CrossMark

Correction to: effectiveness of a prolonged incarceration and rehabilitation measure for high-frequency offenders

N. Tollenaar[1] · A. M. van der Laan[1] ·
P. G. M. van der Heijden[2,3]

Correction to: J Exp Criminol (2014) 10:29–58
https://doi.org/10.1007/s11292-013-9179-y

In the article "Effectiveness of a prolonged incarceration and rehabilitation measure for high-frequency offenders", the results are flawed due to errors that were discovered while carrying out a replication study on newer data. These errors consist of defects in the syntax for processing the data. Because these defects occured early in the data processing, they influenced all results where matching took place. We will first discuss the errors and describe their unique effect on the results. Then we discuss the actual net combined effect on the final results.

The errors consisted of the following syntactical bugs and a design flaw:

– For the simultaneous control group, wrong start and end dates of the incarceration belonging to the prison term index case were matched. This caused the calculated

---

✉ N. Tollenaar
n.tollenaar@minvenj.nl

A. M. van der Laan
a.m.van.der.laan@minvenj.nl

P. G. M. van der Heijden
p.g.m.vanderheijden@fss.uu.nl

[1]   Research and Documentation Centre (WODC), Turfmarkt 147, Den Haag 2511 DP, Netherlands

[2]   Utrecht University, Utrecht, Netherlands

[3]   University of Southampton, Southampton, UK

pre-detention period to partially overlap the actual detention period. This had a strong downward influence on its pre-treatment conviction frequency per year free[1] measures. On the other hand, the calculated time-at-risk *after* treatment was shorter than the actual time-at-risk, and caused the conviction frequencies post-treatment to be on average much larger. The error had virtually no effect on the ISD effect estimate of recidivism prevalence and minor effects on the descriptive results of the background characteristics. However, this error caused a major overestimate of the recidivism frequency outcomes in the simultaneous control group, thereby overestimating the effect of ISD. Finally, because the index case had to be redefined using the correct incarceration dates, the effective *n* decreased from 6,652 to 6,247.

– For the ISD-group, errors were made in the construction of the exposure time prior to incarceration in ISD, resulting in exposure times higher than the intended four years for 77% of the group. This too large exposure time was used in the denominator of the pre-treatment conviction frequency per year free measure causing this measure for these individuals to be too small. On average, correcting for this increased this frequency by 1.4. In the descriptive statistics, this mainly had influence on that specific variable. In the matching, correcting these errors caused the matched pre-treatment frequency for both control groups to be higher. The prevalence outcomes were virtually unaffected. Although the correction increased the starting level on pre-treatment measure in the ISD group, the absolute value of the treatment effect in the analysis that combines propensity score matching and difference-in-differences (PSM-DD) were mainly unaffected. However, the frequency outcomes in propensity score matching (PSM) were severely affected.

– Due to failure to account for operator precedence in syntax, the upper time limit of the selection of prison records for the historical control group was erroneously set to 31-12-2014 instead of the intended 1-10-2004 (the introduction date of the ISD). This left too many records (1-10-2014 up to 31-12-2004) in this group. Correcting this results in a slightly lower *n* (i.e. 3,932 instead of 4,092). In itself, this had a negligible effect on the effect estimates and the descriptive results.

– In combining the *t*-test for differences on the post-intervention frequencies per year free, the wrong standard error was applied in calculating the *t*-statistics. The average standard error was used instead of the correct standard error based on the total variance. As the differences between these standard errors were very small, this error biased the *t*-statistics in Table 3 slightly in both directions and had little effect on the *p*-values. In itself this error did not alter any of the decisions of the statistical tests.

– The data processing syntax also contained a design flaw. The multiple imputation was performed on the total group of all frequent offenders in the Netherlands, including information of convictions after leaving prison or ISD.[2] However, this

---

[1] This statistic is obtained by dividing the number of convictions (or offences) by the observation time in days (in the case of pre-treatment: four years previous to ISD or prison; in the case of post-treatment: the maximum observation time after ISD or prison) minus the total number of days incarcerated in that period and rescaling to one whole year.

[2] Including treatment outcomes as imputation variables is known to improve the quality of imputations and reducing bias in the treatment effect estimate (see e.g. Hill, 2004 referenced in the original article).

Table 3 Imputed means and standard deviations of ISD and control groups

		ISD (A) Mean (sd)	Historical control group (B) Mean (sd)	Simultaneous control group (C) Mean (sd)	(A) vs (B)	(A) vs (C)
Total group	Reconvictions per year free	3.4 (6.1)	4.0 (5.7)	3.5 (5.1)	$t(df=505.36)=-1.41$ $d = 0.09$	$t(df=191.55)=-0.13$ $d = 0.01$
	Recorded offences per year free	5.2 (10.2)	6.8 (10.7)	5.8 (9.7)	$t(df=442.30)=-2.57^{**}$ $d = 0.16$	$t(df=60.35)=-1.06$ $d = 0.06$
Recidivists	Reconvictions per year free	4.7 (6.7)	4.2 (5.8)	4.0 (5.3)	$t(df=992.22)=1.16$ $d = 0.08$	$t(df=311.73)=1.72;$ $d = 0.12$
	Recorded offences per year free	7.2 (11.4)	7.8 (10.9)	6.7 (10.0)	$t(df=760.66)=0.14$ $d = 0.01$	$t(df=117.32)=0.68$ $d = 0.05$

* = $p<0.05$; ** = $p<0.01$; ***=$p<0.001$

These t-statistics are combined from five t-statistics from the five imputations

causes imputation information of outcomes to be shared between observations that belong to different treatment groups; this may cause attenuation of the treatment effect. To correct for this, multiple imputation was performed separately for each treatment and control condition, including the outcome variable, conviction frequency per year free. By itself, this step had little effect on the effect estimates and the descriptive statistics.

The pivotal statistic that is affected – predominantly by the first two errors mentioned above – is the number of convictions (or offences) per year free prior to incarceration.

This resulted in major changes involving Table 3 and Fig. 4, and correct versions of Table 3 and Fig. 4 are provided in this corrigendum. The outcomes on different forms of recidivism frequency are substantially affected and this changed a conclusion with respect to the recidivism frequency per year free.

In the original article, two alternative methods were compared to investigate whether the post-intervention reconviction frequency per year free was lower for the ISD-group, namely propensity score matching (PSM), and a combination of propensity score matching and difference-in-differences (PSM-DD). In the 2014 version of the paper both methods pointed in the same direction, namely that ISD reduces this frequency.

After correcting for all errors simultaneously, the corrected results show that:

–   The PSM effect estimates of the conviction frequencies per year free after release are dramatically different. Using only propensity score matching (Table 3), only the total group recorded number of offences is now significantly lower for the ISD relative to the historical control group. All other reconviction and recorded offence frequencies are no longer statistically different; Cohen's d now ranges from 0.01 to 0.19 (this was from 0.21 to 0.27);

–   In the combination of propensity score analysis and difference-in-difference analysis (PSM-DD), the results only substantially change for the simultaneous control group: the estimated effect on number of convictions per year free is 3.2 lower for the ISD group (this was 3.3) compared to the historical control group and 3.8 lower (this was 5.8) compared to the simultaneous control groups;

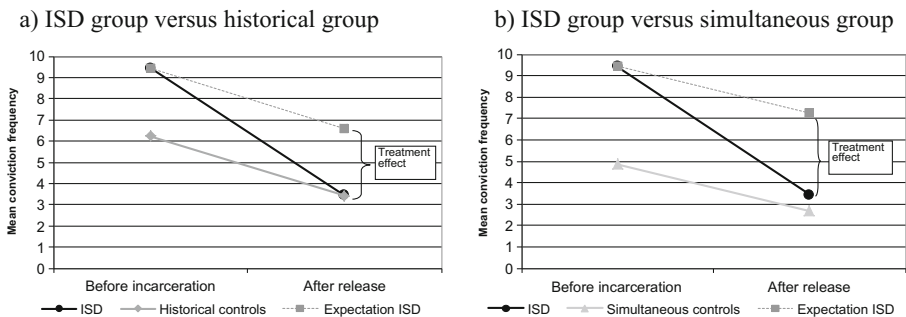–   The distribution of propensity scores does not alter visibly;

Fig. 4 Difference-in-difference estimates of the ISD-group versus the two control groups on average number of convictions per year free

– The incapacitation effect is estimated to be on average 5.6 (was 5.7) criminal convictions; 9.1 (this was 9.2) recorded offences were prevented. The preventive effect of ISD is on average 2.4 (this was 2.5) convictions and four (unchanged) recorded offences per chronic offender per year;
– The relative distribution of prevented offence type does not change substantially;
– The recidivism prevalence changes marginally:

the corrected one-minus-survival plots are virtually indistinguishable from the erroneous plots;
the ISD recidivism reducing estimate using the historical control group is 17.6% (was 16%) with an effect size of Cohen's h of 0.46 (was 0.42). The tests are significant at the same level;
the ISD recidivism reducing estimate using the simultaneous control group is the same; only Cohen's h changed from 0.29 to 0.30. The tests are now significant at the 0.0001 level (this was 0.001);

– Descriptive statistics of background characteristics change marginally;
– The coefficients of the propensity score model change marginally.

On recidivism frequency, the recalculated results outcomes now reveal a substantial discrepancy between the results of row 1 in Table 3 (reconviction frequency per year free of the total group, PSM) and Fig. 4 for the number of reconvictions per year free (PSM-DD). We would expect the estimated effects for both methods to be approximately the same.

Why do we find this large discrepancy between the results of the PSM and the PSM-DD? Several explanations might hold:

• There is a large residual bias in the PSM analysis without DD;
• There is a hidden bias in the PSM-DD analysis that is not constant over time;
• The parallel slopes assumption of DD does not hold and the expected slope may be more extreme in the ISD-group, regardless of the treatment effect.

The latter seems a plausible explanation because judges and public prosecutors select HFOs on a high pre-trial recidivism frequency. As recidivism frequency is a measure subject to measurement error (i.e., the process in which recidivism is detected is highly subject to chance), there can be a considerable amount of regression towards the mean. As the control group is not matched on the pre-treatment measure in a PSM-DD analysis, its pre-treatment scores are less extreme than those for the ISD-group. Therefore, the groups will suffer from different amounts of regression towards the mean, biasing the effect estimate. How to correct for this bias in a PSM context may be subject for further research.

There are other, minor changes to the text and the tables in the 2014 paper that are less relevant. Interested readers are invited to contact us if they are interested in these changes. We will then send a corrected version of the 2014 paper.