

Evaluation of the EQUIPS Domestic Abuse Program

Sara Rahman and Suzanne Poynton

Aim: To estimate an unbiased effect of commencing the EQUIPS Domestic Abuse Program (DVEQUIPS) on general re-offending and domestic violence (DV)-related re-offending.

Method: To deal with omitted variable bias, instrumental variables approaches are used to examine general and DV re-offending within 12 months free time after program referral for DVEQUIPS starters and offenders who were referred but did not start (non-starters). The program was delivered as a closed group which permitted the use of the peer non-commencement rate as an instrumental variable (IV), to exploit differences in participation based on the number of other offenders available to start the program at the time of an offender's referral.

Results: Instrumental variables analyses indicated that selection bias was not a major problem, supporting the use of logistic regression to compare re-offending outcomes for starters and non-starters. The treatment effects estimated for offenders who started the program were not significantly different from zero for either general or DV-related re-offending within 12 months of referral.

Conclusion: There is no evidence of a treatment effect for those who start DVEQUIPS within 12 months of referral when compared with those offenders who were referred but did not start.

Keywords: Domestic violence, program evaluation, instrumental variables

INTRODUCTION

With nearly one in four Australian women reporting being a victim of partner violence since the age of 15 (Australian Bureau of Statistics, 2017) and an estimated annual cost of \$22 billion (KPMG, 2016), domestic violence (DV) has become a major policy area in Australia, and is now identified as one of 12 Premier's Priorities for NSW. Consequently, there is now greater focus on the need to identify effective responses and interventions to reduce DV re-offending and its subsequent impact on victims, families and the wider community. Rehabilitation programs, which aim to change offenders' violent behaviour, are a key component of DV prevention policy. Historically DV perpetrator intervention programs have been driven by the Duluth model (Pence & Paymar, 1993).¹ Based on feminist theory, this model focuses on male offenders' accountability for abuse and considers gendered stereotypes as the root cause of the controlling and dominating behaviour exhibited by men over their partners. More recent DV perpetrator treatment programs have utilised cognitive-behavioural therapy (CBT) techniques. CBT-based approaches

view offending as a learned behaviour which can be addressed through identification of erroneous cognitions and beliefs that lead to problematic behaviour and the development of appropriate cognitive and emotional skills to apply in high-risk situations. DV perpetrator intervention programs will often combine elements of both these theoretical approaches (Day, Chung, O'Leary, & Carson, 2009; Eckhardt, Murphy, Black & Suhr, 2006).

In Australia, perpetrator rehabilitation programs are delivered by criminal justice agencies and community organisations. Programs from community organisations are generally voluntary, do not have strict eligibility criteria, and participants are referred from various sources. The programs delivered by criminal justice agencies are typically run within correctional facilities and community corrections centres and target offenders with sentences of 12 months or more as the programs tend to be five months or longer in duration. Sometimes participation is a condition of the offender's sentence or parole order (Mackay, Gibson, Lam, & Beecham, 2015). Programs can differ in content (the extent to which elements of different theoretical approaches

are used), duration, the level of victim involvement (whether parallel support programs are mandated for offenders' partners, or partners attend sessions with offenders) and whether they are delivered to groups or to individuals. Longer programs, such as those delivered by corrective services agencies, tend to be delivered in group settings, with facilitators guiding offenders through structured sessions that progressively move through themes or modules. This structure means that offenders are expected to commit to the program for a long time, and are expected to complete many sessions before they can be considered to have completed the program.

In NSW, EQUIPS (which stands for Explore, Question, Understand, Investigate and Practise, Plan, Succeed) Domestic Abuse (hereafter referred to as DVEQUIPS) is a behaviour change program offered to medium to high-risk DV offenders (as assessed by the Level of Service Inventory-Revised (LSI-R)) who are serving custodial or community-based sanctions and have a current intimate partner violence (IPV) offence or a history of such offences. The program involves 20 two-hour sessions (40 hours) grouped into five modules delivered on a weekly or biweekly basis by trained facilitators approved by Corrective Services NSW (CSNSW). Some of these modules are based on Duluth concepts (e.g. sexual respect and relationship skills, identifying abuse) while others involve elements of CBT (e.g. offence mapping, managing emotions beliefs and attitudes). DVEQUIPS commenced in 2015 and superseded an earlier CSNSW DV program known as the Domestic Abuse Program (DAP). While resembling DAP in terms of the core modules delivered, DVEQUIPS is an offence-specific program which forms part of a larger suite of EQUIPS programs offered to offenders under the supervision of community corrections. This change enabled DVEQUIPS to be combined with EQUIPS programs addressing other criminogenic needs (such as EQUIPS Foundation or EQUIPS Addiction) to increase dosage, and potentially treatment effects, particularly amongst high-risk offenders (Corrective Services NSW, 2016).

RELATED LITERATURE

Four systematic reviews of the DV literature have generally produced mixed evidence on the effectiveness of DV perpetrator interventions. Two of the earliest reviews were able to identify only a small number of rigorous experimental studies. Smedslund and colleagues (2007) conducted a meta-analysis of six randomised controlled trials of CBT-based interventions; only one study found evidence for a statistically significant treatment effect. Feder, Wilson, and Austin (2008), broadening their criteria to include both experimental and quasi-experimental studies, reviewed 10 evaluations of DV programs undertaken in North America. While these experiments provided evidence for a small beneficial treatment effect when outcomes were measured using officially recorded crime, there was little evidence for an impact based on victim-reported abusive behaviour by their partners.

A systematic review by Eckhardt et al. (2013) included 39 DV evaluations which used clearly identified comparison groups or methods to control for selection effects when comparing outcomes for program starters with non-starters. The perpetrator interventions assessed in this review included traditional Duluth programs and CBT programs, as well as alternative programs and brief interventions. The effectiveness of each of these models was reported separately where possible. Eckhardt et al. (2013) found no difference in program effectiveness amongst the more rigorous experimental studies but a positive effect of perpetrator interventions from studies employing a quasi-experimental design, particularly where the DV program was compared to a no intervention group. The quasi-experimental studies suffered from methodological flaws, such as low sample sizes, poor control groups and lack of controls for confounding factors. Approaches other than Duluth or CBT, which addressed motivation and readiness to change, showed promising results. However, the interventions in the studies reviewed tended to suffer from the same methodological problems as the evaluations of Duluth and CBT approaches (such as low sample sizes and no information about how samples were derived). A further systematic review (Miller, Drake, & Nafziger, 2013), identified 11 rigorous evaluations of DV perpetrator intervention programs. Amongst these, six evaluated the impact of Duluth-based treatment models on re-offending and only one found a statistically significant effect on recidivism outcomes. Evaluations assessing the efficacy of other types of DV treatment programs showed reductions in re-offending but none were statistically significant, although all suffered from small sample sizes.

Despite the relatively large number of DV offender programs which have been delivered across Australia over the last decade (Mackay et al., 2015), very few rigorous outcome evaluations have been undertaken. For example, Day et al. (2010) evaluated the Gold Coast DV Integrated Response service, a Duluth-based DV offenders program. This evaluation analysed self-reported measures of offenders' abusive behaviour and attitudes before and after the program, as well as re-offending 12 months post-completion. However, the study group consisted of only 38 men, of which 35 completed the baseline survey and 12 completed the follow-up survey. They found moderate declines in abusive behaviour over time and better re-offending outcomes for completers compared to non-completers, however the sample size was too small for the authors to conclude whether or not the program was effective. Similarly, a South Australian study by Migliore, Ziersch, and Marshall (2014) examined re-offending outcomes of DV perpetrators who were the subject of protection orders and were subsequently referred to the DV Perpetrator Program. The authors compared offending rates of program completers with non-completers, as well as a group of offenders who received an order but were not referred. They found no statistically significant difference in re-offending after the order was issued when completers were compared with

non-completers, but some evidence for a slight improvement in offending outcomes when completers were compared to the no-treatment group. However, no attempts were made to account for other confounding factors such as offender demographics or prior offending.

One previous evaluation of particular relevance to the current study is the work by Blatch, O'Sullivan, Delaney, van Doorn, and Sweller (2016). This study examined general and violent recidivism outcomes for 953 offenders referred to CSNSW DAP between October 2007 and 30 June 2010. An intention-to-treat design was used where offenders referred to DAP were matched to non-referred offenders using propensity score matching. Offenders' time to a general and violent re-offence was analysed using Cox regression and the number of re-offences was analysed using Poisson regression. They found that DAP reduced the time to first general reconviction by 15 per cent and the time to a violent reconviction by 27 per cent. One limitation of this study is that re-offending was measured from the start date of the supervised order to maintain the same definition of follow-up periods between offenders, which meant that re-offences that occurred before participants were referred to the program were counted, potentially diluting the actual treatment effect. Additionally, no efforts were made to account for unobserved selection effects between completers, starters and non-starters in the regression models.

While the reviews above provide limited evidence for the effectiveness of DV perpetrator programs in reducing re-offending the authors consistently acknowledge the challenges involved in synthesizing findings from DV perpetrator intervention evaluations because of differences in application and implementation of the interventions, issues of attrition, selection into programs and methodological differences. The tendency to find no effect when more rigorous methods are applied underscores the need to carefully consider the design of quasi-experimental studies to ensure the techniques used are able to adequately address and minimise any bias arising from possible selection effects.

THE CURRENT STUDY

The aim of this study is to obtain an unbiased estimate of the causal impact of the community-based DVEQUIPS program on recidivism. As the community-based program is largely voluntary, it is possible that offenders who start the program and those who do not start the program differ systematically on unobservable factors related to re-offending risk (such as motivation to change). The potential presence of these factors poses a problem for evaluation as the differences in re-offending arising from these unobserved factors could wrongly be attributed to DVEQUIPS if no efforts are made to account for them. This study seeks to address these difficulties by employing an instrumental variables approach to estimate an unbiased treatment effect of the program.

METHOD

DATA

This study uses data from two sources: (1) the CSNSW Offender Information Management System (OIMS) and (2) the NSW Bureau of Crime Statistics and Research (BOCSAR) Re-Offending Database (ROD). OIMS includes data on all offenders who are currently or have previously been under the supervision of CSNSW, including details regarding treatment program referrals, program attendance rates and order types and lengths. ROD 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).

For this study, details for all offenders referred to the community-based DVEQUIPS in 2015 were extracted from OIMS.² Additional data were provided on all offenders referred to DAP between 2008 and 2014. As offenders could be referred more than once to DVEQUIPS, DAP or both between 2008 and 2015, each offender's earliest referral to DVEQUIPS was selected.³ Further, offenders may have attended multiple instances of the program's delivery as part of the same referral. In these cases only the first instance was selected. This resulted in a sample of 1,312 unique offenders referred to DVEQUIPS in 2015. This dataset was then linked to ROD to obtain demographic, offending and re-offending data. The ROD data used in this study contains all NSW court appearances finalised by 31 March 2017. The index appearance was identified by choosing the closest index appearance to the earliest sentence date on all active supervised orders an offender was on at the time of referral. A new offence was defined as any proven offence with an offence date occurring prior to 31 December 2016. This definition allows for a minimum of 12 months of follow-up time after the recorded referral date for all offenders in the sample and a three month lag for these new offences to be finalised in court.⁴

SAMPLE

The cohort of interest is all male offenders referred to DVEQUIPS in 2015. Out of the 1,312 offenders in this 'referral' dataset, 487 (37.1%) started the program within 12 months of referral (i.e. within the minimum follow-up period), where starting is defined as completing four or more sessions (the first module of the program). Fewer offenders (286; 21.8% of all referrals) completed the program, where completion is defined as attending 16 or more sessions. Due to the low rate of commencement in this dataset, the treatment group for this study is offenders who started DVEQUIPS within 12 months of referral. Those who were referred but did not start the program within 12 months comprise the control group in the analyses.

Offenders who started the program after 12 months from referral are excluded from the study sample (i.e. neither included in the treatment or control group). This is because offenders who started

the program after more than 12 months from their referral date would be 'treated' after their outcome has been measured. This leaves a final sample of 1,273 offenders for the analysis, 487 (38.3%) who started the program within 12 months of referral and 872 (61.7%) offenders who were referred but did not start (non-starters).

One limitation of this dataset is that it does not capture the results of the suitability assessments used when referring offenders to DVEQUIPS. These assessments form part of the pre-program suitability interview conducted by Community Corrections Officers. This means that some of the offenders in the sample who were identified as being eligible for DVEQUIPS may have been deemed unsuitable in the pre-program suitability interview. The extent to which this issue may affect our conclusions regarding program impact is explored further in the supplementary analyses presented in the Appendix.

VARIABLES

Re-offending outcome

Recidivism is the outcome examined and is defined in this study as any new proven offence committed after the program referral date.⁵ Usually, the finalisation date of the offender's index appearance is used as the point from which re-offending is measured, however this might introduce bias by counting any offences occurring prior to contact with the program. Another potential concern is that re-offending prior to program contact could ostensibly affect referral and participation, for example through offender availability or willingness to participate. Ideally, re-offending should be measured from the program start date to eliminate concerns about counting re-offences which occur prior to treatment.⁶ However, there is no comparable date for individuals in the control group (i.e. those who were referred but never started). Consequently, we use the referral date as the starting date for measuring re-offending as it is available for individuals in both the treatment and control groups.

For the analyses we measure recidivism using two binary variables:

- Re-offending within 12 months free time from referral: whether or not an individual committed a proven offence (excluding breaches of justice orders except for ADVOs) within 12 months excluding any time spent in custody;⁷ and
- Re-offending with a DV offence within 12 months free time from referral: whether or not an individual committed a proven DV-related offence (any offence flagged as DV-related excluding breaches of justice orders but including breaches of ADVOs) within 12 months excluding any time spent in custody.

Control variables

A wide range of variables that could potentially influence treatment allocation (referral to DVEQUIPS) and/or the likelihood

of re-offending were considered for inclusion in the instrumental variable model. These variables are detailed below.

Offender characteristics

- Age in years of the offender (categorised into groups of 18-24, 25-34, 35-44, and 45 and above);
- Indigenous status: whether the offender identified as being of Aboriginal or Torres Strait Islander descent at any court appearance recorded in ROD (categorised into yes or no);
- Socioeconomic disadvantage as measured by the ABS Socio-economic Index for Areas (SEIFA) for postcode of residence of the offender (ABS, 2011a) (categorised into quartiles, where quartile 1 is the most disadvantaged, and quartile 4 the least disadvantaged, and a category for unknown socioeconomic disadvantage);
- Remoteness for postcode of residence of the offender (categorised into major cities, inner regional, outer regional or remote or very remote, or unknown), obtained by applying the Accessibility Remoteness Index of Australia to the offender's postcode of residence (ABS, 2011b);
- Level of Service Inventory – Revised (LSI-R) score: most recent risk category (coded into low or medium-low, medium, medium-high or high, or unknown).

Characteristics of index finalisation

- Index finalisation date;
- Number of proven concurrent charges (continuous from 0-6+);
- Any proven offence of different types, by ANZSOC code (coded into yes or no);
- Penalties received at index finalisation (coded into yes or no).

Prior criminal history

- Total number of previous court appearances (coded into 0-4, 5-9 and 10 or more prior appearances);
- Proven offences in the five years prior to index finalisation (variables for all offence categories, coded into yes or no).

Characteristics of offenders' active supervised orders (obtained from OIMS)

- The time remaining on offenders' supervised orders, defined as the time from an offenders' referral date to the last order expiry date (coded into three categories: fewer than 20 weeks, between 20 and 40 weeks and more than 40 weeks);
- Whether offenders were on a parole order at time of referral;
- Location of referral (i.e. Community Corrections Office).

STATISTICAL ANALYSIS

The 'treatment effect' of DVEQUIPS could be estimated by comparing the rates of re-offending for offenders who started the program and those who did not, controlling for any differences

in observed risk factors. However, given that the program was voluntary, the offenders who started the program could differ from the offenders who did not start the program in ways that cannot be observed in this data. For example, offenders who started could be more motivated to change their behaviour than those who did not, which would also make them less likely to re-offend regardless of program participation.

An instrumental variables approach is employed to deal with this potential problem of omitted variable bias. In the present context this approach (detailed in Angrist, Imbens, & Rubin, 1996) requires us to find an instrumental variable (known as an instrument or IV) that influences whether an individual gets access to treatment but which has no direct effect on re-offending other than through its effect on access to treatment. Essentially, the use of an IV allows us to separate the effect of treatment on offending from the effect of other factors on offending that are correlated with treatment participation (e.g. offender motivation).

The ability to obtain an unbiased estimate of the treatment effect using this approach hinges on the IV's validity. A valid IV is one that is strongly correlated with treatment allocation but is not related to the outcome except through its impact on the treatment variable. The proposed IV for this study reflects the DVEQUIPS commencement rate at the same time and location at which an individual offender is referred to the program (which we refer to as the peer non-commencement rate). That is, at the community corrections office where an individual offender is referred, we calculate the proportion of all referrals to the DVEQUIPS program at this location that had not started the program by the date the individual was referred. The IV is represented by equation (1) below. Essentially, the IV is calculated for each offender i at the point of their referral at time t by taking the number of their peers (those previously referred from the same location j) and calculating the proportion among them who had yet to start the program at time t . This measures the availability of other offenders to start the program at the time of an offender's referral to the program.⁸

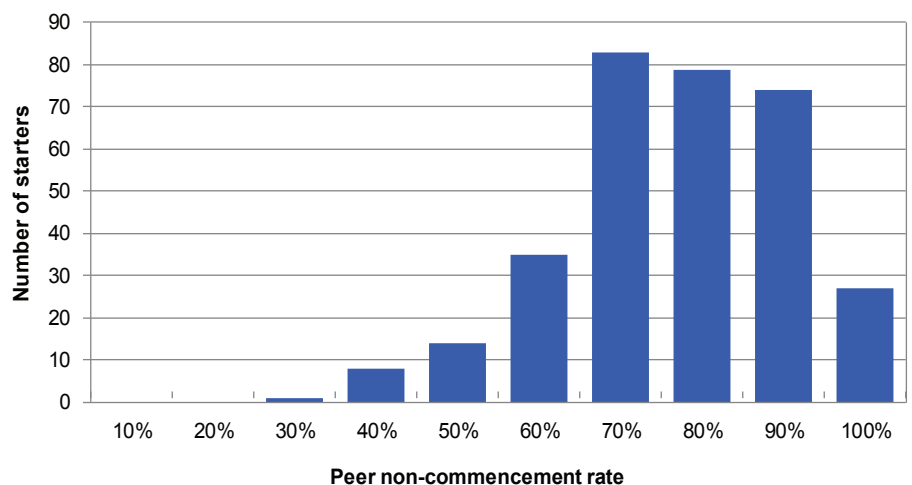
$$(1) \quad IV_{ijt} = \frac{\sum \text{Not started}_{jt}}{\sum \text{Referred}_{jt}}$$

This IV was chosen because DVEQUIPS is delivered in a closed group-based setting, with specific modules delivered in sequence over 20 weeks. This means that offenders cannot commence the program as soon as they are referred but must wait for a new group to begin. Also, with a limited number of DVEQUIPS facilitators at each location, a new group will not commence until a minimum number of referrals are received. An offender's chance of commencing the program soon after referral will therefore depend on there being a sufficient number

of his peers also available to start a program at that location at the same time.⁹ We primarily test the relevance of this IV in the estimated models using a partial F -test (detailed in Bound, Jaeger, & Baker, 1995).¹⁰ If the test statistic exceeds the recommended value of 10, the IV is considered relevant (i.e. a strong instrument). Some evidence for the relevance of the IV is also shown in Figure 1 which illustrates the relationship between the peer non-commencement rate (categorised by rounding to the nearest 10%) and the number of offenders who start DVEQUIPS. As expected, there is a strong positive relationship between the IV and starting DVEQUIPS. However, the partial F -statistic is a better measure of the IV's relevance once adjusted for the model covariates, as it is able to measure the ability of the IV to predict treatment after adjusting for other factors affecting participation (i.e. whether the IV is able to predict participation conditional on all relevant observed factors).

While the relevance assumption can be easily tested with a partial F -test, the exclusion restriction (the assumption that there is no relationship between the IV and the outcome except through the treatment variable) cannot be tested directly. This assumption is generally considered satisfied if there are no direct and obvious relationships between the variable and re-offending outside of its effect on starting. A threat to the exclusion restriction in this case would be the existence of any factor specific to a location which influences both the accumulation of DVEQUIPS referrals over time and is related to an individual's propensity to reoffend. This is plausible for several reasons: (1) A community corrections office with a higher rate of non-starters could also have officers who are less engaged with offenders and who have inferior supervisory skills; (2) variables which could affect rates of re-offending whilst under supervision; (3) there may be differences in the fidelity of program delivery between locations which could affect its effectiveness at reducing re-offending; and, (4) the volume of offenders under supervision and referrals at particular locations may affect the capacity of community corrections officers to engage with offenders. To account for these confounding effects,

Figure 1. Number of DVEQUIPS starters, by peer non-commencement rate (proposed IV)



location fixed effects (which would be relatively stable over the time period examined) were also included in the IV models.

The models were estimated as follows. First, a two-stage least-squares (2SLS) estimation method was used for the instrumental variables analysis. The IV was included in a first stage linear regression (the treatment equation) to obtain the predicted probability of starting the program for those in the referral group. These predicted probabilities were then used in place of starting DVEQUIPS in a second-stage linear regression predicting re-offending (the outcome equation). The 2SLS approach enables the use of test statistics for the relevance of the IV (partial *F*-test) as well as endogeneity (*C*-statistic), which is a diagnostic for determining whether selection bias is a problem in this evaluation.

As the treatment variable (starting DVEQUIPS) and the outcome variable (whether or not an offender re-offended) are both binary, using 2SLS in this context means that both regressions were effectively linear probability models, which do not constrain the predicted values of the models between 0 and 1. Therefore, a recursive bivariate probit approach suggested by Heckman (1978) was also undertaken. This approach involves the estimation of the treatment and outcome equations as simultaneous probit models (which recognise starting DVEQUIPS and the re-offending outcomes as variables bounded by zero and one) with jointly distributed errors to account for selection effects. The correlation parameter estimated in these models indicates the extent to which the equations are related and unobserved selection effects are present. Both the 2SLS and probit models were estimated

using clustered standard errors for the community correction, offices from which offenders were referred to take into account any possible correlation in re-offending rates within sites.

Endogeneity and model assessment

Tests for the presence of endogeneity were also undertaken to inform the choice between IV and ordinary logistic regression methods. For the 2SLS case, endogeneity was assessed using the difference-in-Sargan *C*-statistic (Baum, Schaffer and Stillman 2003), which tests whether the difference in estimates produced by the 2SLS and a linear regression (where the endogenous variable is treated as exogenous) outweighs the loss of efficiency from using 2SLS. In the case of the recursive bivariate probit models, chi-square tests of the correlation parameter *p* are used to assess for endogeneity (i.e. whether the error terms of the two equations are significantly correlated).

RESULTS

DESCRIPTIVE STATISTICS

Table 1 shows the demographic, index offence and prior offending characteristics of the DVEQUIPS starter and non-starter groups. There were significant differences in the demographic characteristics (age, socioeconomic disadvantage and remoteness area) and LSI-R risk category of offenders in the two groups. Those who started DVEQUIPS had fewer concurrent offences at their index contact than non-starters. More DVEQUIPS starters had received community service orders at

Table 1. Demographic, index offence and prior offending characteristics, DVEQUIPS starters (within 12 months of referral) vs. non-starters

Variable	Did not start DVEQUIPS within 12 months (n=786)	Started DVEQUIPS within 12 months (n=487)	<i>p</i>
Age (categories)			.004
18-24 years	176 (22.4%)	80 (16.4%)	
25-34 years	293 (37.3%)	188 (38.6%)	
35-44 years	240 (30.5%)	144 (29.6%)	
45 years and above	77 (9.8%)	75 (15.4%)	
Indigenous status			.295
Non-Indigenous/unkown	474 (60.3%)	308 (63.2%)	
Indigenous	312 (39.7%)	179 (36.8%)	
Socioeconomic disadvantage (SEIFA 2011) for defendant's postcode			.005
1 (most disadvantaged)	251 (31.9%)	141 (29.0%)	
2	264 (33.6%)	170 (34.9%)	
3	144 (18.3%)	123 (25.3%)	
4 (least disadvantaged)	55 (7.0%)	28 (5.7%)	
5 (unknown)	72 (9.2%)	25 (5.1%)	
Remoteness area of defendant's postcode			<.001
Metropolitan	392 (49.9%)	271 (55.6%)	
Inner regional	213 (27.1%)	153 (31.4%)	
Outer regional, remote or very remote	110 (14.0%)	38 (7.8%)	
Unknown	71 (9.0%)	25 (5.1%)	

Table 1. Demographic, index offence and prior offending characteristics, DVEQUIPS starters(continued)

Variable	Did not start DVEQUIPS within 12 months (n=786)	Started DVEQUIPS within 12 months (n=487)	p
LSI-R risk category			.002
Low to medium-low	95 (12.1%)	33 (6.8%)	
Medium	445 (56.6%)	304 (62.4%)	
Medium-high to high	207 (26.3%)	113 (23.2%)	
Unknown	39 (5.0%)	37 (7.6%)	
Number of proven concurrent charges at index contact (including principal offence), mean (standard deviation)	4.05 (3.23)	3.62 (2.73)	.014
Offences at index (any concurrent offence)			
Acts intended to cause injury at index contact	543 (69.1%)	358 (73.5%)	.091
Assault	449 (57.1%)	298 (61.2%)	.152
Serious assault	210 (26.7%)	133 (27.3%)	.817
Abduction, harassment and other offences against the person	52 (6.6%)	30 (6.2%)	.748
Theft	58 (7.4%)	26 (5.3%)	.154
Drug offences	56 (7.1%)	35 (7.2%)	.967
Property damage	209 (26.6%)	115 (23.6%)	.236
Offences against justice procedures	525 (66.8%)	318 (65.3%)	.583
Breach of custodial order	55 (7.0%)	36 (7.4%)	.790
Breach of community order	253 (32.2%)	160 (32.9%)	.805
Breach of ADVO	351 (44.7%)	211 (43.3%)	.642
DV-related assault	414 (52.7%)	267 (54.8%)	.454
Penalties received at index contact (any)			
Intensive correction order	36 (4.6%)	32 (6.6%)	.125
Supervised suspended sentence	158 (20.1%)	102 (20.9%)	.717
Community service order	54 (6.9%)	68 (14.0%)	<.001
Supervised bond	448 (57.0%)	275 (56.5%)	.853
Time remaining on supervised order (i.e. to expiry of last supervised order)			.006
Fewer than 20 weeks	97 (12.7%)	37 (7.7%)	
20 to 39 weeks	224 (29.3%)	127 (26.6%)	
40 weeks or more	444 (58.0%)	314 (65.7%)	
Number of finalised court appearances (with proven offence/s) as a juvenile or adult, mean (SD)	3.07 (2.34)	2.74 (2.31)	.015
Penalties received in the five years prior to finalisation (any)			
Prison	255 (32.4%)	119 (24.4%)	.002
Intensive correction order	17 (2.2%)	14 (2.9%)	.423
Supervised suspended sentence	132 (16.8%)	69 (14.2%)	.212
Community service order	114 (14.5%)	75 (15.4%)	.662
Bond	512 (65.1%)	300 (61.6%)	.202
Supervised bond	321 (40.8%)	187 (38.4%)	.387
Proven offences in the five years prior to finalisation (any)			
Acts intended to cause injury	499 (63.5%)	292 (60.0%)	.207
Assault	436 (55.5%)	257 (52.8%)	.347
Serious assault	211 (26.8%)	128 (26.3%)	.826
Theft	180 (22.9%)	71 (14.6%)	<.001
Drug offence	196 (24.9%)	93 (19.1%)	.016
Property damage offence	281 (35.8%)	163 (33.5%)	.407
Breach of custodial order	79 (10.1%)	45 (9.2%)	.635
Breach of community order	298 (37.9%)	145 (29.8%)	.003
Violent offence	508 (64.6%)	296 (60.8%)	.166
Property offence	220 (28.0%)	96 (19.7%)	<.001
DV offence	483 (61.5%)	300 (61.6%)	.957
DV-related assault	329 (41.9%)	200 (41.1%)	.781

Table 2. Unadjusted re-offending rates, DVEQUIPS starters and non-starters

Outcome variable	Did not start DVEQUIPS (n=786)	Started DVEQUIPS (n=487)	p
Re-offending within 12 months free time	371 (48.8%)	206 (42.9%)	.043
Re-offending with a DV offence within 12 months free time	184 (24.2%)	126 (26.3%)	.419

the index court finalisation than non-starters. Predictably, starters also tended to have more time remaining on their supervised orders than non-starters at the time of referral. There are also notable differences in the prior offending history of those who start DVEQUIPS and those who do not. Those who start DVEQUIPS are less likely to have a prison sentence, theft, drug, property and breach of community order offences proven in the five years prior to their index appearance. DVEQUIPS starters also have fewer prior court appearances with at least one proven offence compared to their counterparts who do not start.

Table 2 presents general and DV-related re-offending rates within 12 months of free time for offenders who started DVEQUIPS and non-starters. Fewer offenders who started DVEQUIPS re-offended with any offence within 12 months free time compared to non-starters. The observed re-offending rate among those who started DVEQUIPS was approximately six percentage points lower than among non-starters. Observed, unadjusted rates of DV re-offending among those who started DVEQUIPS, and those who did not start DVEQUIPS were not significantly different. Of course, these comparisons do not take into account any systematic differences between the groups on observed or unobserved variables.

INSTRUMENTAL VARIABLES ANALYSIS

Re-offending with a general offence

Table 3 reports the results of the logistic regression (Column A), 2SLS (Column B) and bivariate probit (Column C) models estimating the effect of DVEQUIPS on rates of general re-offending. The top half of the table presents the results of the second-stage regression (predicting re-offending) and the associated measures of model fit. The bottom half of the table presents the results of the first-stage regressions (predicting starting DVEQUIPS within 12 months of referral) for the 2SLS and bivariate probit models, along with the tests of endogeneity and the partial F-test of the IV in the 2SLS estimation.

We start with the 2SLS specification of the model (Column B). As expected, the IV has a positive effect on starting the program. Specifically, a one percentage point increase in the peer non-commencement rate leads to a 0.52 percentage point increase in the probability that an offender will start DVEQUIPS. This is consistent with our expectations that an offender is more likely

to start the program if there are more offenders at that location who have been referred but are yet to start the program. The partial F-statistic of 29.37 exceeds the rule of thumb of 10 and confirms that it is a relevant IV. However, the difference-in-Sargan C-statistic is not statistically significant (1.38, $p=.240$), which indicates that unobserved factors do not significantly change estimates obtained through single-equation regressions. The treatment effect estimated through this method is negative but not statistically significant. This indicates that there is no evidence of a significant difference in 12 months re-offending likelihood for offenders who started DVEQUIPS compared with non-starters.

The bivariate probit specification of the model (Column C) provides a valid alternative for estimating treatment effects while accounting for unobserved differences. As is the case with the 2SLS model in Column B, a one percentage point increase in the peer non-commencement rate positively impacts an offender’s probability of starting the program. However the test of endogeneity (the significance of the correlation of residuals, p) was not significant, indicating that unobserved selection effects that affect both an offender’s likelihood of starting DVEQUIPS and re-offending are not present. Again, the coefficient for starting DVEQUIPS estimated by this model is negative but not statistically significant from zero, which suggests no difference between the treatment and control groups in the probability of re-offending within 12 months free time.

The lack of evidence for systematic endogeneity found in the 2SLS and bivariate probit models indicates that the logistic regression estimates (which are more efficient and produce smaller standard errors) are reliable. The results of these models are presented in Column A. Encouragingly, the direction of the coefficient is in the same direction as the 2SLS and bivariate probit models. However, despite the use of a more efficient estimation method, this result is not significantly different from zero at the .05 level. We checked these results for robustness to two potential sources of bias: first, whether differences in eligibility arose between the groups from some offenders not having an IPV offence, and second, whether completion of other EQUIPS programs which affected starting DVEQUIPS differed across groups. In neither case did these results change. Therefore, we consider these results relatively robust to the exclusion of potentially unsuitable offenders who may have been referred despite having no IPV offence and the exclusion of offenders who completed other treatment programs. Note however that there may be other reasons why offenders could be excluded (e.g. mental health or substance use issues) which we cannot observe. Therefore, the robustness of our results to these other sources of unobserved bias is unknown.

The control variables included in the models above encompass a number of factors, some of which remained significant independent predictors of starting DVEQUIPS and re-offending. Older offenders were significantly more likely to start the program, and also less likely to re-offend within 12 months free time. Offenders in the highest LSI-R categorisation (medium-high to high) were more likely to re-offend than those in the referent

Table 3. Logistic regression, two-stage least-squares (2SLS) and bivariate probit estimations of re-offending with a general offence within 12 months free time from referral to DVEQUIPS

Dependent variable: Re-offending within 12 months free time	Column A: Logistic regression		Column B: 2SLS		Column C: Bivariate probit	
	Coef.	(St. err.)	Coef.	(St. err.)	Coef.	(St. err.)
Started DVEQUIPS	-0.20	(0.14)	-0.30	(0.22)	-0.49	(0.36)
Age (relative to 18-24 years)						
25-34 years	-0.17	(0.18)	-0.04	(0.04)	-0.12	(0.12)
35-44 years	-0.50 *	(0.24)	-0.11 *	(0.05)	-0.32 *	(0.14)
45 years and above	-1.01 ***	(0.28)	-0.20 **	(0.06)	-0.61 **	(0.19)
LSI-R risk category (relative to low to medium low)						
Medium	0.44	(0.27)	0.13 *	(0.07)	0.35	(0.19)
Medium-high to high	0.66 *	(0.28)	0.18 **	(0.07)	0.47 *	(0.19)
Unknown	0.44	(0.40)	0.16	(0.11)	0.37	(0.27)
Received a community service order at index appearance	0.09	(0.23)	0.06	(0.05)	0.13	(0.15)
Number of prior court appearances with at least one proven offence (relative to 0-4 prior appearances)						
5 to 9 prior appearances	0.27	(0.16)	0.07 *	(0.03)	0.20 *	(0.10)
10 or more prior appearances	0.69 ***	(0.15)	0.17 ***	(0.03)	0.47 ***	(0.10)
Received a prison sentence in the five years prior to index contact	0.41 *	(0.17)	0.06	(0.04)	0.19	(0.12)
Proven theft offence in the five years prior to index contact	0.27	(0.19)	0.03	(0.05)	0.12	(0.14)
Time remaining on supervised order at time of referral (relative to fewer than 20 weeks)						
20 to 39 weeks		(0.22)	-0.01	(0.06)	-0.04	(0.14)
40 weeks or more		(0.25)	0.01	(0.07)	0.01	(0.16)
Constant	-0.90 **	(0.33)	0.28 ***	(0.07)	-0.59 **	(0.20)
Location fixed effects	YES		YES		YES	
N	1192		1151		1151	
AIC	1513.01		1592.63		2780.49	
Dependent variable: Starting DVEQUIPS within 12 months of referral	Coef.	(St. err.)	Coef.	(St. err.)	Coef.	(St. err.)
IV: Peer non-commencement rate			0.52 ***	(0.10)	1.55 ***	(0.29)
Age (relative to 18-24 years)						
25-34 years			0.03	(0.03)	0.03	(0.11)
35-44 years			0.03	(0.04)	0.06	(0.13)
45 years and above			0.11 **	(0.04)	0.31 *	(0.12)
LSI-R risk category						
Medium			0.14 *	(0.05)	0.49 **	(0.19)
Medium-high to high			0.13 *	(0.06)	0.46 *	(0.21)
Unknown			0.22 **	(0.08)	0.73 **	(0.26)
Received a community service order at index appearance			0.14 **	(0.04)	0.40 **	(0.13)
Number of prior court appearances with at least one proven offence (relative to 0-4 prior appearances)						
5 to 9 prior appearances			-0.03	(0.05)	-0.10	(0.14)
10 or more prior appearances			0.04	(0.04)	0.13	(0.11)
Received a prison sentence in the five years prior to index contact			-0.08 *	(0.04)	-0.29 *	(0.12)
Proven theft offence in the five years prior to index contact			-0.12 *	(0.05)	-0.39 **	(0.15)
Time remaining on supervised order at time of referral (relative to fewer than 20 weeks)						
20 to 39 weeks			0.08	(0.05)	0.26	(0.16)
40 weeks or more			0.12 *	(0.05)	0.37 *	(0.18)
Constant			-0.51 ***	(0.11)	-3.09 ***	(0.35)
Partial F-statistic			29.37 ***			
C-statistic (Endogeneity test)			1.38	(p=0.240)		
Correlation parameter					0.23	0.22

* p<.05, ** p<.01, *** p<.001
Coef. = coefficient; Std. err. = standard error.

category of low to medium-low. Offenders who had a prior prison sentence and offenders who had more than 10 previous court appearances with at least one proven offence were also more likely to re-offend.

Re-offending with a DV offence

The results of the models estimating the treatment effect of DVEQUIPS with DV re-offending as the second-stage outcome variable are presented in Table 4. As with the estimates for general re-offending, the top half of the table details the results of the second-stage equation (likelihood of DV re-offending) and measures of model fit, and the bottom half presents the first-stage estimates (likelihood of starting DVEQUIPS) and the measures of instrument validity and endogeneity.

Column B presents the results of 2SLS estimations of the treatment effect of DVEQUIPS on DV re-offending. Looking at the first coefficient in the second column of the lower half of the table, the IV has a positive and significant effect on starting the program (i.e. a one percentage point increase in the peer non-commencement rate increases the likelihood of an offender starting DVEQUIPS within 12 months of referral), and the strength of the IV is confirmed by the partial *F*-statistic of 29.37. As is the case with general re-offending, the difference-in-Sargan *C*-statistic is not statistically significant (0.19, $p=.663$). This indicates that there is little evidence for preferring instrumental variables approaches to the logistic regression model. The estimated treatment effect in this model is small but not significantly different to zero.

Further evidence for whether endogeneity is an issue for DV re-offending can be gathered from the estimates of the bivariate probit model in Column C. The strength of the IV is again demonstrated through a statistically significant positive relationship with starting the program. Given the appropriateness of the IV and the satisfaction of the model's assumptions, we turn to the significance of the correlation parameter, for evidence for endogeneity. The lack of significance corroborates that we should not be concerned about unobserved selection effects in estimating the treatment effect. Consistent with the 2SLS estimation, the treatment effect is again small and not statistically significant.

The lack of evidence for systematic endogeneity found in the 2SLS and bivariate probit models indicates that the logistic regression estimates can be relied on but once again, no evidence for a treatment effect is evident after accounting for other relevant covariates. Robustness checks (see the Appendix) where offenders without a known IPV offence at index contact are excluded do not yield significantly different results. Age is a significant factor in predicting DV re-offending, with offenders aged 45 and above less likely to commit a new DV offence relative to young offenders. Receiving a community service order and receiving a prison sentence in the five years prior to the index contact were both associated with a higher probability of DV re-offending.

Starters within six months of referral

As previously mentioned, offenders with a long lag between referral and starting the program may have re-offended prior to starting the program and bias could be introduced by counting these re-offences in estimating the treatment effect of DVEQUIPS. Therefore, the same models were re-estimated excluding 63 offenders who started the program more than six months after referral. A summary of the results of these re-estimations are presented in Table 5 below.

Starting with re-offending within 12 months free time among those who started within six months, we observe a strong effect of the program on reducing general re-offending within 12 months free time estimated through IV methods, with a strong partial *F*-statistic shown in the 2SLS case. However, the effect estimated in both the 2SLS and bivariate probit models is not statistically significant, and no evidence for systematic endogeneity is found. The coefficient for the treatment effect is also negative in the logistic regression models but is not statistically significant.

For DV-related re-offending within 12 months free time, both IV methods find no significant effect of starting the program on reducing re-offending. As in all the other estimations, endogeneity is not a problem, and therefore the more efficient logistic regression estimates might be used. These indicate a positive effect of starting the program on DV-related re-offending but this is not significant at the 0.05 level.

DISCUSSION

The aim of this study was to obtain an unbiased estimate of the causal impact of the community-based DVEQUIPS program on general and DV re-offending rates. Two-stage least squares (2SLS) and bivariate probit models were initially used to estimate the effect of the program in the potential presence of omitted variable bias. The IV used (i.e. the peer non-commencement rate) was relatively strong but no evidence for endogeneity was found. Logistic regression was therefore preferred for estimating the treatment effect. The results of the logistic regression models show no significant treatment effect for those who start DVEQUIPS for both general re-offending within 12 months free time from referral and re-offending with a DV offence within 12 months free time. Restricting the treatment group in the analyses to only those offenders who started the program within six months does not alter this result.

These findings are inconsistent with those reported in the earlier evaluation of the Domestic Abuse Program (DAP), the predecessor to DVEQUIPS (Blatch et al. 2016). One potential reason for this discrepancy is the difference in participant profile and completion rates of the two programs. In the DAP study, 42 per cent of offenders referred were assessed as at a low or medium-low risk of re-offending (according to the LSI-R) compared with just 11 per cent of offenders in the DVEQUIPS referral sample; meanwhile 81 per cent of persons referred to

Table 4. Logistic regression, two-stage least-squares (2SLS) and bivariate probit estimations of re-offending with a DV-related offence within 12 months of referral to DVEQUIPS

Dependent variable: DV re-offending within 12 months free time	Column A: Logistic regression		Column B: 2SLS		Column C: Bivariate probit	
	Coef.	(St. err.)	Coef.	(St. err.)	Coef.	(St. err.)
Started DVEQUIPS	0.12	(0.17)	0.09	(0.16)	0.09	(0.45)
Age (relative to 18-24)						
25-34	-0.21	(0.17)	-0.05	(0.03)	-0.15	(0.12)
35-44	-0.15	(0.20)	-0.03	(0.04)	-0.10	(0.12)
45 and above	-0.61 *	(0.28)	-0.11 *	(0.05)	-0.39 *	(0.18)
LSI-R risk category (relative to low to medium-low)						
Medium	0.34	(0.31)	0.05	(0.05)	0.25	(0.20)
Medium-high	0.44	(0.32)	0.06	(0.05)	0.31	(0.19)
Unknown	0.57	(0.45)	0.10	(0.09)	0.43	(0.29)
Received a community service order at index appearance	0.49 *	(0.22)	0.09	(0.05)	0.32 *	(0.15)
Number of prior court appearances with at least one proven offence (relative to 0-4 prior appearances)						
5 to 9 prior appearances	-0.03	(0.19)	0.01	(0.03)	0.03	(0.11)
10 or more prior appearances	0.33	(0.19)	0.06	(0.04)	0.23	(0.12)
Received a prison sentence in the five years prior to index contact	0.37 *	(0.17)	0.08 *	(0.03)	0.24 *	(0.11)
Proven theft offence in the five years prior to index contact	0.02	(0.18)	0.01	(0.04)	-0.01	(0.12)
Time remaining on supervised order at time of referral (relative to fewer than 20 weeks)						
20 to 39 weeks	-0.29	(0.28)	-0.06	(0.05)	-0.19	(0.16)
40 weeks or more	0.04	(0.30)	-0.01	(0.05)	-0.02	(0.16)
Constant	-1.72 ***	(0.35)	0.14 ***	(0.05)	-1.14 ***	(0.21)
Location fixed-effects	YES		YES		YES	
N	1164		1151		1151	
AIC	1248.63		1241.33		2531.14	
Dependent variable: Starting DVEQUIPS within 12 months of referral	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.
Peer non-commencement rate			0.52 ***	(0.10)	1.55 ***	(0.30)
Age (relative to 18-24)						
25-34			0.03	(0.03)	0.04	(0.10)
35-44			0.03	(0.04)	0.08	(0.13)
45 and above			0.11 **	(0.04)	0.32 **	(0.12)
LSI-R risk category (relative to low to medium-low)						
Medium			0.14 *	(0.05)	0.48 **	(0.19)
Medium-high			0.13 *	(0.06)	0.45 *	(0.21)
Unknown			0.22 **	(0.08)	0.72 **	(0.26)
Received a community service order at index appearance			0.14 **	(0.04)	0.39 **	(0.13)
Number of prior court appearances with at least one proven offence (relative to 0-4 prior appearances)						
5 to 9 prior appearances			-0.03	(0.05)	-0.10	(0.14)
10 or more prior appearances			0.04	(0.04)	0.12	(0.11)
Received a prison sentence in the five years prior to index contact			-0.08 *	(0.04)	-0.29 *	(0.12)
Proven theft offence in the five years prior to index contact			-0.12 *	(0.05)	-0.40 **	(0.15)
Time remaining on supervised order at time of referral (relative to fewer than 20 weeks)						
20 to 39 weeks			0.08	(0.05)	0.27	(0.16)
40 weeks or more			0.12 *	(0.05)	0.37 *	(0.18)
Constant			-0.51	(0.11)	-3.09 ***	(0.36)
Partial F-statistic			29.37			
C-statistic (Endogeneity test)			0.19	(p=.663)		
Correlation parameter					-0.004	(0.29)

* p<.05, **p<.01, ***p<.001
Coef.= coefficient; Std. err.= standard error.

Table 5. Summary of models predicting impact of starting DVEQUIPS on general and DV re-offending for offenders starting within six months of referral

Outcome Model	Re-offending within 12 months free time			DV re-offending within 12 months free time		
	Logistic	2SLS	Bivariate probit	Logistic	2SLS	Bivariate probit
Effect of starting DVEQUIPS (Standard error)	-0.16 (0.14)	-0.37 (0.22)	-0.65 (0.37)	0.14 (0.19)	-0.03 (0.14)	-0.18 (0.37)
Location fixed effects	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
N	1171	1127	1127	1144	1127	1127
AIC	1483.52	1602.54	2729.81	1232.22	1213.98	2482.78
Partial F-statistic		35.09			35.09	
C-statistic (Endogeneity test)		1.28 (p=0.259)			0.19 (p=0.664)	
Correlation parameter (Standard error)			0.12 (0.57)			-0.17 (0.54)

DAP commenced the program compared with only one-third of referrals to DVEQUIPS. Treatment effectiveness may vary depending upon the level of risk of the participants targeted by program and/or the overall dosage received.

A further possibility is that the use of an intention-to-treat design comparing referrals and non-referrals (adjusting for observed characteristics of offenders) as was adopted in the DAP evaluation is not sufficient to deal with problems of omitted variable bias. Referral to either DAP or DVEQUIPS does not appear to be systematic since the number of offenders referred to these programs is always much lower than the number of DV offenders under community supervision. If referrals rely to some extent on the discretion of the community corrections officer supervising (or assessing) the offender than it would be reasonable to expect that referrals and non-referrals would likely differ not only on observables but also unobservables, in particular the willingness of the perpetrator to engage with treatment and/or authorities. Using methods that do not adequately address these issues would bias the results of the analysis in favour of a treatment effect.

While our results indicate no difference in the likelihood of re-offending between referrals that start DVEQUIPS and referrals who do not commence the program, it is important to note that the counterfactual in this case are offenders who remained under the supervision of community corrections. What this finding suggests then is that DVEQUIPS participation does not confer any additional benefit over and above the effect of supervision and engagement with community corrections (see Wan, Poynton, van Doorn & Weatherburn 2014) for further discussion of the effectiveness of supervision). This result is not surprising given the limited evidence base for DV perpetrator rehabilitation programs as found by prior reviews.

One important caveat to our results is that re-offending was measured from the referral date and offenders referred to the DVEQUIPS program appear to experience significant delays in commencing the program. Measuring our outcome prior to

the commencement of any treatment was necessary given the absence of a comparable date for offenders in our control group who did not start the program. This would, however, dilute any treatment effect of the program (if one exists) and reduce our ability to detect a significant effect if one exists. Further analyses restricting the sample to only those who commenced within six months of referral also was unable to detect a significant effect of DVEQUIPS on re-offending and amongst this sample only a very small proportion of offenders (5%) reoffended before commencing the program.

The inclusion in the treatment group of people who did not complete treatment and the low treatment completion rate raises the question of whether the effects of treatment are being obscured by treatment ‘dilution’. It is possible that rates of re-offending were reduced among the minority (21.7%) of the referral group who completed treatment and that this effect was obscured by those who did not complete treatment. There are several points, however, that need to be borne in mind when considering this possibility. The first is that removing program non-completers from the analysis would have violated the intention-to-treat principle and left the evaluation open to selection bias. The second is that the benefits of treatment may be limited to those who complete it but the costs are not. Participation rates of the program are therefore an important consideration when determining the overall value of a program as they reflect the broader realities of program delivery. Thirdly, the fact that no difference in re-offending was observed in the treatment group despite most offenders (55.9%) having completed 16 two-hour treatment sessions suggests that if there is a treatment effect, it is likely to be comparatively small.

Further research using more robust experimental approaches should be undertaken to confirm the results of the analysis presented here. What is clear from our work, however, is the need for a better understanding of the referral process for community corrections programs, including the reasons why offenders are assessed as ineligible or suitable for programs and any factors that contribute to delays in commencing therapeutic programs

after successful referral. Better coverage and more efficient delivery of therapeutic services for DV offenders under community supervision has the potential to enhance the effectiveness of community-based programs and achieve improved outcomes for victims and their families.

NOTES

1. Note that 'perpetrators' is used here to distinguish between those who are accused of DV and those who have been proven to have committed DV offences.
2. There is insufficient follow-up time to observe the re-offending rates of offenders in custody.
3. For offenders who were referred multiple times and did not start the program on any referral, the earliest referral to the program was chosen. For offenders with multiple referrals who started the program multiple times, the earliest referral where they started the program was selected. Offenders who had previously been referred to DAP who started the program and were then referred to DVEQUIPS in 2015 were excluded from the treatment group.
4. This may be affected by court delays; however we expect this only to be a major problem relating to the NSW District Criminal Court (and therefore only to affect indictable or strictly indictable matters).
5. Repeat victimisation may be a better measure of offender behavioural change given knowledge of underreporting of DV incidents, however we did not have access to this type of data for this study.
6. Ninety-three (23.6%) offenders in the treatment group re-offended before the start date.
7. 33 offenders did not have 12 months free time after referral and did not re-offend within that free time and are treated as missing.
8. This was calculated prior to the sample selection criteria being set to ensure it accurately reflected referrals and non-commencements at each location.
9. Some offenders are referred to the program from one location and participate in the program at another location.
10. In the presence of clustered standard errors, which are employed in this study, the Kleibergen and Paap (2006) identification tests are often preferred to the partial F-test. However, for simplicity we report the partial F-tests for all the 2SLS estimations in this report as they are not contradicted by the Kleibergen-Paap tests.

REFERENCES

- Australian Bureau of Statistics. (2011a). Australian Standard Geographical Classification (ASGC) (cat. no. 1216.0). Retrieved 18 Jan. 2018 from <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1216.0July%202011?OpenDocument>
- Australian Bureau of Statistics. (2011b). Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia 2011. Retrieved 18 Jan. 2018 from <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012011?OpenDocument>
- Australian Bureau of Statistics. (2017). *Personal Safety, Australia, 2016*. (cat. no. 4906.0). Retrieved 11 Nov 2017 at: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4906.0Main+Features12012?OpenDocument>
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables, *Journal of the American Statistical Association*, 91(434).
- Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, 3(1), pp. 1–31. Retrieved 21 Nov 2017 at: <http://www.stata-journal.com/sjpdf.html?articlenum=st0030>
- Blatch, C., O'Sullivan, K., Delaney, J. J., van Doorn, G., & Sweller, T. (2016). Evaluation of an Australian domestic abuse program for offending males. *Journal of Aggression, Conflict and Peace Research*, 8. doi:10.1108/JACPR-10-2015-0194.
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association*, 90(430), p. 443.
- Corrective Services NSW. (2016). *Compendium of Offender Behaviour Change Programs in New South Wales*. Retrieved 17 Jan 2018 at: <http://www.correctiveservices.justice.nsw.gov.au/Documents/programs/CSNSW%20Compendium%20of%20Offender%20Behaviour%20Change%20Program.PDF>
- Day, A., Chung, D., O'Leary, P., & Carson, E. (2009). Programs for Men who Perpetrate DV: An Examination of the Issues Underlying the Effectiveness of Intervention Programs. *Journal of Family Violence*, 24(3), 203–212. doi: 10.1007/s10896-008-9221-4.
- Day, A., Chung, D., O'Leary, P., Justo, D., Moore, S., Carson, E., & Gerace, A. (2010). *Integrated responses to DV: Legally mandated intervention programs for male perpetrators*. Trends and Issues in Crime and Criminal Justice, No. 404. Canberra: Australian Institute of Criminology. Retrieved 17 Jan 2018 at: http://www.aic.gov.au/media_library/publications/tandi_pdf/tandi404.pdf
- Eckhardt, C. I., Murphy, C., Black, D., & Suhr, L. (2006). Intervention Programs for Perpetrators of Intimate Partner Violence: Conclusions from a Clinical Research Perspective. *Public Health Reports* (1974-), 121(4), 369–381.
- Eckhardt, C. I., Murphy, C. M., Whitaker, D. J., Sprunger, J., Dykstra, R., & Woodard, K. (2013). The Effectiveness of Intervention Programs for Perpetrators and Victims of Intimate Partner Violence. *Partner Abuse*, 4, 196–231.

Feder L, Wilson D.B., Austin S. (2008) Court-mandated interventions for individuals convicted of domestic violence. *Campbell Systematic Reviews* 12.

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System, *Econometrica*, 46(4), p. 931.

Hua, J. & Fitzgerald, J. (2006). *Matching court records to measure re-offending*. Crime and Justice Bulletin no. 95. Sydney: NSW Bureau of Crime Statistics and Research.

Kleinbergen, F. & Paap, R. (2006). Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of Econometrics*, 133(1), 9-126.

KPMG. (2016). *The cost of violence against women and their children in Australia*. Retrieved 17 Jan 2018 at: https://www.dss.gov.au/sites/default/files/documents/08_2016/the_cost_of_violence_against_women_and_their_children_in_australia_-_summary_report_may_2016.pdf

Mackay, E., Gibson, A., Lam, H., & Beecham, D. (2015). *Perpetrator interventions in Australia: State of knowledge paper*. Australia's National Research Organisation for Women's Safety, Sydney.

Migliore, C., Ziersch, E., & Marshall, J. (2014). *Interventions Orders and the Intervention Response Model: Evaluation Report 3 (Statistical Overview and Outcome Evaluation)*. Office of Crime Statistics and Research, South Australian Attorney-General's Department.

Miller, M., Drake, E., & Nafziger, M. (2013). *What Works to Reduce Recidivism by DV Offenders?* (Report No. 13-01-120). Washington State Institute for Public Policy, Olympia.

Pence, E., & Paymar, M. (1993). Education groups for men who batter: The Duluth model. New York, NY, US: Springer Publishing Co.

Smedslund G, Dalsbø TK, Steiro A, Winsvold A, & Clench-Aas J. (2011). Cognitive behavioural therapy for men who physically abuse their female partner. *Cochrane Database of Systematic Reviews* 18: 2.

Wan, Y., Poynton, S., van Doorn, G. & Weatherburn, D. (2014). *Parole Supervision and Re-offending*. Trends and Issues in Crime and Criminal Justice No. 485. Canberra: Australian Institute of Criminology.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Jennifer Galouzis from Corrective Services NSW for providing the data from OIMS and Mark Ramsay and Mai Ho for linking data to ROD. Additional thanks are due to Chris Blatch for feedback, the two anonymous reviewers for their comments, the BOCSAR research team with thanks to Steve Yeong and Joanna Wang for providing advice and input on the analysis. Additional thanks to Paul Nelson for comments on an earlier draft of this work and Florence Sin for desktop publishing this report.

APPENDIX

SUPPLEMENTARY ANALYSES

The following analyses examine two possible reasons for the low start rate observed in the DVEQUIPS referral cohort. The first is the possibility that non-starters tended to not have intimate partner violence (IPV) offences, causing them to be flagged as unsuitable in the pre-program suitability interview. The second is whether defining starting by attending four sessions (as opposed to one session) has a significant impact on the proportion of offenders who started the program.

IPV Offending

As previously discussed, the data used for this study records offenders referred to the program before they underwent their pre-program suitability interview. Therefore, there could be unobserved differences between the suitability of offenders who were referred to the program which would be flagged in these interviews. One of the factors that might cause someone to be flagged as unsuitable for DVEQUIPS is if they did not have an IPV offence as the program is geared towards addressing gendered violence, and not having an IPV treatment need is explicitly stated as a reason for unsuitability in the program guidelines.

To check whether this was the case, the referral data was linked to the COPS database to obtain the relationship between victim and offender for each incident related to offenders' index appearance. This involved obtaining the police H-number associated with the offences they were sentenced for at finalisation and subsequently linking the related data on the relationship between the offender and the victim for each relevant incident. This data is not recorded consistently at an incident level. We cannot observe the offender-victim relationship in the majority of incidents; the offender-victim relationship was not recorded for 3,184 out of the 6,218 (51.2%) incidents linked to the offender group.

However, offenders tended to have multiple incidents related to their index offences. This meant that most offenders had at least one incident where the offender-victim relationship was recorded. After aggregating the incident-level data for each of the 1,164 DVEQUIPS referrals whose data could be linked to the COPS database, at least one offender-victim relationship could be found for 942 offenders. In other words, 222 offenders did not have any incidents where the offender-victim relationship was recorded.

Out of these 942 offenders who had at least one incident where the offender-victim relationship data was available, 351 offenders out of 373 (94.1%) who started the program and 476 out of the 569 (83.7%) who did not start the program had at least one recorded incident against an intimate partner (Table A1). A chi-squared test indicated that these differences were statistically significant ($\chi^2= 22.94$; $p < .001$). Therefore, we repeated the multivariate analyses only for offenders with a known IPV offence (Table A2). These returned similar results as those in the body of the report: that is, the partial *F*-statistic for the IV was above 10,

the C-statistic remained insignificant (i.e. there was no evidence for endogeneity) and the estimates of the treatment effect were not significantly different to zero for both general and DV-related re-offending.

Despite these non-significant differences in the presence of a known IPV offence between those who start and did not start DVEQUIPS, caution should be exercised given the scale of missing data at the incident level.

Definition of program commencement

Another potential reason for the low commencement rate observed in the data is how commencement was defined. In

this evaluation DVEQUIPS commencement was defined as completing a minimum of four out of 20 sessions, or equivalent to at least one module of the program. Redefining starting rates as attending one or more sessions improves the starting rate from 38.2 per cent to roughly 46.6 per cent, still less than half of the sample and well below the 81 per cent commencement rate recorded in the Blatch et al. (2016) study. The results of the models estimated using this definition of starting DVEQUIPS indicates that starting one session of the program is strongly associated with an increase in general and DV re-offending (Table A3) in the logistic regression estimates.

The results in Table A3 should also be interpreted with caution.

Table A1. DVEQUIPS starters and non-starters, and whether they had at least one offence against an intimate partner (offenders with no recorded data removed)

Whether had a known IPV offence	Did not start	Started	Total
No	93 (16.3%)	22 (5.9%)	115 (12.2%)
Yes	476 (83.7%)	351 (94.1%)	854 (87.9%)
Total	569	373	942

Table A2. Summary of models predicting impact of starting DVEQUIPS on general and DV re-offending (offenders with no known IPV offence excluded)

Outcome Model	Re-offending within 12 months free time			DV re-offending within 12 months free time		
	Logistic	2SLS	Bivariate probit	Logistic	2SLS	Bivariate probit
Estimate of treatment effect (Standard error)	-0.22 (0.15)	-0.21 (0.21)	-0.32 (0.43)	0.15 (0.19)	0.12 (0.15)	0.28 (0.39)
Location fixed effects	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
N	1089	1048	1048	1060	1048	1048
AIC	1383.56	1417.52	2556.80	1144.69	1141.34	2322.56
Partial F-statistic		34.58			34.58	
C-statistic (Endogeneity test)		0.65 (p=0.420)			0.38 (p=0.536)	
Correlation parameter (Standard error)			0.12 (0.26)		0.54	-0.11 (0.25)

Table A3. Summary of models predicting impact of starting DVEQUIPS within 12 months of referral on general and DV re-offending, starting defined as completing one or more sessions

Outcome Model	Re-offending within 12 months free time			DV re-offending within 12 months free time		
	Logistic	2SLS	Bivariate probit	Logistic	2SLS	Bivariate probit
Estimate of treatment effect (Standard error)	1.10*** (0.16)	-0.27 (0.18)	-0.67 (0.39)	1.58*** (0.01)	0.01 (0.12)	0.07 (0.37)
Location fixed effects	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
N	1230	1187	1187	1202	1187	1187
AIC	1563.69	1665.01	2908.73	1315.66	1309.71	2674.59
Partial F-statistic		53.36			53.36	
C-statistic (Endogeneity test)		1.74 (p=0.187)			0.03 (p=0.875)	
Correlation parameter (Standard error)			0.46 (2.77)			0.12 (0.24)

They do not necessarily indicate that starting a session DVEQUIPS leads to more re-offending. One reason is that there could be selection effects within the starting cohort which affect re-offending which the design does not compensate for. Our IV approach deals with selection bias between starters and non-starters, but does not disentangle any specific unobservable selection effects within the starter group, for example any unobserved variables which correlate with dropping out of the program early and re-offending. The potential for further bias means that the definition of treatment should be considered carefully in evaluating these programs; certainly in the case of a 20-session program, we expect that a higher threshold for program commencement is a better measure of having started the program than attending a single session.

Participation in other EQUIPS modules

The last sensitivity check on the findings was conducted by examining participation in other EQUIPS modules among those referred to DVEQUIPS. As the EQUIPS set of programs was designed to be administered as part of a treatment pathway for offenders, DVEQUIPS referees may also have been referred to and completed another EQUIPS module around the time they were referred to DVEQUIPS. This might have precluded them from participating in DVEQUIPS in the 12 months following their referral causing them to fall within the control group for this study (did not start DVEQUIPS within 12 months of referral) and

potentially have a lower risk of re-offending as a result of their completion of the other programs. This would bias our results against a treatment effect for DVEQUIPS.

To investigate this possibility, we obtained data from OIMS on whether offenders referred to DVEQUIPS completed another EQUIPS program in the 12 months following referral to DVEQUIPS. Table 3 shows the incidence of completion of another EQUIPS module among DVEQUIPS starters and non-starters. While not perfectly correlated with DVEQUIPS participation status, a greater proportion of non-starters had completed another EQUIPS module in the 12 months after referral to DVEQUIPS ($\chi^2= 21.61$; $p < .001$) compared to DVEQUIPS starters.

Consequently, we re-estimated our models excluding those who completed other EQUIPS programs to ensure that our estimates of the effect of DVEQUIPS is robust to any potential effects of the greater incidence of completion of other EQUIPS programs among the control group. We report the results of these estimations in Table A5. Note that as in all our previous estimations, we find no evidence for endogeneity, and therefore we prefer the more efficient logistic regression estimates of the treatment effect. For general re-offending, the estimated treatment effect is negative but not significant, mirroring our estimates in our original specification. For domestic violence re-offending, we find a positive but insignificant effect, as we previously do. Hence, we conclude that the results are not significantly affected by bias arising from the completion of other programs.

Table A4. DVEQUIPS starters and non-starters, and whether they had completed any other EQUIPS program in the 12 months after the referral date

Other EQUIPS program completion in 12 months from referral date	Did not start	Started	Total
No	570 (81.4%)	405 (91.4%)	975 (85.3%)
Yes	130 (18.6%)	38 (8.6%)	168 (14.7%)
Total	786	487	1273

Table A5. Summary of models predicting impact of starting DVEQUIPS within 12 months of referral on general and DV re-offending, excluding completers of other EQUIPS programs

Outcome	Re-offending within 12 months free time			DV re-offending within 12 months free time		
	Logistic	2SLS	Bivariate probit	Logistic	2SLS	Bivariate probit
Estimate of treatment effect	-0.11	0.14	0.32	0.22	0.25	0.60
(Standard error)	0.17	0.25	0.41	0.17	0.23	0.47
Location fixed effects	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
N	923	895	895	900	895	895
AIC	1141.47	1181.90	2149.62	883.10	898.30	1893.77
Partial F-statistic		25.11			25.11	
C-statistic (Endogeneity test)		0.42 ($p=0.517$)			0.95 ($p=0.330$)	
Correlation parameter			-0.23			-0.28
(Standard error)			(0.25)			(0.30)