consented to referral for sex offender treatment and were found suitable for CUBIT, but who failed to complete treatment prior to their release (i.e. who either never commenced treatment, or who commenced but did not complete the program).

In this particular institutional context, there was an intention to treat all offenders in the sample. For the purpose of identifying a treatment effect, defining treatment by program completion avoids some of the issues that commonly plague intention-to-treat designs; in particular, where the 'treatment' group is diluted by subjects who were effectively untreated (who were referred to the program, but did not commence; or who commenced but did not complete treatment). IV methods are specifically designed to be robust to the unobservable selection effects related to commencement and completion that usually motivate an intention-to-treat approach. If treatment gains are only realised by offenders who complete CUBIT, then defining 'treatment' by completion should lift the statistical power of our attempts to identify any causal impact of the program, all else being equal (provided the IV is sufficiently strongly correlated with CUBIT completion[4]).

Another advantage of defining treatment status by program completion is that it invites the exploration of two alternative IVs—one reflecting treatment completion rates amongst an offender's peer group as a whole, and another reflecting completion rates amongst peers who commenced treatment. The

latter has the potential to more readily satisfy both the exclusion restriction and the relevance condition required of an IV, and it is not possible to construct an IV analogous to this latter alternative when treatment status is defined by program commencement. The IVs, and these validity conditions, are discussed in detail below.

IV approach

The IV approach adopted in this paper allows for consistent estimation of the impact of treatment on re-offending in the presence of unobservable selection effects. The method works by isolating an exogenous factor—an IV—that is correlated with participation in treatment, but which is assumed to be otherwise unrelated to recidivism (conditional on other important observable characteristics). Statistical analysis is used to exploit the variation in treatment outcomes induced by the IV, and in essence, to compare re-offending outcomes for treated and untreated offenders whose exposure to treatment was sensitive to this exogenous factor.

Suppose the underlying structural model for re-offending is the following:

$$Y_i = \gamma T_i + X_i B + u_{it} \quad \ldots\ldots (1)$$

where Y_i is a re-offending outcome for individual i, T_i is a binary variable indicating whether or not an individual i completed treatment, X_i is a vector of the complete set of other relevant control variables, and u_i is random error. Here gamma is the parameter of interest—the impact of treatment on re-offending.

In practice we are likely to be able to observe only a subset $(\tilde{X}_i)$ of the complete set of relevant factors (X_i) related to recidivism. In this case, we are only able to estimate the model:

$$Y_i = \gamma' T_i + \tilde{X}_i B' + u'_{it} \quad \ldots\ldots (1')$$

Importantly, if there are unobservable selection effects—unobservable characteristics related to recidivism risk, which are also correlated with treatment completion—then treatment will be endogenous (that is, correlated with the error term) in the estimated model. The same issue arises in the presence of certain types of measurement error. In this context, standard ordinary least squares (OLS) regression procedures will yield biased estimates of the parameter (γ) of interest (see Wooldridge, 2010, section 4.3.1).

IV methods can help overcome this issue, as they allow for consistent estimation of the treatment impact, even where there are unobservable selection effects (see Wooldridge, 2010, Chapter 5 for an overview). To be valid, an IV must be identified which satisfies the following assumptions:

1. Relevance: The IV must be (sufficiently strongly) correlated with treatment.
2. The exclusion restriction: The IV must not be correlated with the outcome of interest (conditional on the other exogenous control variables), except through its effect on participation in treatment.

The first assumption is testable, but the second must be assumed.

Unfortunately, IV methods are costly in terms of statistical efficiency. Intuitively, the treatment effect is ultimately identified through variation in the IV, which—unlike random allocation—is imperfectly correlated with actual treatment outcomes. Statistically speaking, this makes it more difficult to identify a significant effect, particularly given the small study sample; it implies IV methods have limited power to assess the impact of treatment on sex re-offending in particular (because it is relatively rare). However, once a valid IV has been identified, it is possible to test the extent to which residual unobservable selection effects are likely to bias estimates of the treatment effect in a standard regression framework. Such tests seek statistical evidence of suspected endogeneity (under the null hypothesis, the suspected endogenous variable is in fact exogenous). If there is no obvious cause for concern, then the more efficient estimate derived from equation (1'), which supposes treatment is exogenous, may provide the most informative results.

Proposed IVs

Recall that the IVs proposed in this study aim to reflect the CUBIT completion rate around the time of an inmate's own prospective participation in treatment. Specifically, I consider CUBIT completion rates[5] amongst what I refer to as the inmate's *peer group*—all those individuals in the SOTP database, identified as suitable for CUBIT, who had earliest possible release dates within a 6-month window either side of an individual's own earliest possible release date (excluding the individual himself). The earliest date of release is used in place of the actual custodial episode end date in defining this peer cohort, because the latter may not be exogenous to CUBIT completion (individuals may be released early on parole precisely because they completed treatment).

The first IV I consider is constructed by calculating the fraction of the peer-group who completed CUBIT prior to release. That is:

$$IV_{1,i} = \frac{\sum_{j \in P_i} 1\{Completed_j = 1\}}{|P_i|}$$

Where $IV_{1,i}$ denotes the instrumental variable (the peer-group completion rate for individual i), P_i denotes the set of individual i's peers, and Completed_j is a binary variable indicating whether individual $_j$ went through treatment prior to release.

In order to be valid as an IV, this variable must satisfy the exclusion restriction: that is, it must be unrelated to an individual's recidivism risk, except through its exogenous impact on his propensity to undergo treatment. There is one obvious mechanism through which this restriction may not hold with this proposed IV. Specifically, the prevailing treatment rate may not be exogenous if the CUBIT program operates under binding capacity constraints. In this case, an individual who refuses to participate in treatment would *create* an opportunity for someone

else to participate (and there is a positive probability that that individual will complete the program). The prevailing treatment rate for individuals who refuse treatment would be (causally) higher than for other individuals, all else being equal. This would undermine the exclusion restriction if those individuals who refuse treatment innately present a higher (or equally, lower) risk of re-offending (that is, conditional on treatment).

The IV is only valid if the exclusion restriction holds—and this is not something we are able to test empirically since the model is not over-identified. I present the results using the overall peer-group completion rate as an IV for completeness. However, in light of concerns around the exclusion restriction, I also consider as the preferred IV, the prevailing completion rate *amongst those peers who commenced treatment*.

$$IV_{2,i} = \frac{\sum_{j \in P_i} 1\{Completed_j = 1\}}{\sum_{j \in P_i} 1\{Commenced_j = 1\}}$$

This alternative IV will not be affected by an individual's participation in treatment; both the numerator and denominator of the IV are altered when the opportunity for a new participant arises through another individual's failure to undertake or complete treatment. As a result, I consider this measure the preferred IV in terms of satisfying the exclusion restriction. Of course, it remains to be demonstrated that the preferred IV is sufficiently relevant as a predictor of individual offenders' treatment completion. Intuitively, the relevance of this second IV is more uncertain than the first. It is natural to imagine that the prevailing institutional factors underlying the completion rate amongst commencements will have the strongest impact on the probability an individual will complete treatment conditional on him having commenced in the first place. The relevance of this alternative IV is an empirical question to be tested.[6]

A separate risk to the exclusion restriction arises through the potential for peer effects to influence an individual's propensity to complete treatment and/or re-offend (see Bayer, Hjalmarsson, & Pozen, 2009; Damm & Gorinas, 2013). For example, suppose low-risk offenders exert a positive influence on their peers, generally reducing their recidivism risk, and that low-risk offenders are more likely to complete CUBIT (or vice versa). In this case, the IV for the influential individual's peers would be negatively related to their (peer-affected) re-offending risk (regardless of whether or not they completed treatment). Relatedly, if low-risk offenders who are more likely to complete CUBIT also *support others through treatment*, then the value of the IV for that influential low-risk offender would be positively correlated with their own completion and negatively related to their (innate) re-offending risk. These peer effects would tend to bias the IV estimate of the coefficient on treatment completion downwards, suggesting that CUBIT is more effective at reducing recidivism than is actually the case.

To test the robustness of the IV against this range of possibilities, a conditioning variable is constructed to provide a proxy measure of the relative risk level of an offender's peer group. Bivariate comparisons are used to identify a relevant covariate which is indicative of re-offending risk. Results from specifications including these conditioning variables are presented in the robustness checks section in the Appendix.[7]

Another risk to the exclusion restriction is that the time-variation intrinsic to the IV is correlated with some other time-varying factor which also relates to changes in re-offending rates over time. For example, higher treatment completion in custody may reflect a more effective justice sector in general, and hence be positively correlated with arrest and conviction rates (conditional on crime). This would tend to bias the IV estimate of the coefficient on treatment completion upward, implying CUBIT is less effective than it is in practice. Of course, this issue is not unique to this study; it might plague any analysis which does not control for aggregate trends in crime rates (since by assumption there will be more treated subjects at times when justice policy and practice is more effective). Nonetheless, to ensure robustness of the IV against this possibility, an additional conditioning variable is included in each model reflecting the aggregate NSW sex, violent, or general offending crime rate 2½ years following an offender's earliest release date (approximating the mid-point of the follow-up window used to observe recidivism outcomes).[8] Results from specifications including this conditioning variable are also presented in the robustness checks section in the Appendix.

Statistical analysis

In this paper, two-stage least squares (2SLS) IV methods are initially used to derive a consistent (but potentially inefficient) estimate of the causal impact of treatment on re-offending. In the first-stage, a reduced form equation is estimated predicting completion of treatment for individual i, denoted by the binary variable T_i, as a function of the (exogenous) observable control variables ($\tilde{X}_i$), and an instrumental variable, denoted by Z_i.

$$T_i = \tilde{X}_i B_1 + \beta_z z_i + v_i \quad \text{(First-stage)}$$

In the second-stage, the predicted probability of completion ($\hat{T}_i$) from the first-stage regression is used in a second-stage regression predicting re-offending.

$$Y_i = \tilde{X}_i B_2 + \gamma_{IV} \hat{T}_i + u_i \quad \text{(Second-stage)}$$

The coefficient on the treatment variable (γ_{IV}) is the IV estimate of the Local Average Treatment Effect (LATE) on re-offending. That is, the impact of treatment on recidivism for those offenders whose completion is sensitive to the IV (see Abadie, 2003; Angrist, Imbens, & Rubin, 1996; Imbens & Angrist, 1994). In the context of the present study, the treatment impact is identified for those offenders whose completion of CUBIT effectively hinges on variation in institutional conditions (as reflected in the peer completion rates) over time. To be clear, the LATE is uninformative with respect to the effect of treatment on those treated inmates who were not marginal—who would always have completed CUBIT (for example, due to their unwavering enthusiasm, or because their treatment was prioritised by the

institution). It also says nothing about how effective treatment would have been for people who would never complete treatment (because they are wholly uninterested, or institutionally excluded for whatever reason). The approach also *assumes* that there are no 'defiers'—people who are less likely to complete treatment at times when other peers are more likely to complete treatment, and vice-versa. Monotonicity of the instrument is also assumed; that is, the affect of the instrument on treatment is either always positive, or always negative (irrespective of its level).

Both the second-stage equation predicting re-offending, and the first-stage equation predicting CUBIT completion, will take the form of Linear Probability Models (LPM).[9] Heteroskedasticity-robust standard errors are estimated (but otherwise no adjustment is made to account for any impact of overlapping peer groups on these errors). In assessing the robustness of the results, I estimate a probit regression treating CUBIT completion as exogenous to compare with an analogous OLS LPM. I also estimate a seemingly unrelated bivariate probit (or biprobit) model to compare with the 2SLS IV results, but a full exploration of the bivariate probit approach is beyond the scope of this paper. The consistency of the treatment impact estimator in a bivariate probit model (which estimates the average treatment effect) rests critically on the relatively strong assumption that the model is correctly specified (for an overview see Baum, Dong, Lewbel, & Yang, 2012). A more comprehensive model predicting CUBIT completion (rather than the reduced form first-stage regression intrinsic to 2SLS) would be warranted in a more thorough exploration of this modelling approach.

Covariates are selected for inclusion in the models based on several considerations. Some variables (an offender's age and Aboriginal and Torres Strait Islander status) were initially included based on the weight of existing evidence suggesting their relationship with re-offending risk (for example, for literature on sex offender recidivism risk see Broadhurst & Loh, 2003; Gelb, 2007). Other variables (such as sentence length and ex-ante parole prospects) were initially included in response to evidence (regarding features of CUBIT operations) suggesting they might be correlated with program completion, possibly in addition to recidivism risk. Other potential covariates—such as parole supervision outcomes—were explicitly *excluded* due to a risk they might be affected by program completion (see discussion in the descriptive statistics section). Lastly, additional covariates were initially selected based on a benchmark OLS LPM predicting re-offending assuming CUBIT completion was exogenous; covariate groups which were statistically significant at the (relatively generous) .10 level were maintained in the initial specifications. An analogous probit model is also estimated, to check the robustness of the results on the independent variables against the choice of functional form. The preferred IV specification is ultimately refined to be more parsimonious, including just those covariates statistically significant at the .05 level in the second-stage IV regression.

All statistical analyses are performed using Stata version 13.1. The user-written command 'ivreg2' (Baum, Schaffer, & Stillman, 2010) is used to perform the 2SLS analysis. The 'robust' option is used to estimate heteroskedasticity-robust standard errors in the LPMs. (Alternative, bootstrapped standard errors are provided in simplified models presented in the robustness checks section.) An F-test is used to assess the relevance of the IV in the first-stage regression (a commonly applied rule-of-thumb is that the value of this statistic should exceed 10, though more refined critical values are available; James H Stock, Wright, & Yogo, 2002; James H. Stock & Yogo, 2005).

The 'endog' option is used to test for evidence that the IV approach is necessary to address the endogenity of treatment in the re-offending models. The test calculates a Chi-square statistic (with one degree of freedom—the number of instruments being tested) based on the difference of Sargan-Hansen (Hansen, 1982; Sargan, 1958) statistics for equations where completion is treated as exogenous and endogenous, respectively, and is robust to heteroskedastic errors. The null hypothesis assumes that treatment status can be treated as conditionally exogenous. If the null is not rejected, then there is little evidence to suggest there would be problematic bias in using OLS, in which case following an IV approach would simply impede efficiency. In this context, the Akaike Information Criterion (AIC; Akaike, 1974) value may be used to determine the preferred model amongst the benchmark alternatives (the OLS LPM and probit models) which each assume CUBIT completion is exogenous conditional on the covariates.

Control variables

The following range of control variables are considered for inclusion in the re-offending models:

1. Offender socio-demographic characteristics
    a. Age in years upon release from custody.
    b. Indigenous status: whether the offender identified as being of Aboriginal or Torres Strait Islander descent at any court appearance recorded in ROD.
    c. Postcode level of disadvantage according to the Australian Bureau of Statistics (2011b) Socio-Economic Indices for Areas (SEIFA) Index of Relative Socio-Economic Disadvantage (IRSD).
2. Characteristics relating to the offender's index custodial episode

    Many potential control variables relating to the index custodial episode could be affected by participation in CUBIT. For example, offenders who complete treatment might subsequently be awarded lower security classification levels, and be more likely to be released early onto parole. As a result, direct measures of these characteristics have been excluded from the suite of potential control variables. The following related control variables are considered for inclusion:

a. Minimum non-parole period. This variable captures the time from episode start date to the earliest possible release date, less any balance of parole to be served at the outset in the case of a breach at the time of imprisonment.
b. Sentence length. This variable captures the time from episode start date to the sentence expiry date, less any balance of parole to be served at the outset.
c. Parole prospects. This is a categorical variable indicating a period of parole ordered by the court, or the ex-ante prospect of parole being granted through the State Parole Authority (SPA). In the first instance, for offenders who were awarded parole, the variable was derived based on whether they realised parole under the directive of the court or SPA. Offenders whose non-parole period aligned with the sentence length were assigned no parole prospects. Otherwise, offenders with sentences shorter than three years duration were determined to have had court-ordered parole prospects, and offenders with sentences three years or longer were determined to have had SPA parole prospects (reflecting the legislation as set out in the *Crimes (Sentencing Procedure) Act 1999* and the *Crimes (Administration of Sentences) Act 1999*).
d. Maximum security classification level. This variable records the highest level of security classification assigned to an offender during the current custodial episode (generally, that which was assigned at the start of the episode; Corrective Services NSW, 2015a).

3. Offenders' criminal offence history and penalties received

Index offence:

a. Most Serious Offence type. This is a categorical variable indicating the Most Serious Offence associated with the index custodial episode: sex, other violent (non-sex) or other type of offence. These data were extracted from the OIMS database by staff at Corrective Services NSW, and the Most Serious Offence was determined based on the penalty attributed to each offence. The ANZSOC offence types were grouped to align with the definitions of sex, violent and general offences used in the re-offending models (see the data sources section).

Control variables for offenders' prior criminal offence history are based on the count of finalised court appearances (including youth justice conferences) during the index custodial episode or within 5 years prior to the index custodial start date where one or more of a particular type of offence was proven, or a particular type of penalty received. BOCSAR court records date back to 1994, so data are recorded as missing for any individual who entered custody prior to 1995. Moreover, counts will be downwardly biased for individuals entering custody in the years immediately after 1994. No adjustment is made for this.

Three different potential control variables are used to capture counts of prior finalised court appearances where there was a proven:

b. sex offence
c. violent (including sex) offence, or
d. any offence.

Five additional potential control variables capture counts of prior finalised court appearances where the following penalties are handed down:

a. imprisonment (full-time prison sentence, including juvenile control orders)
b. other detention (periodic detention, Intensive Correction Order, or a home detention sentence)
c. suspended sentence, or
d. 'section 9' bond (section 9 of the *Crimes (Sentencing Procedure) Act 1999 (NSW)* stipulates that a court may "instead of imposing a sentence of imprisonment... make an order directing the offender to enter into a good behaviour bond").

4. Official risk-assessment measures

a. The Level of Service Inventory-Revised (LSI-R) is an official actuarial-based assessment tool administered in order to estimate an individual's risk of general recidivism (Andrews & Bonta, 1995; Watkins, 2011). It provides an aggregate risk score based on 54 items (where a higher score indicates a higher level of risk), through combining results over ten different domains: criminal history (10 items), education/employment (10 items), financial (2 items), family/marital (4 items), accommodation (3 items), leisure/recreation (2 items), companions (5 items), alcohol/drug problem (9 items), emotional/personal (5 items), attitudes/orientation (4 items).
The LSI-R includes dynamic risk factors that may change over time; in particular, factors which might change as a result of completing treatment. This study attempts to preclude the use of LSI-R scores which may have been collected following CUBIT completion. For each individual, the most recent LSI-R score collected prior to one year before their release from the index custodial episode was extracted from ROD (ROD includes the history of LSI-Rs administered from 2007 onwards). If ROD included no such score, the LSI-R score associated with the OIMS extract was used if available, provided it was administered prior to one year prior to the offender's release.
LSI-R raw scores, and scores categorised into riskbands, are each considered for inclusion as control variables. The 'Low' riskband category is defined by a score of 0-13, 'Medium-Low' risk by a score of 14-23; 'Medium' risk by a score of 24-33; 'Medium-High' risk by a score of 34-40 and 'High' risk by a score of 41 or more.

b. The Static-99 (Hanson & Thornton, 2000) is an actuarial-based assessment tool commonly administered to estimate the risk of sexual reconviction in adult males. It is based on a range of static factors (unaffected by

therapeutic intervention): an offender's age, historical living arrangements, extent of past sex and violent offending, and characteristics relating to the nature of past sex offending (for details see Harris et al., 2003). The original Static-99 assessment tool was revised slightly in 2009 to attribute greater variation in predicted risk associated with an offender's age (Helmus, Babchishin, Hanson, & Thornton, 2009), with the revised version referred to as the Static-99R. The revised scheme, with scores ranging from a low of -3 to a high of 10, is used to interpret all Static-99 results here. Static-99R scores, and scores categorised into riskbands, are each considered for inclusion in the analysis. Riskbands are defined as follows: 'Low' risk applies to scores from -3 to 1; 'Low-Moderate' risk to scores of 2 or 3; 'Moderate-High' risk to scores of 4 or 5; and 'High' risk to scores of 6 or more (Helmus et al., 2009).

5. Recorded characteristics of offenders' sex offending

   The following details derived from SOTP database records are also considered for inclusion as potential control variables:

   a. the gender profile of an offender's victims: female, male or both,
   b. the age profile of an offender's victims: whether victims included adults, children or both, and
   c. the relationship the offender held with victims: whether they were immediate family, extended family, acquaintances, or strangers.

RESULTS

DESCRIPTIVE STATISTICS

This section presents descriptive reoffending statistics for the complete study sample (n=386). Further descriptive statistics for the study sample are presented in the Appendix to this report. The Appendix to this report includes descriptive statistics on sample offenders' socio-demographic characteristics, characteristics of the index custodial episode, offenders' prior criminal offence history and penalties received, official risk assessment results, and recorded characteristics of offenders' sex offending; compared across subsamples of treated (N=229) and untreated (N=157) offenders. This section presents headline reoffending statistics for the study sample by treatment status, and describes the proposed IVs. Further descriptive statistics relating re-offending rates to observable characteristics are included in the Appendix.

Recidivism

Table 1 summarises key recidivism rates for the study sample, including results of a bivariate statistical comparison of re-offending rates across treated and untreated cohorts.

- Of the 347 offenders available to assess sex re-offending, 11.8 per cent committed a proven sex offence within five years of free time following their release from custody. There was virtually no difference in this rate across the treated and untreated groups (the relevant Chi-square statistic is close to zero).
- Of the 369 offenders available to assess violent re-offending, 29.5 per cent re-offended with a proven violent offence within five years of free time following their release from custody. The violent re-offending rate was lower for people who completed CUBIT (26.5 per cent versus 34.0 per cent), but this difference was not statistically significant at .05 level (Chi-square = 2.4; *p*-value = .120).
- Of the 379 offenders available to assess general re-offending, 47.2 per cent re-offended within the five-year window of free time following their release from custody. General re-offending rates were considerably higher for the group who did not complete CUBIT than for the treated group; 56.5 per cent versus 40.9 per cent, respectively, with the difference statistically significant at the .05 level (Chi-square = 8.9; *p*-value = .003).

These data obviously in no way reflect the seriousness of the re-offence beyond the distinction provided by the three headline categories. For context, the following descriptive data on the ANZSOC category of re-offenders' first re-offence (within each category) provide some indication of the range of seriousness.

Table 1. CUBIT-suitable sample: Treatment and re-offending rates

		Re-offending within 5 years free time following release by offence type					
		Sex		Violent		General	
Treatment status	N	subsample N	% offended	subsample N	% offended	subsample N	% offended
Treated	229	206	11.7	219	26.5	225	40.9
Untreated	157	141	12.1	150	34.0	154	56.5
Total	386	347	11.8	369	29.5	379	47.2
Chi-square statistic			0.0		2.4		8.9**
p-value			.908		.120		.003

Note. Chi-square statistic is based on a test of independence of re-offending rates by treatment status.
* $p<.05$, ** $p<.01$, *** $p<.001$.

- Amongst the subset of offenders who re-offended with a sex offence within five years of free time following their release from custody, 55 per cent first re-offended with an aggravated sexual assault, 28 per cent with a child pornography offence, 9 per cent with a non-aggravated sexual assault, and the remaining 9 per cent with a non-assaultive sexual offence.
- Amongst the subset of offenders who re-offended with a violent offence, 30 per cent first re-offended with a sex offence, 20 per cent with a serious assault resulting in injury, 17 per cent with a common assault, 13 per cent with a stalking offence, and 8 per cent by breaching a violence order (the remaining 12 per cent included serious assault not resulting in injury, threatening behaviour, aggravated robbery, abduction and kidnapping, harassment and private nuisance and property damage by fire or explosion).
- Amongst the subset of offenders who re-offended with a general offence, 16 per cent first re-offended with a sex offence and 24 per cent with a violent (non-sex) offence. Another 24 per cent first re-offended with traffic and vehicle regulatory offences (such as drink or drug driving, or driving without a licence); 10 per cent by offending against justice procedures, government security and government operations; 7 per cent with an illicit drug offence; and 6 per cent with theft and related offences. The remaining 13 per cent included break and enter, deception, property damage, public order, prohibited weapons, and other miscellaneous offences.

Proposed IVs

Figure 1 illustrates the data underlying the IVs employed in this paper. The top panel (Panel A) illustrates the size of each individual's peer group (the number of individuals found suitable for CUBIT, with earliest release dates within a 12-month window centred around the index offender's own earliest release date), as a function of the earliest release date for that individual. Also shown are the number of those suitable peers who commenced the program, and those who ultimately completed treatment. These data are translated into overall commencement rates and completion rates (the first proposed IV) in the second panel (Panel B). The third panel (Panel C) depicts the percentage of peers who commenced the program who ultimately completed treatment (the preferred IV).

The size of the pool of CUBIT-suitable inmates ebbs and flows over time, but it has generally increased since the early days of operation, consistent with growth in program capacity (Figure 1, Panel A). Recall that CUBIT supported 20 beds from 1999 to 2001, then 40 beds, and then from September 2005, potential capacity equivalent to around 60-beds (through shifting to a rolling open-group format whereby new inmates can enter the program whenever a place becomes available) (Ware & Bright, 2008).

CUBIT completion rates were at the lowest levels for cohorts with earliest release dates around late 2004; during this era, only around one-in-three of the CUBIT-suitable sample completed the treatment program prior to their release from custody (Figure 1, Panel B). Both commencement and completion rates have risen since the CUBIT restructure in September 2005, and the wedge between them has narrowed. In the later years of the study period, almost all CUBIT-suitable offenders who commenced CUBIT went on to complete the program (Figure 1, Panel C). Corrective Services NSW is currently engaged in a comprehensive research program investigating these aspects of CUBIT operations, so further insight around these data is likely to become available in the near future (Howard, Manuscript in preparation).

In general, commencement and completion rates appear to move in sync with the volume of peers found suitable for CUBIT. This positive correlation might be expected if a greater number of suitable offenders are recruited through the referral system during periods of expansion, or at times when program staff are particularly proactive and effective (and vice-versa). As discussed earlier, it is possible that the characteristics of the sample of persons found suitable for CUBIT might change over time as the catchment net widens and contracts (for example, if program staff attempting to recruit additional participants, target less high-risk offenders), and that these changes might be correlated with treatment rates. This provides additional motivation to ensure the IV results are robust to the inclusion of control variables reflecting the peer-group risk level. To this end, the models were also re-estimated including the size of the peer-group as a conditioning variable as a robustness check, but it was nowhere near statistically significant in any of the specifications, and the results of this exercise are omitted.

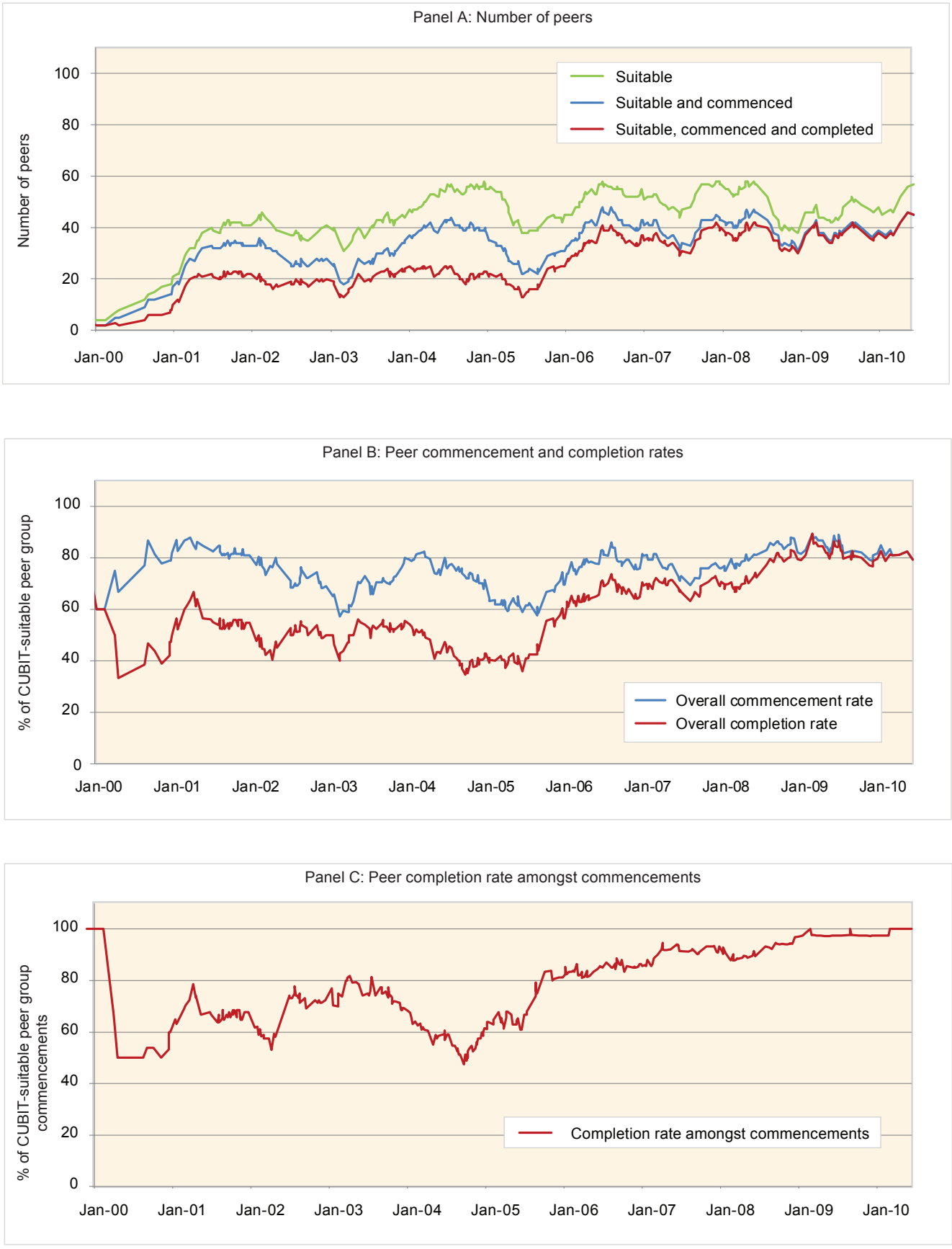
Occasionally, the commencement and completion rates appear to be negatively correlated with the volume of suitable peers (during the 2004 period in Figure 1, for example), perhaps suggestive of periods of particularly binding capacity constraints. This discourages the use of the first proposed IV (the completion rate amongst the whole peer cohort) and lends weight to the preferred alternative (the completion rate amongst peer-group commencements).

MODEL ESTIMATION

Tables 2, 3 and 4 report the model estimation results for re-offending with a sex offence, violent offence, and general offence respectively.

The results in Column D are key; they present coefficient estimates from the 2SLS model using the preferred IV—the prevailing completion rate amongst commencements. Recall that the IV estimate of the treatment effect is the estimated coefficient on treatment status (CUBIT completion) in the second-stage regression predicting re-offending; these results are presented in the top half of the table. The bottom half of the table presents estimates from the first-stage regression predicting CUBIT completion using the IV (it is these predicted values which are used to indicate treatment status in the second-stage regression). The key diagnostic to consider from this first-stage regression is the F-statistic on the excluded instrument, which

Figure 1. Prevailing treatment rates (proposed IV)



Note. Peer group sub-sample captures offenders found suitable for CUBIT with earliest possible release dates 6 months either side of an particular offender

indicates the extent to which the IV is a sufficiently relevant predictor of treatment (and therefore valid for use in the second-stage regression).

The other columns present results from supplementary models. Benchmark regressions results treating CUBIT treatment as exogenous are reported in each table in Columns A and B. The former estimates a probit specification; the latter estimates an LPM using OLS. Note that since the IV models employ an LPM functional form, their results should be benchmarked against Column B. Column C reports 2SLS IV estimates, treating CUBIT completion as endogenous, using the overall peer-group completion rate as the IV instead of the completion rate amongst commencements. These results are included for completeness, but recall that the exclusion restriction is more difficult to argue with this IV. Columns D and E both focus on 2SLS using the preferred IV (the contemporaneous completion rate amongst those peers who commenced CUBIT). In each table, the specification in Column D uses a generous suite of control variables, as in the previous models (Columns A, B and C in the same table). To test for the stability of the results, this model is refined in Column E to exclude covariate groups which are not statistically significant at the .05 level in the second-stage re-offending regression.

Re-offending with a sex offence

First consider Table 2, which reports the model estimation results for re-offending with a sex offence.

Column D presents IV results using the preferred IV—the prevailing completion rate amongst commencements. The first-stage regression results (in the lower half of the table) suggest that this preferred IV is a relevant predictor of an individual's own completion (F-statistic = 26.3); a 1 percentage point increase in the completion rate amongst an offender's peer-group commencements, is associated with a 0.963 percentage point increase in the probability that that offender will also (start and) complete treatment. In the second-stage regression, the IV point-estimate of the treatment effect (the estimated coefficient on CUBIT completion) is positive, but the standard errors are large; the coefficient is insignificant at the .05 level.

The Chi-square test of the endogeneity of CUBIT completion in this 2SLS model is statistically significant at the .05 level (Chi-square = 4.4, *p*-value = .035), suggesting that there are unobservable characteristics correlated with both CUBIT completion and, separately, sex re-offending risk. This implies that, despite potentially being more efficient, the results presented in Columns A and B which treat CUBIT completion as exogenous are probably unreliable; the IV approach is necessary.[10]

The model presented in Column D is therefore our preferred model. This model includes a generous suite of control variables: the victim relationship profile, victim age profile, the category of the offender's Most Serious Offence at the index custodial episode, whether the offender has a history of break and enter offences, the offender's sentence length, the offender's ex-ante parole prospects, and the offender's age group and ATSI status. The more statistically significant covariates in this model (Column D) and its OLS benchmark (Column B) all have the expected sign. Generally speaking, risk of re-offending with a sex offence is higher for individuals who had offended against strangers (compared to immediate family), who consistently offended against children (rather than adults, or both children and adults), and who had a proven break-and-enter offence in the past five years (conditional on the other controls). Sentence length, parole prospects, age and ATSI status show little relationship with re-offending with a sex offence amongst this sample (conditional on the other controls).

To test the stability of the results in Column D, a more parsimonious version of this model is presented in Column E, excluding the control variables which are insignificantly different from zero in the second-stage regression at the .05 level. The IV remains relevant in the first-stage regression (F-statistic =17.3) and the estimated coefficient on CUBIT completion remain insignificantly different from zero in the second-stage regression. Further robustness checks on these results (including a version of the model with bootstrapped errors) are presented in the Appendix in Table A1. In particular, these checks illustrate that the results are qualitatively robust to the inclusion of the conditioning variables relating to peer group re-offending risk and time-variation in rates of sex offending designed to better ensure the robustness of the exclusion restriction. For completeness, it is also worth noting that the results are broadly unchanged if the overall CUBIT completion rate is instead employed as the IV (Column C), rather than the completion rate amongst commencements (which more readily satisfies the exclusion restriction).

Re-offending with a violent offence

Now consider Table 3, which reports the model estimation results for re-offending with a violent (including sex) offence.

Column D presents 2SLS results for this alternative outcome variable using the preferred IV—the prevailing completion rate amongst commencements. The first-stage regression suggests that the IV is a relevant predictor of an individual's own completion in this model (F-statistic = 31.6); a 1 percentage point increase in the completion rate amongst an offender's peer-group commencements, is associated with a 0.971 percentage point increase in the probability that that offender will also (start and) complete treatment. The estimate of the treatment effect—the coefficient on CUBIT completion in the second-stage re-offending equation—is positive, small, and insignificantly different from zero.

In contrast to the sex re-offending model, the endogeneity test is nowhere near statistically significant in the violent re-offending model (Chi-square = 0.5, *p*-value = .465), implying that there is no evidence the IV approach is necessary. That is, observable characteristics may provide sufficient protection against any selection bias in treatment status that is correlated with the propensity to re-offend with a violent offence.

Table 2. Model estimation: Re-offending with a sex offence

	Column A		Column B		Column C		Column D		Column E	
					2SLS: Second-stage regression					
Dependent variable: Re-offence with a sex offence within 5 years of free time following release"	**Benchmark probit**		**Benchmark OLS**		**Primary IV: Prevailing overall completion rate**		**Preferred IV: Prevailing completion rate amongst commencements**			
Independent variables	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)
Completed CUBIT	-0.201	(0.215)	-0.028	(0.033)	0.260	(0.174)	0.217	(0.124)	0.292	(0.174)
Victim relationship profile: *Relative to immediate family*										
Stranger	1.215 **	(0.393)	0.193 ***	(0.054)	0.173 **	(0.056)	0.176 **	(0.055)	0.163 **	(0.053)
Acquaintance	0.515	(0.391)	0.060	(0.044)	0.037	(0.048)	0.041	(0.046)	0.038	(0.046)
Extended family	0.530	(0.490)	0.067	(0.073)	0.019	(0.077)	0.026	(0.073)	0.036	(0.079)
Victim age profile: *Relative to adults only / adults and children*										
Children only	0.681 **	(0.251)	0.107 *	(0.048)	0.072	(0.055)	0.077	(0.052)		
Victim profile: includes missing data	0.878 *	(0.424)	0.124 *	(0.057)	0.129 *	(0.064)	0.128 *	(0.062)	0.100	(0.056)
Most Serious Offence at index: *Relative to a sex offence*										
Violent offence	-1.223 *	(0.509)	-0.152 **	(0.048)	-0.142 **	(0.051)	-0.143 **	(0.050)	-0.145 ***	(0.044)
Other	-0.398	(0.456)	-0.055	(0.075)	-0.059	(0.080)	-0.058	(0.078)	0.031	(0.093)
Prior break and enter offence	0.641 *	(0.321)	0.127 *	(0.063)	0.120	(0.062)	0.121 *	(0.061)		
Missing prior history data	0.083	(0.504)	0.019	(0.093)	0.008	(0.098)	0.010	(0.096)		
Sentence length (years): *Relative to less than 3 years*										
3-4 years	-0.422	(0.498)	-0.069	(0.080)	-0.050	(0.090)	-0.053	(0.087)		
5-9 years	-0.218	(0.498)	-0.052	(0.081)	-0.079	(0.091)	-0.075	(0.088)		
10+ years	0.271	(0.590)	0.032	(0.105)	-0.053	(0.126)	-0.041	(0.118)		
Parole prospects: *Relative to court-ordered parole*										
SPA parole prosepct	-0.174	(0.477)	-0.055	(0.054)	-0.046	(0.061)	-0.047	(0.059)		
No parole prospects	1.014	(0.807)	0.280	(0.256)	0.233	(0.282)	0.240	(0.276)		
Aboriginal and/or Torres Strait Islander	0.030	(0.233)	0.008	(0.039)	0.014	(0.040)	0.013	(0.040)		
Age category: *Relative to group aged under 30*										
Aged 30-39	-0.043	(0.313)	-0.021	(0.050)	-0.061	(0.058)	-0.055	(0.055)		
Aged 40-49	0.313	(0.314)	0.036	(0.056)	0.015	(0.060)	0.018	(0.059)		
Aged 50+	-0.172	(0.344)	-0.032	(0.056)	-0.040	(0.062)	-0.039	(0.060)		
Constant	-1.887 ***	(0.557)	0.080	(0.104)	-0.033	(0.130)	-0.016	(0.118)	-0.123	(0.101)
N	347		347		347		347		347	
F/Chi2-statistic on model	38.7 **		1.7 *		1.5		1.5		2.5 *	
AIC	253.4		199.0							
Dependent variable: Completed CUBIT					**2SLS: First-stage regression**					
IV: peer completion rate					0.817 ***	(0.203)	0.963 ***	(0.188)	0.770 ***	(0.185)
Victim relationship profile: *Relative to immediate family*										
Stranger					0.070	(0.088)	0.085	(0.087)	0.069	(0.083)
Acquaintance					0.065	(0.084)	0.069	(0.083)	0.045	(0.082)
Extended family					0.152	(0.114)	0.151	(0.112)	0.176	(0.110)
Victim age profile: *Relative to adults only / adults and children*										
Children only					0.130 *	(0.064)	0.136 *	(0.063)		
Victim profile: includes missing data					-0.005	(0.098)	0.007	(0.097)	-0.045	(0.089)
Most Serious Offence at index: *Relative to a sex offence*										
Violent offence					-0.077	(0.086)	-0.100	(0.087)	0.009	(0.092)
Other					-0.049	(0.125)	-0.061	(0.117)	-0.017	(0.104)
Prior break and enter offence					0.017	(0.087)	0.011	(0.086)		
Missing prior history data					0.095	(0.113)	0.126	(0.111)		
Sentence length (years): *Relative to less than 3 years*										
3-4 years					-0.024	(0.126)	-0.022	(0.130)		
5-9 years					0.115	(0.124)	0.120	(0.128)		
10+ years					0.337 *	(0.145)	0.351 *	(0.149)		
Parole prospects: *Relative to court-ordered parole*										
SPA parole prosepct					0.009	(0.109)	0.018	(0.113)		
No parole prospects					0.170	(0.213)	0.175	(0.216)		
Aboriginal and/or Torres Strait Islander					-0.026	(0.061)	-0.030	(0.061)		
Age category: *Relative to group aged under 30*										
Aged 30-39					0.141	(0.078)	0.145	(0.077)		
Aged 40-49					0.059	(0.083)	0.058	(0.082)		
Aged 50+					0.001	(0.090)	-0.007	(0.088)		
Constant					-0.129	(0.189)	-0.409	(0.211)	-0.030	(0.156)
F-statistic on excluded instrument					16.2 ***		26.3 ***		17.3 ***	
Chi-square statistic on endogenity					3.3	(*p*=.069)	4.4 *	(*p*=.035)	3.6	(*p*=.058)

Notes. OLS = Ordinary Least Squares; 2SLS = Two-stage least squares; IV = Instrumental variable; st.err. = standard error; AIC = Akaike Information Criterion; SPA = State Parole Authority.
* $p<.05$, ** $p<.01$, *** $p<.001$.

Table 3. Model estimation: Reoffending with a violent (including sex) offence

	Column A		Column B		Column C		Column D		Column E	
Dependent variable: Re-offence with a violent offence within 5 years of free time following release	**Benchmark probit**		**Benchmark OLS**		**2SLS: Second-stage regression** **Primary IV: Prevailing overall completion rate**		**Preferred IV: Prevailing completion rate amongst commencements**			
Independent variables	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)
Completed CUBIT	-0.214	(0.162)	-0.065	(0.047)	0.154	(0.211)	0.047	(0.159)	0.123	(0.171)
LSI-R riskband: *Relative to High*										
Low	-0.500	(0.496)	-0.212	(0.149)	-0.265	(0.163)	-0.239	(0.153)	-0.256	(0.154)
Medium-Low	-1.591 ***	(0.477)	-0.422 ***	(0.120)	-0.445 ***	(0.124)	-0.434 ***	(0.120)	-0.456 ***	(0.125)
Medium	-0.451	(0.392)	-0.180	(0.127)	-0.219	(0.133)	-0.200	(0.128)	-0.233	(0.132)
Medium-High	-0.681	(0.406)	-0.255	(0.136)	-0.305 *	(0.146)	-0.281 *	(0.139)	-0.304 *	(0.145)
Missing	-0.519	(0.350)	-0.202	(0.115)	-0.256 *	(0.129)	-0.230	(0.121)	-0.247 *	(0.124)
Prior violent offence finalisations: Relative to none or one										
2	0.084	(0.192)	0.026	(0.055)	0.027	(0.055)	0.027	(0.054)	0.028	(0.055)
3+	0.571 **	(0.220)	0.193 *	(0.081)	0.229 **	(0.087)	0.211 *	(0.083)	0.222 **	(0.084)
Missing	0.150	(0.395)	0.041	(0.118)	0.057	(0.125)	0.049	(0.120)	0.023	(0.109)
Sentence length (years): *Relative to less than 3 years*										
3-4 years	-0.098	(0.384)	-0.017	(0.109)	0.006	(0.112)	-0.005	(0.108)		
5-9 years	0.109	(0.396)	0.040	(0.110)	0.029	(0.111)	0.034	(0.108)		
10+ years	0.062	(0.478)	0.029	(0.132)	-0.032	(0.146)	-0.002	(0.138)		
Parole prospects: *Relative to court-ordered parole*										
SPA parole prosepct	-0.385	(0.357)	-0.117	(0.095)	-0.115	(0.093)	-0.116	(0.092)		
No parole prospects	0.431	(0.686)	0.152	(0.195)	0.104	(0.205)	0.127	(0.197)		
Aboriginal and/or Torres Strait Islander	0.593 ***	(0.167)	0.196 ***	(0.058)	0.198 ***	(0.058)	0.197 ***	(0.057)	0.196 ***	(0.058)
Age category: *Relative to group aged under 30*										
Aged 30-39	0.137	(0.218)	0.052	(0.076)	0.034	(0.076)	0.043	(0.075)	0.034	(0.077)
Aged 40-49	-0.204	(0.236)	-0.052	(0.075)	-0.051	(0.076)	-0.051	(0.074)	-0.048	(0.076)
Aged 50+	-0.612 *	(0.263)	-0.154 *	(0.078)	-0.148	(0.078)	-0.151 *	(0.077)	-0.157 *	(0.075)
Constant	0.246	(0.452)	0.570 ***	(0.152)	0.489 **	(0.171)	0.529 ***	(0.160)	0.421 **	(0.148)
N	369		369		369		369		369	
F/Chi2-statistic on model	76.6 ***		6.2 ***		5.8 ***		6.1 ***		7.5 ***	
AIC	409.3		426.4							
Dependent variable: Completed CUBIT					**2SLS: First-stage regression**					
IV: peer completion rate					0.826 ***	(0.181)	0.971 ***	(0.173)	0.930 ***	(0.175)
LSI-R riskband: *Relative to High*										
Low					0.237	(0.156)	0.250	(0.160)	0.209	(0.171)
Medium-Low					0.098	(0.142)	0.108	(0.142)	0.109	(0.151)
Medium					0.166	(0.132)	0.163	(0.132)	0.134	(0.145)
Medium-High					0.217	(0.134)	0.205	(0.136)	0.184	(0.149)
Missing					0.254 *	(0.120)	0.263 *	(0.122)	0.210	(0.134)
Prior violent offence finalisations: *Relative to none or one*										
2					-0.009	(0.062)	-0.010	(0.061)	-0.023	(0.062)
3+					-0.171 *	(0.078)	-0.177 *	(0.077)	-0.191 *	(0.079)
Missing					-0.041	(0.111)	-0.013	(0.107)	0.254 **	(0.087)
Sentence length (years): *Relative to less than 3 years*										
3-4 years					-0.058	(0.138)	-0.061	(0.142)		
5-9 years					0.067	(0.137)	0.070	(0.142)		
10+ years					0.324 *	(0.154)	0.331 *	(0.160)		
Parole prospects: *Relative to court-ordered parole*										
SPA parole prosepct					0.035	(0.123)	0.048	(0.127)		
No parole prospects					0.174	(0.161)	0.187	(0.166)		
Aboriginal and/or Torres Strait Islander					-0.016	(0.059)	-0.017	(0.058)	-0.010	(0.058)
Age category: *Relative to group aged under 30*										
Aged 30-39					0.097	(0.077)	0.106	(0.077)	0.140	(0.078)
Aged 40-49					-0.004	(0.083)	-0.003	(0.082)	0.002	(0.084)
Aged 50+					-0.049	(0.090)	-0.054	(0.088)	0.006	(0.087)
Constant					-0.172	(0.192)	-0.456 *	(0.218)	-0.319	(0.198)
F-statistic on excluded instrument					20.8 ***		31.6 ***		28.2 ***	
Chi-square statistic on endogenity					1.2	(*p*=.273)	0.5	(*p*=.465)	1.2	(*p*=.274)

Notes. OLS = Ordinary Least Squares; 2SLS = Two-stage least squares; IV = Instrumental variable; st.err. = standard error; AIC = Akaike Information Criterion; SPA = State Parole Authority.
* *p*<.05, ** *p*<.01, *** *p*<.001.

Note that this model includes observable control variables reflecting: the offender's LSI-R score riskband, the number of prior court finalisations in the past five years where violent offences were proven, the offender's sentence length, the offender's ex-ante parole prospects, and the offender's age group and ATSI status. The sign on each of the more statistically significant coefficients in this model (and the benchmark OLS model in Column B) is in the expected direction; a higher risk of re-offending with a violent offence is evident for offenders who are classified by the LSI-R as at high risk rather than lower risk categories, who have three or more court finalisations in the past five years where a violent offence was proven (relative to just one), who identify as Aboriginal or Torres Strait Islanders, and who are younger in age at the time of release (aged under 30 versus aged 50 and above). A slightly more parsimonious version of the model is presented in Column E, and the results are, in the main, unchanged.

The lack of any evidence pointing towards the endogeneity of treatment with respect to violent re-offending suggests that regressions treating CUBIT completion as exogenous could potentially be relied upon to provide more efficient estimate of any treatment impact. Columns A and B present results—using probit and OLS LPM functional forms, respectively—from models which assume that an individual's propensity to complete CUBIT is unrelated to their violent re-offending risk, after controlling for observable characteristics. The AIC implies that the probit specification is preferred for predicting violent recidivism outcomes. In any case, the estimated coefficient on treatment completion is negative, but statistically insignificant at the .05 level in both specifications.

Note that the same conclusions can be drawn if the overall CUBIT completion rate is initially employed as the IV (Column C) rather than the completion rate amongst commencements. Further robustness checks on these results (including a version of the preferred IV model with bootstrapped errors) are presented in the Appendix in Table A2. Most importantly, these checks demonstrate that the qualitative results are robust to the inclusion of variables relating to peer group re-offending risk and violent crime trends designed to better ensure the robustness of the exclusion restriction.

Re-offending with a general offence

Table 4 reports the model estimation results for general re-offending.

Column D presents IV results for the 2SLS model predicting general re-offending using the preferred IV—the prevailing completion rate amongst commencements. The IV is a relevant predictor of an offender's own completion in the first-stage regression (F-statistic = 30.5), with a 1 percentage point increase in the completion rate amongst an offender's peer-group commencements associated with a 0.990 percentage point increase in the probability that that offender will also (start and) complete treatment. In the second-stage re-offending model, the coefficient on CUBIT completion is very close to zero and statistically insignificant; no impact of treatment on general re-offending is detected in this specification.

However, according to an endogeneity test, there is no evidence that IV methods are necessary in modelling general re-offending in this sample, conditional on relevant observable characteristics (Chi-square = 0.7, p-value = .417). The observable characteristics incorporated into this model include: the offender's LSI-R score riskband, victim age profile, whether property offences had been proven in the previous five years, whether a suspended sentence had been received in the previous five years, the offender's minimum non-parole period and sentence length, the offender's ex-ante parole prospects, and the offender's age group and ATSI status. The sign on each of the more statistically significant coefficients is in the expected direction; a higher risk of re-offending with a violent offence is evident for offenders who are classified by the LSI-R as at high risk rather than lower risk categories, who have three or more court finalisations in the past five years where a violent offence was proven (relative to just one), who identify as Aboriginal or Torres Strait Islanders, and who are younger in age at the time of release (aged under 30 versus aged 50 and above).

As in the case of violent re-offending, if observable characteristics adequately control for any selection bias with respect to treatment and general recidivism risk, then regressions treating CUBIT completion as exogenous might provide more efficient estimates of any treatment effect. In both the probit and OLS specifications treating CUBIT completion as exogenous (Columns A and B, respectively), the estimated coefficient on treatment completion is negative, and statistically significant at the .05 level. This suggests that completing CUBIT reduces general re-offending risk. The AIC value is lower for the probit model than the OLS model, so the probit specification is interpreted as our preferred model (Column A).

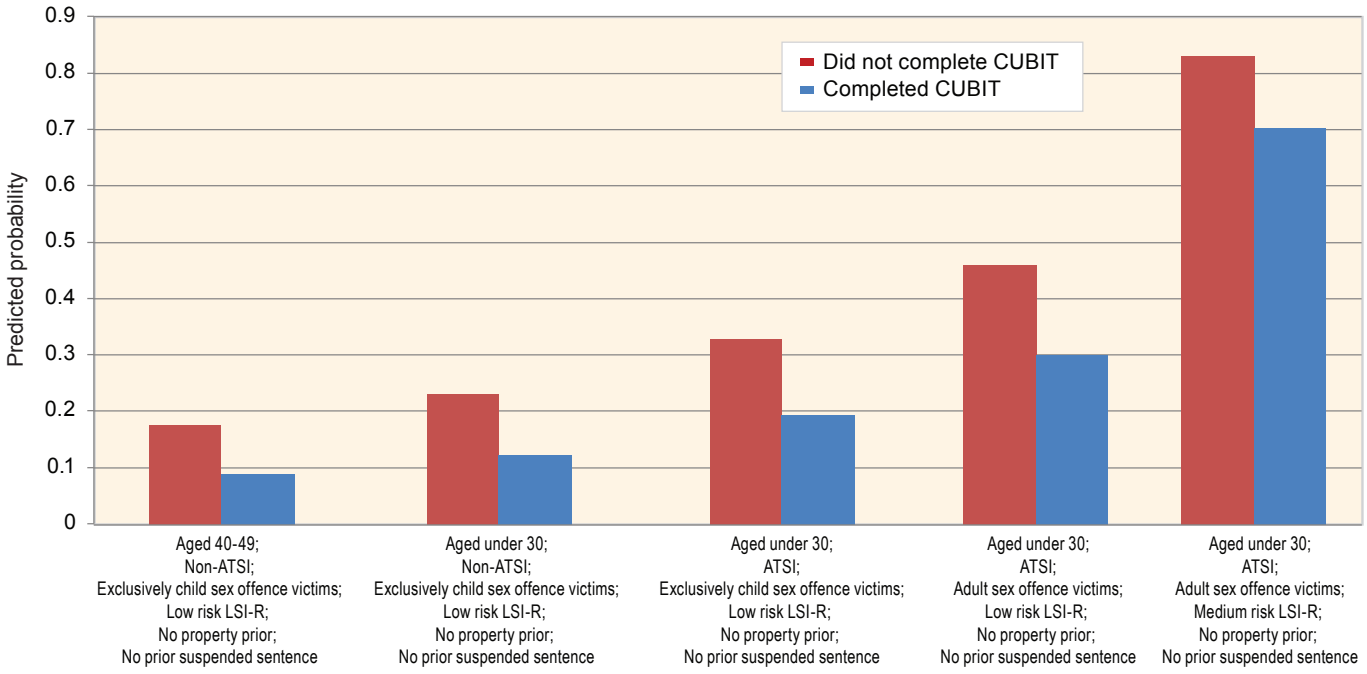
In a probit model, the marginal impact of any one independent variable on the predicted probability of re-offending depends on the value of the other covariates. The magnitude of the treatment impact according to the probit specification in Column A is illustrated under a range of different scenarios in Figure 2. All scenarios presented consider an offender with: a 5-9 year sentence, a non-parole period of 3-4 years, and the prospect of parole being attained through SPA. Other characteristics are iteratively altered from a lower-risk to a higher-risk category in each scenario (from left to right). To take an example, the fourth case considers CUBIT-suitable offenders who are: aged under 30, ATSI identified, whose sex offence victims include adults, who are classified as being at a medium-low risk of re-offending on the LSI-R, and who have not had a proven property offence in the past five years, nor received a suspended sentence. The probit model predicts that there is a 46 per cent probability that offenders with these characteristics will re-offend with a general offence in the five years following their release from custody if they do not complete CUBIT. This probability is estimated to fall by 16 percentage points to 30 per cent, if they complete the CUBIT program prior to release from custody.

Table 4. Model estimation: Reoffending with a general offence

	Column A		Column B		Column C		Column D		Column E	
Dependent variable:					**2SLS: Second-stage regression**					
Re-offence with a general offence within 5 years of free time following release	**Benchmark probit**		**Benchmark OLS**		**Primary IV: Prevailing overall completion rate**		**Preferred IV: Prevailing completion rate amongst commencements**			
Independent variables	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)	Coeff.	(st.err.)
Completed CUBIT	-0.427 **	(0.155)	-0.134 **	(0.048)	0.117	(0.209)	-0.010	(0.162)	-0.026	(0.165)
LSI-R riskband: *Relative to High*										
Low	-0.861	(0.525)	-0.245	(0.146)	-0.294 *	(0.150)	-0.270	(0.144)	-0.300 *	(0.146)
Medium-Low	-1.058 *	(0.447)	-0.321 **	(0.123)	-0.342 **	(0.124)	-0.331 **	(0.120)	-0.363 **	(0.123)
Medium	-0.286	(0.423)	-0.082	(0.122)	-0.128	(0.124)	-0.105	(0.121)	-0.108	(0.124)
Medium-High	0.073	(0.438)	0.038	(0.124)	-0.015	(0.126)	0.011	(0.122)	0.012	(0.124)
Missing	-0.391	(0.382)	-0.117	(0.109)	-0.176	(0.117)	-0.146	(0.111)	-0.153	(0.112)
Victim age profile: *Relative to adults only / adults and children*										
Children only	-0.342	(0.177)	-0.115	(0.059)	-0.143 *	(0.062)	-0.129 *	(0.059)	-0.128 *	(0.058)
Missing	-0.344	(0.206)	-0.105	(0.063)	-0.089	(0.067)	-0.097	(0.063)	-0.093	(0.064)
Prior property offence	0.526 **	(0.193)	0.177 **	(0.064)	0.188 **	(0.066)	0.183 **	(0.064)	0.221 ***	(0.061)
Prior suspended sentence	0.831 *	(0.364)	0.249 **	(0.095)	0.244 *	(0.103)	0.247 *	(0.097)	0.278 **	(0.093)
Missing prior history data	0.323	(0.416)	0.090	(0.118)	0.112	(0.131)	0.101	(0.121)	-0.012	(0.098)
Minimum non-parole period (years): *Relative to less than 3 years*										
3-4 years	0.265	(0.246)	0.078	(0.064)	0.096	(0.067)	0.087	(0.063)		
5-9 years	0.220	(0.291)	0.071	(0.087)	0.076	(0.084)	0.073	(0.083)		
10+ years	-0.796	(0.530)	-0.194	(0.146)	-0.176	(0.154)	-0.185	(0.147)		
Sentence length (years): *Relative to less than 3 years*										
3-4 years	-0.462	(0.396)	-0.137	(0.105)	-0.115	(0.112)	-0.126	(0.106)		
5-9 years	-0.688	(0.453)	-0.193	(0.116)	-0.219	(0.121)	-0.206	(0.115)		
10+ years	-0.411	(0.564)	-0.107	(0.159)	-0.192	(0.173)	-0.149	(0.162)		
Parole prospects: *Relative to court-ordered parole*										
SPA parole prosepct	0.293	(0.368)	0.077	(0.088)	0.073	(0.090)	0.075	(0.087)		
No parole prospects	0.453	(0.713)	0.148	(0.212)	0.075	(0.223)	0.112	(0.214)		
Aboriginal or Torres Strait Islander	0.295	(0.174)	0.102	(0.056)	0.107	(0.058)	0.104	(0.055)		
Age category: *Relative to group aged under 30*										
Aged 30-39	0.146	(0.227)	0.041	(0.075)	0.012	(0.079)	0.027	(0.075)	0.052	(0.074)
Aged 40-49	-0.191	(0.238)	-0.072	(0.078)	-0.079	(0.078)	-0.076	(0.076)	-0.056	(0.075)
Aged 50+	-0.601 *	(0.261)	-0.203 *	(0.086)	-0.204 *	(0.085)	-0.204 *	(0.083)	-0.193 *	(0.082)
Constant	0.792	(0.501)	0.739 ***	(0.152)	0.668 ***	(0.167)	0.704 ***	(0.158)	0.674 ***	(0.139)
N	379 ***		379 ***		379 ***		379 ***		379 ***	
F/Chi2-statistic on model	116.9 ***		8.3 ***		7.6 ***		7.8 ***		10.7 ***	
AIC	455.3		476.7							
Dependent variable:										
Completed CUBIT					**2SLS: First-stage regression**					
IV: peer completion rate					0.876 ***	(0.187)	0.990 ***	(0.179)	0.950 ***	(0.176)
LSI-R riskband: Relative to High										
Low					0.177	(0.157)	0.193	(0.160)	0.167	(0.169)
Medium-Low					0.065	(0.143)	0.077	(0.143)	0.075	(0.150)
Medium					0.167	(0.132)	0.164	(0.133)	0.131	(0.144)
Medium-High					0.188	(0.138)	0.177	(0.140)	0.134	(0.152)
Missing					0.230	(0.120)	0.239	(0.122)	0.189	(0.132)
Victim age profile: *Relative to adults only / adults and children*										
Children only					0.123 *	(0.059)	0.125 *	(0.058)	0.080	(0.058)
Missing					-0.048	(0.072)	-0.046	(0.072)	-0.063	(0.071)
Prior property offence					-0.097	(0.067)	-0.093	(0.066)	-0.090	(0.064)
Prior suspended sentence					-0.060	(0.121)	-0.063	(0.119)	-0.146	(0.118)
Missing prior history data					-0.049	(0.133)	-0.014	(0.129)	0.237 *	(0.093)
Minimum non-parole period (years): *Relative to less than 3 years*										
3-4 years					-0.079	(0.080)	-0.068	(0.079)		
5-9 years					-0.020	(0.097)	-0.022	(0.096)		
10+ years					-0.091	(0.165)	-0.113	(0.163)		
Sentence length (years): *Relative to less than 3 years*										
3-4 years					-0.031	(0.129)	-0.036	(0.131)		
5-9 years					0.126	(0.146)	0.125	(0.149)		
10+ years					0.391 *	(0.175)	0.404 *	(0.177)		
Parole prospects: *Relative to court-ordered parole*										
SPA parole prosepct					0.056	(0.116)	0.064	(0.117)		
No parole prospects					0.240	(0.140)	0.254	(0.142)		
Aboriginal and/or Torres Strait Islander					-0.007	(0.061)	-0.010	(0.060)		
Age category: *Relative to group aged under 30*										
Aged 30-39					0.111	(0.077)	0.121	(0.076)	0.143	(0.076)
Aged 40-49					0.004	(0.082)	0.011	(0.081)	0.017	(0.082)
Aged 50+					-0.047	(0.091)	-0.047	(0.090)	0.010	(0.088)
Constant					-0.263	(0.198)	-0.533 *	(0.226)	-0.346	(0.197)
F-statistic on excluded instrument					21.9 ***		30.5 ***		29.2 ***	
Chi-square statistic on endogenity					1.7	(*p*=.193)	0.7	(*p*=.416)	0.5	(*p*=.500)

Notes. OLS = Ordinary Least Squares; 2SLS = Two-stage least squares; IV = Instrumental variable; st.err. = standard error; AIC = Akaike Information Criterion; SPA = State Parole Authority.
* *p*<.05, ** *p*<.01, *** *p*<.001.

Figure 2. General reoffending within five years free time: Probit model predicted probabilities



Note. All scenarios assume a sentence of 5-9 years with a non-parole period of 3-4 years, the prospect of parole through SPA, and no property offences proven nor suspended sentences received in the previous five years

Given the sample characteristics, the probit model (assuming treatment is exogenous) implies an average marginal treatment effect of 13.2 percentage points; that is, completing CUBIT is estimated to reduce the risk of re-offending from 55.1 per cent to 42.0 per cent. It is reassuring to note that a very similar figure arises in the benchmark OLS model (Column B). In the OLS model, completion of CUBIT is associated with a 13.4 percentage point reduction (on average) in the probability of re-offending during the first five years following release from custody.

Note that we are able to arrive at the same conclusion under a number of variations to the IV estimation method. In particular, no evidence of endogeneity of treatment with respect to general re-offending is detected in a more parsimonious version (Column E) of the main IV model, nor in any of the robustness checks presented in the Appendix (Table A3). In particular, these checks show that the results are qualitatively robust to the inclusion of conditioning variables, relating to peer group re-offending risk and general crime trends, designed to better ensure the validity of the exclusion restriction. The results are also unaffected if the overall completion rate is used as the IV (Column C) in place of the completion rate amongst commencements. Lastly, variations presented in the Appendix (Table A4) demonstrate that the findings from the probit and OLS regressions persist in a more parsimonious version of the model, and are robust to the inclusion of parole supervision control variables.

DISCUSSION

CUBIT is a residential, particularly intensive form of sex offender treatment, targeted at offenders with moderate to high risk/needs. The results presented in this paper suggest that, for offenders found suitable for CUBIT, participating in and completing treatment reduces 5-year general recidivism risk by around 13 percentage points on average (from 55 to 42 per cent for the sample as a whole). This conclusion is drawn from simple regressions which treat CUBIT completion as exogenous, controlling for a range of important observable risk factors. Robustness checks demonstrate that these results hold in a more parsimonious specification, and are qualitatively robust to the choice of functional form between a probit or LPM (around the 'average' domain at which marginal effects from these models might be compared).

Ordinarily, there would be some valid concern that findings from this sort of model might reflect unobservable selection effects—unobservable characteristics correlated with the likelihood that an offender will complete treatment, and which separately relate to recidivism risk (irrespective of treatment). However, IV methods are explored in this paper, and results from that approach (Table 6, Columns C-E) lend confidence to a causal interpretation of the simple regression predicting general re-offending treating CUBIT treatment as exogenous. Specifically, tests of the endogenity of treatment completion in

the IV-based general re-offending models were nowhere near statistically significant, suggesting that the observable variables are sufficient to control for selection bias with regard to general re-offending risk amongst this sample. Still, the results from the regression models should be interpreted with some caution, as the endogeneity tests do not explicitly provide evidence to suggest treatment is conditionally exogenous in the general re-offending LPM, but rather fail to provide evidence of endogenity.

Meanwhile, this study finds no evidence that CUBIT reduces sexual recidivism. However, this could reflect limitations of the statistical methods employed, rather than the absence of an impact necessarily. The statistical power of the analyses of sexual recidivism in this paper is limited by the relatively small sample, and low incidence of reconviction for a sex offence (recall that the recidivism rate for sex re-offending is 11.8 per cent over five years in this sample). Even if CUBIT treatment were randomly assigned, with this sample, there would be just a 60 per cent chance of detecting statistical significance at the .05 level on a treatment effect which halved the sexual recidivism rate from 15.0 to 7.5 per cent. In the main estimation results, none of the sex re-offending models, including those assuming treatment is (conditionally) exogenous, showed a statistically significant coefficient on CUBIT completion at the .05 level. In fact, many of the sex re-offending models struggle to achieve statistical significance overall.

The statistical power of the sex re-offending analyses is further hampered by following an IV approach (because efficiency is lost in predicting completion in the first-stage regression), but unfortunately this appears to be necessary in the context of sex re-offending. The IV results (Table 2, Columns C-E) suggest that observable characteristics are probably inadequate to confidently assume treatment completion is exogenous to sex re-offending risk in the models. Endogeneity tests were not convincingly statistically insignificant in models that included controls for a range of observable potential risk factors. Amongst the CUBIT-suitable cohort, offenders with unobservable characteristics that place them at lower risk of sexual recidivism appear to be more likely to complete treatment. The IV methods are used specifically to control for this selection bias, but unfortunately they are less efficient than a standard OLS regression.

Notwithstanding the challenges posed by the limited sample size, it is interesting to note that the point-estimate on the treatment variable in the sex re-offending IV specifications is positive and quite large in magnitude relative to the standard errors.[11] If the true Local Average Treatment Effect of CUBIT is to reduce sexual recidivism, this point-estimate suggests that the impact might not be very large. Of course, it is not inconceivable that intensive sex-offending-focused therapy might aggravate sexual recidivism risk in some circumstances, as investigated by Lovins, Lowenkamp, and Latessa (2009) and Barbaree, Langton, and Peacock (2006) (op. cit. Feix, 2006), and studied in ethnographic detail by Lacombe (2008). In the present context, with the preferred IV (the contemporaneous completion rate amongst CUBIT commencements), the cohort sensitive to the IV might include inmates who, without considerable encouragement, would tend to drop out of the program or to be discharged due to a lack of treatment gains.[12] The LATE estimated in the IV specifications could conceivably be capturing the potentially adverse impact of completing CUBIT for this particular cohort. Regardless, this might be an avenue worth pursuing in future research.[13]

This paper also finds no evidence to conclude that completing CUBIT reduces re-offending with a violent offence. This is despite the fact that the IV results suggest methods to control for endogeneity may not be necessary in modelling violent re-offending. Although limited statistical power remains an issue in the context of violent re-offending, it is somewhat less deleterious than in the sex re-offending analyses (with random assignment, there would be a 99 per cent chance of identifying a halving of the violent offending rate from 40 to 20 per cent at the .05 level, and a one-in-two chance of detecting a 10 percentage point decline, from 35 to 25 per cent). In any case, in all the violent re-offending model specifications tested, including those assuming CUBIT completion is conditionally exogenous to violent recidivism risk (where standard errors are not insurmountably large), the estimated coefficient on CUBIT completion is statistically insignificant at the .05 level.

A key limitation of this investigation stems from the fact that we are technically unable to directly control for the extent to which offenders are supervised under parole following their release from custody. Including parole in the models could confound the estimate of the treatment effect, because parole is not exogenous in the current context. To the contrary, for many offenders in the sample, the likelihood of attaining parole is directly affected by their participation or otherwise in custody-based sex offender treatment programs; offenders who complete treatment are more likely to be released early onto parole, all else being equal (in particular, controlling for the *ex-ante prospect* of parole). If offenders on parole are less likely to commit new offences due to the parole supervision itself, then there is a risk that this effect is being captured by the coefficient on treatment completion in the main specifications presented in this paper. This would tend to bias results to imply treatment is more effective in reducing recidivism than is actually the case. On the other hand, if offenders who are being supervised under parole are more likely to be caught committing new offences (all else being equal), then the bias would work in the opposite direction. It is somewhat encouraging to note that the results on the treatment impact for general re-offending are qualitatively unchanged if the actual duration of parole is included as a control variable (this is demonstrated through the robustness checks in the Appendix). Nonetheless, future research might benefit from separately considering treatment, re-offending *and parole* as jointly determined, to better address this issue. Unfortunately, to employ the methods used in the present paper, a researcher would need to identify a supplementary valid IV.

The impact evaluation results presented in this paper rely on the validity of the IVs used in the analysis. Our preferred IV is the second IV tested: the completion rate amongst commencements

for an offender's peer cohort. We argue that this variable is unrelated to recidivism (sex, violent, and general, respectively), except through its correlation with treatment completion—that this IV satisfies the exclusion restriction. To support this argument, several robustness checks were carried out (in the Appendix) to include additional control variables whose absence might undermine the exclusion restriction. Specifically, we tested the inclusion of controls providing a proxy measure of the peer group risk (in an attempt to capture peer effects which might influence the propensity to undertake treatment and, separately, recidivism risk, or any time-varying changes in unobservable risk amongst the CUBIT-suitable cohort), and a variable capturing sex, violent or general crime trends across NSW (which might capture some omitted time-varying macro-level variable related to both re-offending risk and CUBIT completion rates). In each specification, the qualitative results which rest on the relevance of the IV were robust to the inclusion of these additional controls. Nonetheless, the exclusion restriction remains an untestable assumption.

In conclusion, the available evidence suggests that participation in CUBIT may reduce general recidivism risk considerably, but no evidence is found to suggest a reduction in violent re-offending risk nor sex re-offending risk—though investigations on the latter front in particular are hampered by methodological limitations. It is possible that the CUBIT program is successful in reducing general re-offending on average, but that it does not work effectively to prevent sex and violent re-offending. It is also possible that CUBIT has greater impact on offenders at lower risk of sexual or violent recidivism (with the impact of treatment therefore borne out in the general re-offending analysis only). On the other hand, equally likely is that our finding in relation to general recidivism is indicative of treatment impacts on sex and violent re-offending that our statistical methods are simply unable to identify in the present sample. It should be dutifully acknowledged that this evaluation makes no attempt to address the all-important question of: "What kinds of treatment work for what kinds of offenders under what conditions?" (Marques et al., 2005, p. 80). A process evaluation of CUBIT currently underway at Corrective Services NSW (including Howard, Manuscript in preparation) is likely to improve our understanding of the myriad of factors that influence program effectiveness.

ACKNOWLEDGEMENTS

The author would like to extend a sincere thank you to Mark Howard from Corrective Services NSW for his time and generosity in sharing findings from his ongoing research to help inform this study. For complementary input, thanks are also due to Simon Corben, Jennifer Galouzis, Jayson Ware, and the many staff whose intimate involvement and knowledge of CUBIT operations provided the background to this work. The author is also grateful to Don Weatherburn, Suzanne Poynton and the research officers at BOCSAR for endless helpful discussions; an anonymous referee for a range of constructive comments and insight on methodology; Mark Ramsay and Nicole Mahoney for assistance collating data; and Florence Sin for desktop publishing this report.

NOTES

1 This information is sourced variously from Ware and Bright (2008), Woodrow and Bright (2011), Corrective Services NSW (2015b) and discussions with staff at Corrective Services NSW.

2 Grady, Edwards, Pettus-Davis, and Abramson (2013) also found little difference in five-year recidivism rates across a matched sample of treatment volunteers (some of whom received treatment) and non-volunteers in a study of sex offenders in North Carolina. This study employed propensity-score matching (PSM) to control for unobservable characteristics, so, as the authors dutifully acknowledge, the risk of biases from unobservable selection effects remain.

3 Dubious matches included those where individuals had no recorded criminal history (that is, no court appearances where an offence was proven within the five years prior to and including the index custodial episode), despite having entered custody from 1994 onwards (when ROD records commenced); cases where individuals had a prior court appearance with a proven offence, but no penalty that would likely lead into a custodial episode; and, cases where individuals who entered custody since 1994 had no prior sex offending history, yet OIMS data suggest that their Most Serious Offence was a sex offence. A few additional cases included data pertaining to a period outside the custodial episode window (for example, a referral date subsequent to the episode end date). These were also flagged and excluded from the analyses. Lastly, the IVs could not be calculated for observations with no record of an offender's earliest possible release date, so these cases were also flagged for omission; this reduced the final sample size by just one observation.

4 The relevance of the IV in predicting the treatment (the endogenous variable) will also affect the statistical power of the analysis (and the bias of the treatment impact estimator). If some treatment gains are realised by individuals who started but did not complete CUBIT, and the IV better predicts commencement than completion, it is possible defining treatment by commencement would better facilitate the identification of any impact. This alternative measure was investigated in the formative stages of this study, but both the relevance assumption and the exclusion restriction intrinsic to the IV approach were more readily satisfied in defining treatment by completion.

5 I also considered using the prevailing CUBIT commencement rate as an IV, but it did not have a strong a relationship with individuals' own completion of treatment.

6 In formative stages of the analysis, I also considered IVs based on the prevailing completion rates for cohorts with earliest release dates *prior* to the index offender. I considered defining the peer group using the 12-month period immediately prior to the offender's earliest possible

release date, and the 12-month period six months earlier still (spanning the period from 18-months to 6 months before the individual's own earliest possible release date). As one might expect, these alternative IVs were less powerful predictors of an individual's treatment outcome. Moreover, they do not entirely ameliorate the risk of reverse causality, since the anticipated end of a custodial episode is an imperfect indicator of the potential timing of treatment.

7 There are also other ways in which the general risk level amongst an offender's peer group might be related to their propensity to complete CUBIT and recidivism risk (regardless of treatment), plus to the peer group completion rate—potentially undermining the IV. For example, staff could conceivably prioritise higher needs offenders when treatment places are limited; alternatively, higher risk offenders might be more inclined to drop out of treatment prior to completion, creating capacity for others. It is also possible that the criteria used to determine whether an offender has moderate-to-high risks/need—and is therefore suitable for CUBIT—has evolved over time. In this case, the observable average risk level amongst an offender's peer group could provide a signal of their own otherwise unobservable risk. The robustness check also attempts to provide some assurance that the exclusion restriction is robust to these possibilities.

8 This is arguably endogenous to an offender's own risk through their realised re-offending. However, this influence would be very small. Consider sex offending, for example. There are 386 offenders in our sample, released from custody over the course of a decade, 41 of whom re-offended with a sex offence within five years of their release. Meanwhile, recorded crime statistics indicate that the order of 800 sexual offences were recorded across NSW on average each month over the relevant period.

9 LPMs have several drawbacks. Notably predicted values may lie outside the unit interval, which is a key reason why binary outcome variables are commonly modelled using probit or logit regression. Unfortunately, the consistency of an analogous IV estimator under the non-linearity assumptions implicit in probit or logit models has not been demonstrated. (Methods such as Newey's (1987) minimum chi-squared estimator do exist for the case when the outcome variable is binary but the endogenous regressor continuous, as per Stata's 'ivprobit' command, but these are not applicable here.) LPMs are an accepted approach for IV estimation in the current context with a binary outcome and a binary endogenous variable (Chiburis, Das, & Lokshin, 2011; Wooldridge, 2010, section 15.2) . Another potential issue arises due to the heteroskedasticity of errors in LPMs, but this can be accounted for though using heteroskedasticity-robust options (specifically, estimating Eicker-Huber-White sandwich estimator of variance; Eicker, 1967; Huber, 1967; White, 1980).

10 For interest's sake, the coefficient on CUBIT completion in the second-stage regression in Column D can be compared to the analogous coefficient in the LPM in Column B (noting, however, that the IV approach estimates a LATE and the OLS approach estimates an ATE). The fact that the IV estimate is considerably larger than the OLS estimate suggests that individuals who have unobservable characteristics associated with a lower sexual recidivism risk, might be more likely to complete treatment (all else being equal; in particular, if the LATE under the IV is equivalent to the true ATE).

11 Note that this does not appear to be driven by any biases introduced by more marked censoring in the sex re-offending analyses. As discussed in the study sample section, broadly similar proportions of treated and untreated offenders are censored from each of the sex, violent and general re-offending analyses.

12 One of the anticipated benefits of the shift from a closed- to open-group format in 2005—after which point CUBIT completion rates steadily increased—was that temporary suspensions would be a more practical way to manage "repeatedly disruptive and unresponsive sexual offenders" (Ware & Bright, 2008, p. 344), and therefore discharges would become less common.

13 It is beyond the scope of this paper to fully develop alternative methods (robust to endogeneity of treatment completion) which might estimate the ATE for the sample as a whole. However, a first-pass bivariate probit specification presented in the robustness section (Table A1, Column G) also estimates a positive coefficient on CUBIT completion in the sex re-offending model.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Andrews, D. A., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised: User's manual*. Toronto: Multi-Health Systems Inc.

Australian Bureau of Statistics. (2011a). *Australian and New Zealand Standard Offence Classification (ANZSOC)* (cat. no. 1234.0). Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/mf/1234.0

Australian Bureau of Statistics. (2011b). *Census of population and housing: Socio-Economic Indexes for Areas (SEIFA), Australia* (cat. no. 2033.0.55.001). Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001

Australian Bureau of Statistics. (2015). *Australian demographic statistics* (cat. no. 3101.0). Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Mar%202015?OpenDocument

Barbaree, H. E., Langton, C., & Peacock, E. (2006). Sexual offender treatment for psychopaths: Is it harmful. In W. L. Marshall, Y. M. Fernandez, L. E. Marshall & G. A. Serran (Eds.), *Sexual offender treatment: Controversial issues* (pp. 159-171). West Sussex, UK: John Wiley & Sons, Ltd.

Baum, C. F., Dong, Y., Lewbel, A., & Yang, T. (2012). Binary choice models with endogenous regressors*. Paper presented at the Stata Conference 2012*, San Diego.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2010, 23 August). ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. Retrieved from http://ideas.repec.org/c/boc/bocode/s425401.html

Bayer, P., Hjalmarsson, R., & Pozen, D. (2009). Building criminal capital behind bars: Peer effects in juvenile corrections. *Quarterly Journal of Economics, 124*(1), 105-147.

Broadhurst, R., & Loh, N. (2003). The probabilities of sex offender re-arrest. *Criminal Behaviour and Mental Health, 13*(2), 121-139.

Chiburis, R. C. (2009, 1 November). scoregof: Stata module to perform Score test of normality for probit and bivariate probit. Retrieved from https://ideas.repec.org/c/boc/bocode/s457465.html

Chiburis, R. C., Das, J., & Lokshin, M. (2011). *A practical comparison of the bivariate probit and linear IV estimators* (Policy Research Working Paper, No. 5601). Retrieved 21 Dec 2015 from The World Bank website: https://openknowledge.worldbank.org/bitstream/handle/10986/3368/WPS5601.pdf?sequence=1

Corrective Services NSW. (2015a). *Classification, Placement and Case Plan Reviews* (1.3: March 2015 Edition). Retrieved 12 Nov 2015 from http://www.correctiveservices.justice.nsw.gov.au/Documents/Related%20Links/open-access-information/offender-classification/14.1-Classification-Placement-and-Case-Plan-Reviews.pdf

Corrective Services NSW. (2015b). *Institutional programs for individuals who have committed sexual offences* (5 September 2015 Edition). Retrieved from Department of Justice website:

Damm, A. P., & Gorinas, C. (2013). Deal drugs once, deal drugs twice: Peer effects on recidivism from prisons. *Essays on Marginalization and Integration of Immigrants and Young Criminals—A Labor Economics Perspective.* Denmark: Aarhus University.

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1*(1), 59-82.

Feix, J. (2006). [Review of the book *Sexual Offender Treatment: Controversial Issues*, by W. L. Marshall, Y. M. Fernandez, L. E. Marshall & G. A. Serran (Eds.)]. *Journal of the American Academy of Psychiatry and the Law Online, 35*(1), 139-142.

Gelb, K. (2007). *Recidivism of sex offenders*. Retrieved 21 Oct 2015 from Sentencing Advisory Council website: https://www.sentencingcouncil.vic.gov.au/sites/default/files/publication-documents/Recidivism%20of%20Sex%20Offenders%20Research%20Paper.pdf

Grady, M. D., Edwards, D., Pettus-Davis, C., & Abramson, J. (2013). Does volunteering for sex offender treatment matter? Using propensity score analysis to understand the effects of volunteerism and treatment on recidivism. *Sexual Abuse: A Journal of Research and Treatment, 25*(4), 319-346.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50*(4), 1029-1054.

Hanson, R. K., Gordon, A., Harris, A. J., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*(2), 169-194.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119-136.

Harkins, L., & Beech, A. (2007). Measurement of the effectiveness of sex offender treatment. *Aggression and Violent Behavior, 12*(1), 36-44.

Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). Static-99 coding rules revised - 2003. Retrieved from http://www.static99.org/pdfdocs/static-99-coding-rules_e.pdf

Helmus, L., Babchishin, K. M., Hanson, R. K., & Thornton, D. (2009). Static-99R: Revised age weights. Retrieved from www.static99.org/pdfdocs/static-99randage20091005.pdf

Howard, M. V. A. (Manuscript in preparation). *Process evaluation of the Custody Based Intensive Treatment (CUBIT) programs for sex offenders: Factors associated with program completion*. Corrections Research Evaluation and Statistics, Corrective Services NSW.

Hua, J., & Fitzgerald, J. (2006). *Matching court records to measure reoffending* (Crime and Justice Bulletin No. 95). Retrieved from NSW Bureau of Crime Statistics and Research website: www.bocsar.nsw.gov.au/Documents/CJB/cjb95.pdf

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1*(1), 221-233.

Lacombe, D. (2008). Consumed with sex: The treatment of sex offenders in risk society. *British Journal of Criminology, 48*(1), 55-74.

Lievore, D. (2004). *Recidivism of sexual assault offenders: Rates, risk factors and treatment efficacy*. Retrieved 18 Nov 2014 from Australian Institute of Criminology website: http://www.aic.gov.au/media_library/archive/publications-2000s/

recidivism-of-sexual-offenders-rates-risk%20factors-and%20treatment-efficacy.pdf

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1*(1), 117-146.

Lovins, B., Lowenkamp, C. T., & Latessa, E. J. (2009). Applying the risk principle to sex offenders: Can treatment make some sex offenders worse? *The Prison Journal, 89*(3), 344-357.

Macgregor, S. (2008). *Sex offender treatment programs: Effectiveness of prison and community based programs in Australia and New Zealand*. Retrieved 25 Aug 2014 from Indigenous Justice Clearinghouse website: http://indigenousjustice.gov.au/briefs/brief003.pdf

Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & Van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment,, 17*(1), 79-107.

Murphy, A. (2007). Score tests of normality in bivariate probit models. *Economics Letters, 95*(3), 374-379.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics, 36*(3), 231-250.

NSW Bureau of Crime Statistics and Research (BOCSAR). (2001). *New South Wales Recorded Crime Statistics 2000*. Retrieved from http://www.bocsar.nsw.gov.au/Documents/rcs00.pdf

NSW Bureau of Crime Statistics and Research. (2015). New South Wales Recorded Crime Statistics 2014. Retrieved from http://www.bocsar.nsw.gov.au/Documents/RCS-Annual/rcs2014.pdf

NSW State Parole Authority. (2012). *Operating Guidelines* (June 2012 Edition). Retrieved 8 Dec 2015 from http://www.correctiveservices.justice.nsw.gov.au/Documents/State-Parole-Authority-Operating-Guidelines.pdf

Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence, 10*(1), 85-105.

Rice, M. E., & Harris, G. T. (2003). The size and sign of treatment effects in sex offender therapy. *Annals of the New York Academy of Sciences, 989*(1), 428-440.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica, 26*(3), 393-415.

Schmucker, M., & Lösel, F. (2008). Does sexual offender treatment work? A systematic review of outcome evaluations. *Psicothema, 20*(1), 10-19.

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics, 20*(4), 518-529.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. Cambridge: Cambridge University Press.

Ware, J., & Bright, D. A. (2008). Evolution of a treatment programme for sex offenders: Changes to the NSW Custody-Based Intensive Treatment (CUBIT). *Psychiatry, Psychology and Law, 15*(2), 340-349.

Ware, J., Frost, A., & Hoy, A. (2010). A review of the use of therapeutic communities with sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 54*(5), 721-742.

Watkins, I. (2011). *The utility of Level of Service Inventory-Revised (LSI-R) assessments within NSW correctional environments* (Research Bulletin No. 29). Retrieved 23 Dec 2015 from Corrective Services NSW website: http://www.correctiveservices.justice.nsw.gov.au/Documents/utility-of-level-of-service-inventory-.pdf

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817-838.

Woodrow, A. C., & Bright, D. A. (2011). Effectiveness of a sex offender treatment programme: A risk band analysis. *International journal of offender therapy and comparative criminology, 55*(1), 43-55.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

APPENDIX

Further descriptive statistics for the study sample and details regarding robustness checks are available electronically at www.bocsar.nsw.gov.au

Other titles in this series

No.192 The NSW Intensive Drug and Alcohol Treatment Program (IDATP) and recidivism: An early look at outcomes for referrals

No.191 Participation in PCYC Young Offender Programs and Re-offending

No.190 Does a prison sentence affect future domestic violence reoffending?

No.189 Assessing the risk of domestic violence recidivism

No.188 Can child protection data improve the prediction of re-offending in young persons?

No.187 Does the first prison sentence reduce the risk of further offending?

No.186 The impact of the NSW Intensive Supervision Program on recidivism

No.185 That's entertainment: Trends in late-night assaults and acute alcohol illness in Sydney's Entertainment Precinct

No.184 Trial court delay and the NSW District Criminal Court

No.183 Lockouts and Last Drinks

No.182 Public confidence in the New South Wales criminal justice system: 2014 update

No.181 The effect of liquor licence concentrations in local areas on rates of assault in New South Wales

No.180 Understanding fraud: The nature of fraud offences recorded by NSW Police

No.179 Have New South Wales criminal courts become more lenient in the past 20 years?

No.178 Re-offending on parole

No.177 Understanding the relationship between crime victimisation and mental health: A longitudinal analysis of population data

No.176 The impact of intensive correction orders on re-offending

No.175 Measuring recidivism: Police versus court data

No.174 Forecasting prison populations using sentencing and arrest data

No.173 Youths in custody in NSW: Aspirations and strategies for the future

No.172 Rates of recidivism among offenders referred to Forum Sentencing

No.171 Community Service Orders and Bonds: A Comparison of Reoffending

No.170 Participant Satisfaction with youth Justice Conferencing

No.169 Does CREDIT reduce the risk of re-offending?

No.168 Personal stress, financial stress, social support and women's experiences of physical violence: A longitudinal analysis

No.167 Police use of court alternatives for young persons in NSW

NSW Bureau of Crime Statistics and Research - Level 1, Henry Deane Building, 20 Lee Street, Sydney 2000
bcsr@justice.nsw.gov.au • www.bocsar.nsw.gov.au • Ph: (02) 8346 1100 • Fax: (02) 8346 1298
ISSN 1030-1046 (Print) ISSN 2204-5538 (Online) • ISBN 978-1-925343-22-9