Recidivism and education revisited: evidence for the USA

James Fogarty[a*] and Margaret Giles[b]

[a] Agricultural and Resource Economics,
UWA School of Agriculture and Environment,
University of Western Australia,
Crawley, WA 6009, Australia

[b] School of Business and Law,
Edith Cowan University,
Joondalup WA 6027, Australia

*E-mail address: James.Fogarty@uwa.edu.au

2 August 2018
Working Paper 1806
UWA Agricultural and Resource Economics
http://www.are.uwa.edu.au

Citation: Fogarty, J. and Giles, M (2018) *Recidivism and education revisited: evidence for the USA*, Working Paper 1806, Agricultural and Resource Economics, The University of Western Australia, Crawley, Australia.

Recidivism and education revisited: evidence for the USA

James Fogarty* and Margaret Giles†‡

July 2018

Abstract

In this research we present a replication study of an earlier meta-analysis that investigated the effect of correctional education programs on recidivism. The original study did not correct for publication bias, so testing for and correcting for publication bias is the focus of this replication study. Our findings suggest that publication bias may have lead to a modest over-estimate of the positive effect of correctional education programs on recidivism rates in the original study, but after correcting for publication bias, the evidence that correctional education programs are cost-effective remains strong.

Key Words: Meta-Analysis, Recidivism, Education
JEL Codes: I21, I28

*Agricultural and Resource Economics, University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia (James.Fogarty@uwa.edu.au)

†School of Business and Law, Edith Cowan University, 270 Joondalup Drive Joondalup WA 6027, Australia (M.Giles@ecu.edu.au)

‡The authors would like to sincerely thank both Lois Davis and Robert Bozick for engaging with this replication study and providing comments and feedback on a draft version of the paper. Without their input to the process it would not have been possible to complete this work.

1 Introduction

The literature investigating the effect of prison education and vocational training on recidivism rates is diverse, and as Kelso (1996, p. 29) observes: "Without the benefit of specific knowledge of the methods and philosophies employed, it is difficult to draw any definitive conclusions". Not only are there different methods and philosophies driving the delivery of different courses, there are also disparate ways of conducting evaluations of program efficacy, as well as divergent approaches to reporting evaluation metrics. This heterogeneity complicates the appraisal of 'what works' and frustrates stakeholders, including prison authorities and education funders. Despite significant differences in study design, primary studies that report quantitative results for the effect of prison education programs generally report the difference in the recidivism rates between one or more treatment groups (participating in or completing different types of education and vocational training) and at least one comparison group. This difference in the recidivism rates metric in turn provides a common reference point for conducting a structured evaluation of the literature as a whole.

Recidivism, can, however, be defined in terms of re-offending for a new offence or breach of parole for a previous offence, and the re-offending may involve being re-arrested, re-charged, or re-imprisoned. Recidivism is also defined in terms of a follow-up period which might be any time from release to one or more years since release. Hence, recidivism could be as diverse as a breach of parole within six months of release from prison or re-imprisonment in the same prison after five years in the community following a prior period of incarceration. Often the recidivism measure used simply reflects the record keeping of the justice system or the prison in which the correctional education intervention is being trialed. In cost-benefit analyses, the choice of recidivism measure will dictate the dollar benefits of any prison program, including education and vocational training courses. Re-imprisonment is the most expensive of these recidivism measures, irrespective of whether it reflects a new offence or a breach of parole (prior offence with a non-custodial or community corrections sentence period), and this is the most commonly reported measure. It is also the measure we focus on in this review.

It is valuable to place this study within the context of the findings of earlier systematic reviews and meta-analyses of the effectiveness of correctional education. The earliest relevant systematic review is Lipton, Martinson, and Wilks (1975), which is a review of 231 studies of prisoner rehabilitation programs published between 1945 and 1967, and this review concluded that prison study did not reduce recidivism. Later reviews by Gerber and Fritsch (1995) and MacKenzie and Hickman (1998) produced what Wilson et al. (2000) referred to as "equivocal conclusions with regard to program impacts".

The meta-analysis by Wilson et al. (2000) of 33 primary studies of correctional education in U.S. and Canadian prisons concluded that, whilst "program participants recidivate at a lower rate than nonparticipants", the included studies were methodologically weak, and hence the results of the meta-analysis were not definitive. Both Wells (2000) and Chappell (2003) used meta-analytic techniques in their theses to examine the relationship between correctional education and recidivism, and both found support for the efficacy of correctional education in reducing recidivism. MacKenzie (2006), in a follow-up to the Wilson et al. (2000) study, disaggregated primary studies into various types of education and found all education program types, except life skills programs, were effective at reducing post-release criminality. In their systematic review of 571 evaluations of correctional education, Aos, Miller, and Drake (2006) found evidence that some, but not all programs reduce recidivism.

Most recently, RAND Corporation was sponsored by the U.S. Bureau of Justice Assistance to re-examine the effectiveness of correctional education in the U.S., by undertaking a methodologically rigorous meta-analysis. Now known as the RAND report, or Davis report after lead author Lois Davis, this publication, which we subsequently refer to as Davis et al. (2013), is widely cited by prison educators in the U.S. and internationally. The meta-analysis included fifty U.S. studies of the effectiveness of correctional education published in scholarly journals or appearing in the grey literature over the period 1980-2011. Specifically, the meta-analysis estimated the average effect size – odds of recidivating – from odds ratios or differences in recidivism rates for treatment and comparison groups of prisoners and found strong evidence that participation in correctional education reduced recidivism rates. For included studies, the treatment group was a group of prisoners who undertook a course of education or vocational training in prison and were subsequently released from prison with sufficient time in the community to record recidivism. Courses included preparation for General Equivalency Development or General Equivalency Diploma (GED) and successful GED completion (thought to be equivalent to achieving a high school diploma), adult basic education, vocational training certificate and diploma courses, college-level courses, and undergraduate and postgraduate university courses.

The Davis et al. study involved a comprehensive search for papers, across both the published and grey literature.[1] Further, as it is difficult for researchers to access recidivism data, and as both null results and statistically significant results are of interest in the field, it is unlikely there is a large selection of unpublished papers that were not identified in the literature review. However, as Gelman and Carlin (2014) discuss, when there are low power studies in the literature

[1] This search included a comprehensive list of databases and catalogs, including: Education Resources Information Center (ERIC); Education Abstracts; Criminal Justice Abstracts; National Criminal Justice Reference Service Abstracts; Academic Search Elite; EconLit; Sociological Abstracts; Google Scholar; Rutgers Library of Criminal Justice Grey Literature Database; Vera Institute of Justice; Urban Institute; Washington State Institute for Public Policy; American Institutes for Research; Mathematica Policy Research; John Jay College of Criminal Justice Re-entry Institute; Justice Policy Institute; Center for Law and Social Policy (CLASP); Juvenile Justice Educational Enhancement Program; (JJEEP) RTI International; and Manpower Demonstration Research Corporation (MDRC).

there can still be significant sign and magnitude errors with published studies, and so the issue of publication bias is not just an issue of possible 'missing' studies.

Results for a standard regression publication bias test and a rank based publication bias test are reported in Davis et al. and for the standard regression based test the null hypothesis of no publication bias is rejected, but for the rank based test the null of no publication bias is not rejected. On the basis of the comprehensiveness of the literature search process, and the rank based test results, no publication bias correction is implemented in Davis et al. Rank based tests have low power, and so failing to reject the null with this type of test is not a strong argument against implementing a publication bias correction; especially when the data also fail a widely used publication bias test, and there are a number of studies where the sample size is small.[2] Further, a recent structured review of 128 separate literatures found that in just over half of all literatures reviewed, the effect size was overstated by at least 100%; and in 34% of cases the literature overstated the treatment effect by at least 300% (Ioannidis et al. 2017). In other words, recent evidence suggests publication bias is widespread across many disparate literatures, and the inflation of the effect size due to publication bias is substantial. In the specific context of the effectiveness of correctional education literature, as the evidence regarding the effect of correctional education on recidivism rates prior to the Davis et al. study was mixed, it is valuable to investigate the potential role of publication bias in this literature.

In the remainder of the paper we do three things. First, we re-estimate the Davis et al. model and confirm that we can generate the same results as reported in the original study, when we used the same data set. Next, we review the same primary literature as used in Davis et al. and identify what we think are the most appropriate studies and recidivism rate estimates to use in a meta-regression model and explain our reasoning for selecting these values. Finally, we explore the role of publication bias and effect size heterogeneity and derive new estimates of the effect of correctional education programs, corrected for publication bias and effect size heterogeneity.

2 Data and methods

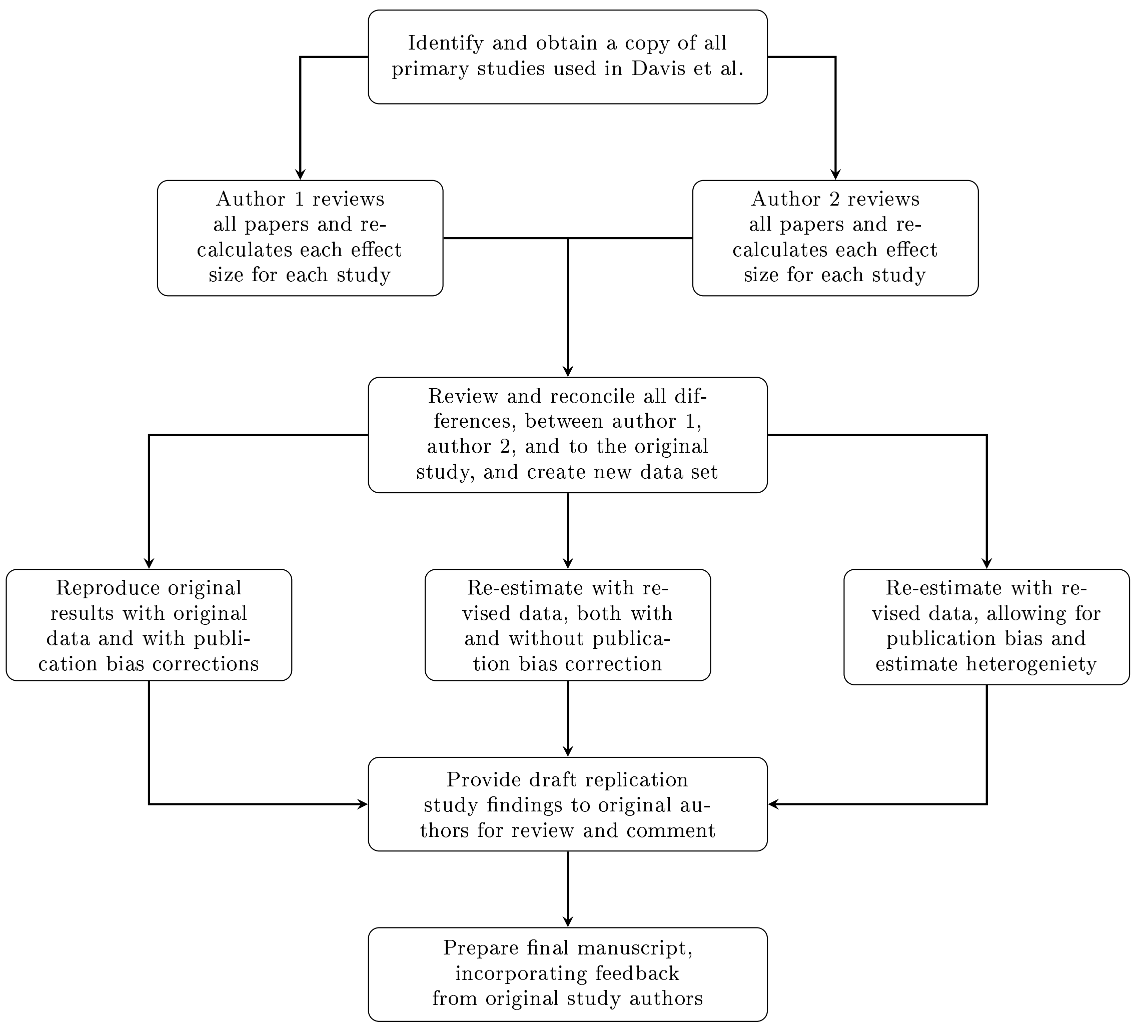
2.1 Data

An overview of the research flow for the replication study is shown in Figure 1, and as can be seen, the first step was to obtain copies of the same papers reviewed in Davis et al. : these papers

[2]Similarly, in Bozick et al., (2018), which covers similar ground to Davis et al., it is reported that the meta-summary of effect sizes fail a publication bias test, but no correction for publication bias is applied. In this case the justification for not using a publication bias correction is evidence from funnel plots. We argue that evidence from funnel plot is not sufficient to not use a publication bias correction, when it is known the data fails a formal publication bias test.

represent the base data set. At the combined data reconciliation step it became evident that there were several differences between the values used in Davis et al. and the values we identified as most appropriate. The main reasons for differences between the values used in Davis et al. and the values we use are: (i) the definition of what should be included as part of a review of education programs; (ii) whether the effect size should be calculated using the raw proportions data in the primary study or from regression model information reported in the primary study; and (iii) the extent of data pooling undertaken to arrive at an effect size estimate. A complete reconciliation of the values used in this study and the values used in Davis et al. is provided in Table A.1 of the Appendix. In many cases there is no clear right approach to selecting summary values, but an overview that argues in favour of the values we use is presented below.

Figure 1: A summary of the steps in the replication review process



The study definition used in Davis et al. was agreed with senior staff at the U.S. Department of Justice and the U.S. Department of Education and for the study an education program was defined as a program "that includes an academic and/or vocational curriculum taught by an instructor, designed to lead to the attainment of a degree, license, or certification. The program could be part of a larger set of services administered to inmates or it could be a stand-alone program." A key reason for this approach was that prisons nearly always offer a range of rehabilitation services, regardless of whether or not the inmate is participating in correctional education, and the approach is a reasonable approach to take. In our review of the primary literature we identified four studies included in Davis et al. where we felt the education component was a minor part of much more comprehensive and substantial prisoner rehabilitation programs, or were studies of education more generally rather than specific evaluations of prison based education and training programs. These studies were: Castellano (1996), Lattimore (1988), Nally (2011), and Van Stelle (1995). Understanding the impact of more comprehensive prisoner rehabilitation programs is valuable; but in our view these studies are not relevant to a review of the effectiveness of specific prison education programs. As such, we exclude these four studies from the final database we use to estimate the publication bias corrected impact of correctional education programs on recidivism rates. A potential risk with our approach is that we have only identified those studies where other rehabilitation services are explicitly mentioned in the paper. We acknowledge that general rehabilitation programs are available in prisons, but for the studies we exclude we think the education component to be small.

For studies based on controlled trials, or studies based on matched data samples, estimating the effect of prison education programs from the raw recidivism proportions data for each group is appropriate. However, many of the studies included in the review are not controlled trials or studies that use matched samples. For these studies, due to the potential for there to be self-selection bias into the treatment group, calculating the effect size from the raw proportions data is problematic. For each study that is not a controlled trial, or a study based on matched samples, the relevant qualifying information is fully, and correctly documented in the Davis et al. supplementary files, but the primary estimates from all studies are included in the main study meta-analysis. We argue for a different approach.

For primary studies that were not controlled trials or did not use matched samples, typically the authors used a regression modelling approach to investigate the effect of correctional education programs. With a regression modelling approach covariates for prisoner attributes are used to 'control' for differences in prisoner attributes. As illustrated in Allen (2006) the method of instrumental variables can also be used to address endogeneity issues associated with self-selection into education programs. If a primary study uses a regression modelling approach to estimate the effect of participation in education programs we argue that the regression model results are preferable to estimates derived from the raw proportions data: at least in theory the regression model controls for subject heterogeneity. For a number of studies that report both regression model results and raw proportions data on recidivism rates Davis et al. rely on the raw proportions data and not estimates derived from the regression model. We prefer effect size estimates derived from the regression model results reported in primary studies.

To support our view that the regression model results rather than the raw proportions data should be used we note that in the primary literature whenever a regression modeling approach is used, the author(s) of the primary study always refer(s) to the results from the regression model when drawing conclusions, not the raw recidivism proportions data. Using the raw proportions data can also lead to a disconnect between the central conclusion of a study and the conclusion that would be drawn based on the raw proportions data. For example, using the proportions data Davis et al. derive an odds ratio for Lattimore (1988) of 0.58 (95% CI 0.38 to 0.89), which suggests a large, positive, and statistically significant effect due to participation in education programs. Lattimore (1988 p. 143), however, report: "...as we were unable to establish a link between vocational training and better post-release employment, we expected to find no differences in the recidivism of our study groups. Results in this Chapter confirmed this a priori hypothesis in that we were unable to find any differences (at the $\alpha = 0.05$ level of significance) in post release recidivism... for members of the study groups". The primary study reached a conclusion of no effect, but if the raw proportions data is used the opposite conclusion is reached. This example demonstrates that at least in some cases it is possible for there to be a substantial, and practical difference between the effect size calculated using raw recidivism proportions data and the effect size calculated from a regression model coefficient. We argue that if a primary study uses a regression model to control for differences in prisoner attributes the effect size based on the regression model results not the raw proportions data is the correct value to use.

The final difference between the approach used in this study and the approach used in Davis et al. relates to the extent of data pooling. A number of primary studies reported results separately for prisoners that participated in an education program and prisoners that completed an education program. Some studies also reported results for vocational programs separately to academic education programs. Other studies also reported results at multiple time periods. In general, the approach taken in Davis et al. was to focus on the results for one-year post-release, or as close as possible to one-year post-release, and use the details for pooled program results when pooled information on program type was available. This is a valid approach, but for this study we prefer to include multiple estimates from a given study, and then use estimation methods that: (i) control for data dependence across estimates from the same study; and (ii) restrict the weight to each estimate such that the weight for a specific observation from a study is equal to the average level of precision for estimates from the study divided by the number of estimates drawn from that study.

We derive effect size estimates from a variety of primary reporting formats. To standardize the effect size estimates to a common format, and move between estimates on the odds scale and estimates the log odds scale, and derive appropriate variance estimates, we use the routines available in Re (2013).

2.2 Methods overview

The first step in the modelling approach was to reproduce the original study results using the same primary data values. When the data are summarised this way it is not possible to reject the null hypothesis of no publication bias. As such, we next calculate the pooled effect size estimate using two alternative publication bias corrections. The first method used to correct for publication bias follows Stanley and Doucouliagos (2014) and involves two steps. First, a weighted least squares (WLS) meta-regression is estimated where the study estimate standard errors are included as a covariate. This is the publication bias test step (Egger test), and a statistically significant point estimate for the standard error covariate is evidence of publication bias. The publication bias corrected estimates are then derived from the estimate of the intercept of a WLS meta-regression where the study estimate variances are included as a covariate. Stanley and Doucouliagos (2014) argue in favour of using the standard errors for the publication bias test and the variance for the correction based on extensive simulation evidence.

The second method used to correct for publication bias follows Ioannidis et al. and this method involves deleting from the meta-regression model all studies that are deemed to have low power to detect a reference effect size, where the reference effect size estimate is derived from a WLS meta-regression model of published estimates. The logic of this approach is that low power studies only identify a statistically significant effect when the observed effect size is extreme, relative to the true population effect, and when such studies are included in a meta-analysis it results in the overstatement of the true effect size. Gelman and Carlin (2014) referred to this as magnitude error. Removing low power studies from a meta-analysis therefore results in an estimate that is 'corrected' for the impact of publication bias.

Finally, we note that failing a publication bias tests is a necessary but insufficient condition to establish that publication bias is an issue. For example, in a simple meta-regression model of the form estimated in Davis et al. heterogeneity in the underlying effect size estimates can also result in failure to reject the null of no publication bias. To address this issue we also test for publication bias in a model that includes covariates for program type.

There is considerable overlap between the econometrics panel data literature, the mixed effects model literature, and the robust variance meta-analysis literature. For example, the cluster robust variance estimator for panel data given in Greene (2000, p. 353) is largely analogous to the within-study correlated effects estimator developed in Hedges et al. (2010), and both estimators are similar to the mixed effect model with a random intercept for study (Laird and Ware 1982).

This research lies within the meta-analysis tradition, and as such we use the Hedges et al. notation and estimator.

2.3 Method of estimation

The research objective is to arrive at consistent estimates of the meta-regression parameters and valid confidence intervals for hypothesis testing. To achieve this we use the Hedges et al. robust variance estimator, with both the variance and the degrees of freedom correction recommended in Tripton (2015). Formally, the meta-regression model we use can be understood as follows. Let $\boldsymbol{y}_j$ denote a vector of length $k_j \geq 1$ containing the k_j study effect size estimates of the effect of prison education programs on recidivism; and let $\boldsymbol{X}_j$ denote the associated matrix of study design attributes that captures information on aspects of the study such as the type of education program, the length of study etc. If the $\boldsymbol{y}_j$ and $\boldsymbol{X}_j$ are stacked it becomes possible to write the meta-regression model as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{e}$ is a stacked vector of the $\boldsymbol{e}_j$ error terms, each with length k_j. There are many possible approaches to estimating $\boldsymbol{\beta}$ in Equation 1, but a general approach that allows for flexible weighting structures to reflect information on estimate quality is:

$$\hat{\boldsymbol{\beta}} = \left[\sum_j \boldsymbol{X}_j' \boldsymbol{W}_j \boldsymbol{X}_j\right]^{-1} \left[\sum_j \boldsymbol{X}_j' \boldsymbol{W}_j \boldsymbol{y}_j\right], \tag{2}$$

where consistent estimates of $\hat{\boldsymbol{\beta}}$ are obtained regardless of the specific form of $\boldsymbol{W}_j$, the weight matrix.

In meta-regression it is common to distinguish between two weighting structures. The first weighting structure is the fixed effects meta-regression weighting structure, and this weighting structure considers only the within-study variance. The second weighting structure is the random effects meta-regression weighting structure and this weighting structure incorporates information on both the within-study variance and the between-study variance.

For meta-regression with dependent (multiple) effect sizes from a given study, using the raw inverse variance weighting scheme results in studies with many estimates receiving a large weight. This in inappropriate. To ensure that studies with a large number of effect size estimates do not receive excessive weight, in this application the total weight to each study is fixed at the study level. As such, the fixed effects meta-regression weighting scheme is given as $w_{ij}^f = 1/k_j\bar{\hat{v}}_{\cdot j}$, where $\bar{\hat{v}}_{\cdot j}$ is the arithmetic average effect size variance estimate across the i estimates from study j.[3] The random effects weighting structure incorporates an estimate of the between-study variance, and in this application the random effects weights are defined as $w_{ij}^r = 1/k_j(\bar{\hat{v}}_{\cdot j} + \hat{\tau}^2)$, where $\hat{\tau}^2$ is an estimate of the between-study variance obtained from the errors of an auxiliary regression estimating equation 1 using the fixed effects weighting structure (see equation (14) and equation (15) of Hedges et al. (2010, p. 46) for details).

Presenting the weighting structures this way highlights two important features of meta-regression weighting structures. First, as $\bar{\hat{v}}_{\cdot j}$ is primarily a function of primary study sample size, it is clear that in meta-regression weights are dependent on the sample size of the primary study, not the number of estimates from that study. Second, as the between-study variance increases, the weights across studies tend to equalize, and hence estimates tend towards OLS estimates.

In robust, or sandwich variance estimation, the least squares errors associated with each observation are used to compute the standard errors of $\hat{\boldsymbol{\beta}}$ (White 1980). The validity of robust variance estimators is based on large sample theory, where in meta-analysis convergence is in terms of the number of primary studies not the total number of effect size estimates. In practice, the use of the raw least squares errors to estimate the covariance matrix results in the systematic understatement of estimate uncertainty: standard errors are too small. When using robust methods it is therefore standard practice to adjust (inflate by some fixed or variable factor) the variance estimates, and the estimator used in this study can be written as:

$$Var(\hat{\boldsymbol{\beta}}) = \left[\sum_j \boldsymbol{X}_j'\boldsymbol{W}_j\boldsymbol{X}_j\right]^{-1}\left[\sum_j \boldsymbol{X}_j'\boldsymbol{W}_j\boldsymbol{A}_j\hat{\boldsymbol{e}}_j\hat{\boldsymbol{e}}_j'\boldsymbol{A}_j\boldsymbol{W}_j\boldsymbol{X}_j\right]\left[\sum_j \boldsymbol{X}_j'\boldsymbol{W}_j\boldsymbol{X}_j\right]^{-1} \tag{3}$$

where the $\hat{\boldsymbol{e}}_j$ are the errors associated with estimating equation 1 with the random effects weighting structure, $\boldsymbol{A}_j$ is the adjustment matrix for study j used to correct for the underestimation of standard errors, $\boldsymbol{W}_j$ is the same weighting matrix for study j used in equation 2, and $\boldsymbol{X}_j$ is the design matrix for study j. The specific correction used for $\boldsymbol{A}_j$ in this analysis is the correction given in equation (10) and equation (11) of Tipton (2015, p. 379), which in turn is motivated by McCaffery et al. (2001). Finally, for inference we also rely on the Satterwaite (1946) degrees of freedom adjustment, as detailed in equation (11) of Tipton (2015, p. 379), and this means that for every covariate in the meta-regression model there is a unique degrees of freedom adjustment.

[3] The meta-regression model of Hedges et al. is general enough to accommodate weighting structures that allow the weights to depend on both the study level precision and the number of observations from each study, but we prefer the weighting structure used here.

Estimation relies on Fisher et al. (2017), and the dependent variable in all regression models is the log odds ratio.

3 Results

The first column of Table 1 provides the results for the direct replication of the original study; the second column of Table 1 shows the results for the regression based test for the presence of publication bias; the third column of Table 1 shows the results using the Stanley and Doucouliagos publication bias correction; and the fourth column of Table 1 shows the results based on the Ioannidis et al. publication bias correction. The first thing to note about the results is that the direct replication estimate of the overall effect of education on recidivism is very close to that reported in Davis et al.: odds ratio 0.61 (95% CI 0.53 to 0.70) versus 0.64 (95% CI 0.59 to 0.70) for the original study. Different software and differences in estimation routines can lead to small differences in results, but we deem our replication sufficiently close to the original result to be satisfied.

The second thing to note, which is again consistent with the original study, is that there is strong evidence of publication bias: the estimate for the standard error covariate is statistically significant. On the odds ratio scale the two approaches to correcting for publication bias give the same result (to two decimal places): odds ratio 0.70 (95% CI 0.61 to 0.81). The direct re-estimation result, corrected for publication bias, therefore suggests that the positive effect of prison education programs is materially smaller than the headline value reported in Davis et al. but the positive effect of prison education programs on recidivism is still statistically significant and substantial.

Table 1: Reproduce original result and publication bias corrections

	Base	Bias test	Corrected 1	Corrected 2
Intercept	-.492	-.183	-.350	-.355
	(.068)	(.101)	(.070)	(.068)
Standard Error		-1.69		
		(.607)		
Variance			-2.80	
			(1.11)	
τ^2	.116	.094	.100	.087
No. Studies	50	50	50	24

Note: Robust, small sample adjusted standard errors in parenthesis.

Next, we work through the same estimation steps, excluding the four studies identified as not meeting our inclusion criteria, and where appropriate, replacing study effect size values with those detailed in the appendix. The results are reported in Table 2, and despite a significant number of differences in the estimates taken from primary studies, and the removal of four studies from the sample; the results in Table 2 are similar to the results reported in Table 1.

The raw estimate of the effect of participation in an education program on the recidivism rate is an odds ratio of 0.63 (95% CI 0.65 to 0.72), and it is still not possible to reject the null hypothesis that publication bias is present. Using the Stanley and Doucouliagos publication bias correction, the estimated effect of education on the odds ratio scale is 0.75 (95% CI 0.57 to 0.75), and using the Ioannidis et al. publication bias correction the estimated effect on the odds ratio scale is 0.72 (95% CI 0.61 to 0.84). So, correcting for publication bias, correcting for studies that we think should not have been included in the original meta-analysis, and using a different approach to identify relevant effect sizes in the primary literature does not result in a material revision to the key finding of the Davis et al (2013) study: prison education and vocational training programs have a material, statistically significant, and positive impact on recidivism rates.

Table 2: Revised data and publication bias corrections

	Base	Bias Test	Corrected 1	Corrected 2
Intercept	-0.456	-0.293	-.419	-.334
	(.060)	(.080)	(.066)	(.076)
Standard Error		-.895		
		(.362)		
Variance			-.686	
			(.520)	
τ^2	.083	.066	.077	.086
No. Studies	46	46	46	21

Note: Robust, small sample adjusted standard errors in parenthesis.

The reason for detecting publication bias could be that there is heterogeneity in the effect size due to structural differences between program types rather than a true publication bias effect. To explore the effect of covariates we limit the effect sizes we include in the sample to estimates where it was possible to classify responses as either participated in a program or participated in and completed a program, and also where a single program type was evaluated: academic fundamentals, which we define as all education programs below post-secondary education; post-secondary education; or vocational education. We therefore deleted observations that were based on pooled program results, or multiple program completions, and before proceeding checked the impact of removing these observations from the sample by re-estimating all models with the restricted sample. The results for the re-estimation are shown in Table 3, and as the results are almost identical to those reported in Table 2, we conclude that the removal of these observations has not had a material impact.

Table 3: Restricted data and publication bias corrections

	Base	Bias test	Corrected 1	Corrected 2
Intercept	-.459	-.285	-.419	-.338
	(.064)	(.082)	(.069)	(.080)
Standard Error		-.968		
		(.376)		
Variance			-.733	
			(.534)	
τ^2	.083	.064	.076	.080
No. Studies	44	44	44	20

Note: Robust, small sample adjusted standard errors in parenthesis.

For the meta-regression modelling we follow a general-to-specific approach, and first estimate a model that allowed for a main effect for completed versus participated; a main effect for type of education program; and an interaction effect. The interaction effect is not significant, so we restrict the remainder of the discussion to models that have main effects only.

In the regression model results the base education type is academic fundamentals, and the base status is completed the program. The reported regression coefficients are interpreted as deviations from the base. To understand how the values in Table 4 should be read, consider the first column of results. The log odds ratio for the effect of completing a basic education program is -0.33, and this effect is statistically significant. The effect for completing a post-secondary education program is found as (-.33) + (-.73) = -1.06, and from the standard error associated with the post-secondary effect variable we conclude that the difference between completing a post-secondary education program and completing a basic education program is statistically significant. The effect of vocational education is found as (-.33) + (-.02) = -.35, and from the standard error associated with this coefficient we conclude that the effect of basic education is not statistically different to vocational education, but both types of program have a positive effect on recidivism. The participate only coefficient says that, on average, the log odds ratio for those that complete a program is .088 less than for those that participate only, but this difference is not statistically significant.

Next we estimate the regression based publication bias test, and these results are reported in the second column of Table 4. The results indicate that publication bias is still an issue, however, by comparing the values in column one and column three of Table 4 it can be seen that the Stanley and Doucouliagos correction for publication bias has little practical impact on the estimates. Specifically, for completing a basic education program the log odds change from -.33 to -.31; for post-secondary education the log odds change from -1.06 to -1.03; and for vocational education the log odds change from -.35 to -.33 (coefficients are deviations from the base, and hence must be added to obtain the actual effect).

Table 4: Heterogeneity or publication bias investigation

	Base	Bias Test	Corrected 1	Corrected 2
Basic education complete	-.330	-.251	-.314	-.225
	(.061)	(.064)	(.061)	(.033)
Post-secondary effect	-.729	-.674	-.715	-.810
	(.124)	(.143)	(.130)	(.113)
Vocational effect	-.017	-.007	-.018	.115
	(.081)	(.076)	(.080)	(.108)
Participate only	.088	.104	.091	.033
	(.085)	(.079)	(.084)	(.092)
Standard Error		-.643		
		(.318)		
Variance			-.468	
			(.426)	
τ^2	.037	.032	.036	.038
No. Studies	44	44	44	19

Note: Robust, small sample adjusted standard errors in parenthesis.

To implement the Ioannidis et al. publication bias correction we set the critical threshold for detecting an effect at the education type level. This means that we use a different threshold standard error value for each education grouping. The results for this model are shown in the final column of Table 4, and for this method of correcting for publication bias the changes are more notable. For completing a basic education program the log odds change from -.33 to -.22; for post-secondary education the log odds change from -1.06 to -1.04; and for vocational education the log odds change from -.35 to -.11.

The raw regression results table does not allow for an easy interpretation of the effect on recidivism rates, especially in terms of establishing whether a value is statistically different from zero. To make matters clear, Table 5 presents the point estimate, and the 95% confidence interval on the odds scale for each education program grouping, for the base model, and the two different approaches to correcting for publication bias.[4] As can be seen, across all results the only material difference is for vocational education when using the Ioannidis et al. publication bias correction, where the effect changes from a statistically significant and practically important effect to an estimate that is not statistically different from zero at conventional levels.

To understand why this occurs we looked in detail at the studies excluded from the sample when the Ioannidis et al. correction is applied. The logic of the Ioannidis et al. correction is that low power studies will only identify a statistically significant effect when it is extreme, hence low power studies result in an inflation of the average effect size. To investigate this specific issue we

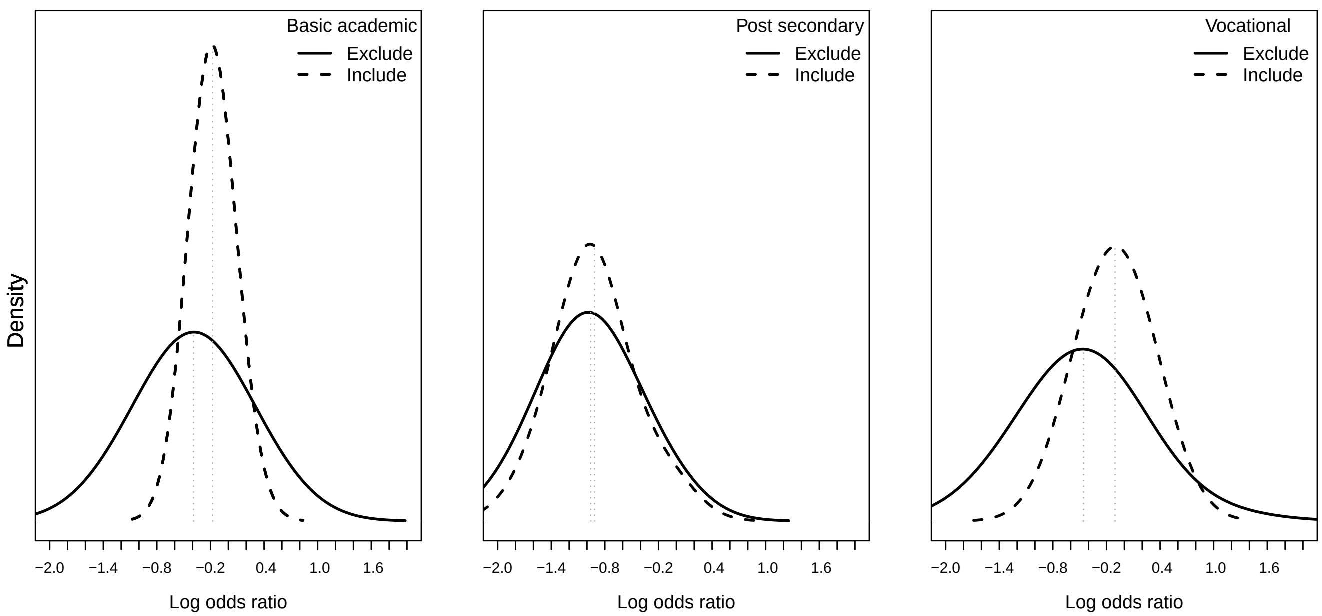
[4]It makes little difference to the results, but as the complete or participate variable is never significant this variable is dropped for these calculations.

Table 5: Odds ratios for the impact of education programs

	Academic fundamentals		Post-secondary		Vocational	
	Est.	95% CI	Est.	95% CI	Est.	95% CI
Base model	.74	.67 to .82	.36	.28 to .47	.74	.65 to .85
Correction 1	.75	.68 to .84	.37	.28 to .50	.75	.66 to .86
Correction 2	.81	.76 to 86	.36	.28 to .47	.91	.72 to 1.15

separate studies into those studies that meet the Ioannidis et al. measure of acceptable power and those that do not, and look at the distribution of effect size estimates for these two groups. As can be seen from Figure 2, across all groups, the mean effect size estimate from studies deemed low power is always stronger than the mean effect size estimate from studies deemed to have adequate power; recall lower number on the log odds scale suggests the effect of prison education programs is stronger. The difference is, however, most pronounced for the vocational education group. It is also notable that across all three groupings the variance is always larger for the sample of low power studies.

Figure 2: Density plot comparison: low power v acceptable power studies



To formally test for difference in the mean effect size across groups we use Bayesian estimation methods. Specifically, we use Kruschke and Meredith (2017) for estimation and the log odds data are modeled using a t-distribution with uninformative priors; the MCMC chain length is 10,000, with no thinning; and the burn-in length is 1,000.

For the academic fundamentals comparison of the effect size estimates reported in low power versus adequate power studies the 95% Highest Density Interval (HDI) is -0.03 to 0.46; for the post-secondary education comparison the 95% HDI is -0.34 to 0.41; and for vocational education comparison the 95% HDI is 0.12 to 0.36. So, on this basis we conclude that only for vocational education program evaluations is there evidence of a significant difference in the reported mean effect size between low power studies and studies deemed to have adequate power. This result is consistent with the result reported in Table 5.

As a final check on the vocational education estimate, we also calculated the simple average of the effect size and effect size standard error for the five studies with the most precise effect size estimates. The average across this sample of studies was a log odds ratio of -.024 and average standard error of .086, which in turn implies an odds ratio not different to one for the five most precise studies. Overall the evidence suggests that the Ioannidis et al. publication bias correction is correctly identifying an effect that is present in only a subset of studies, and this is something that is not detected when using the Stanley and Doucouliagos correction.

4 Conclusions and qualifications

In this study we have re-examined the literature on the effect of correctional education programs using both an established publication bias correction method, and a relatively new approach. For both basic education programs and post-secondary education programs the result is clear. No matter which publication bias correction method is used, there is still a large and statistically significant reduction in the risk of recidivism following participation in these programs.

The evidence regarding vocational education programs is less clear. Using the traditional publication bias correction method, a large, statistically significant reduction in the the risk of recidivism is found following participation in correctional vocation education programs, but when the Ioannidis et al. publication bias correction is used, the estimated reduction is the risk of recidivism, while still relatively large, becomes non-significant at conventional reporting levels. We are however cautious in interpreting this result.

The Ioannidis et al. method of correcting for publication bias resulted is a high number of vocational program evaluations being excluded, and the final model includes only seven studies of correctional vocational education program. This result suggests there are many low power studies of vocational education programs, and hence a tendency for statistically significant results to be outlier observations, but at the same time it also means that the vocational education estimate is relatively imprecise. This matters in the current context as Davis et al. report that from a cost-effectiveness point of view, correctional education programs only need to reduce the three-year reincarceration rate by between 1.9 and 2.6 percentage points to be cost-effective. A reduction of this magnitude is completely plausible based on the evidence summarized in this study, even when the Ioannidis et al. correction is applied. We also note that there is good evidence from other jurisdictions of a positive effect of vocational education programs, for example, Cale et al. (2018).

Overall we conclude that for basic education programs and post-secondary education programs there is strong, clear evidence that the programs reduce recidivism rates, and the reduction is large enough to have confidence that these program are cost-effective. There is also reasonably strong evidence that vocational educations programs reduce recidivism rates by an amount that ensures such programs are cost-effective, but the evidence is not complete, and some additional work specifically focused on vocational education program evaluation would be valuable.

A Appendix

Table A.1: Summary of estimate data reconciliation

Paper	Effect size Davis et al.	Effect size current study	Current study data and reconciliation
Adams et al. (1994)	0.89 [0.77, 1.02] 0.85 [0.67, 1.08] 0.96 [0.88, 1.05]	0.89 [0.77, 1.02] 0.85 [0.67, 1.08] 0.96 [0.88, 1.05]	Both sets of results are based on the Table 2 data in the original study. The results match. The measure is return to prison rates between 14 and 36 months.
Allen (2006)	0.91 [0.89, 0.92] 1.17 [0.91, 1.51]	1.15 [0.21, 6.36] 2.90 [0.53, 16.03]	Davis et al. results are in conflict with the paper conclusion. We use the two stage logit model estimates from Table 8 for reconfined. These estimates are consistent with the author conclusion of no statistically significant results.
Anderson (1981)	0.37 [0.21, 0.67]	0.37 [0.20, 0.66] 0.40 [0.22, 0.75] 0.22 [0.11, 0.46]	Davis et al. use a pooled estimate for participate and complete, which is a match. It is possible to calculate participate and complete measure separately. Return to jail for parole violation within/up to 24 months. Values based on values in the text.
Anderson (1991)	0.69 [0.49, 0.97]	0.76 [0.58, 0.99]	It is not possible to determine the basis for the reported recidivism rates in Davis et al. Here the Chi-square statistic and associated p-value reported in Table 2, and the proportion in each group are used. Return to jail within 12 months.
Anderson (1995)	0.92 [0.85 ,1.00]	1.08 [0.90, 1.30] 0.67 [0.40, 1.09] 0.84 [0.67, 1.07] 0.94 [0.63, 1.42] 0.81 [0.68, 0.97] 0.86 [0.60, 1.24] 0.88 [0.66, 1.19]	Participation numbers are from Table 3, and Recidivism rates from Table 5. The difference is due to pooling. Here different education programs types and participate and complete are treated as separate groups.

Table A.1: Summary of estimate data reconciliation

Paper	Effect size Davis et al.	Effect size current study	Current study data and reconciliation
Batiuk (2005)	0.98 [0.68, 1.42] 0.84 [0.71, 1.01] 0.38 [0.33, 0.43] 0.81 [0.67, 0.99]	0.98 [0.68, 1.42] 0.84 [0.71, 1.01] 0.38 [0.33, 0.43] 0.81 [0.67, 0.99]	Both sets of values agree and are derived from the regression results reported in Table 3.
Blackburn (1981)	0.42 [0.28, 0.64]	0.43 [0.28, 0.65]	Values match.
Blackhawk (1996)	1.05 [0.49, 2.26]	0.73 [0.30, 1.81] 1.46 [0.62, 3.42]	Davis et al. pool the data to create a single group. We separate participate and complete. Pooled values match with Davis et al.
Brewster (2002)	0.73 [0.65, 0.82] 1.24 [1.12, 1.39]	0.73 [0.65, 0.82] 1.24 [1.12, 1.39]	The values are taken from the regression results Table 2 and 3, Model 3, and match.
Burke (2001)	0.37 [0.10, 1.35]	0.36 [0.01, 1.31] 0.31 [0.11, 0.89] 0.24 [0.08, 0.69] 0.36 [0.13, 1.01] 0.40 [0.14, 1.14]	Davis et al. values are one year post release. Data is reported for multiple years and we use the data for each year.
Castellano (1996)	0.28 [0.19, 0.41]	NA	Not a relevant program evaluation. Data not used.
Clark (1991)	0.45 [0.34 ,0.59]	0.45 [0.28, 0.71] 0.33 [0.21, 0.52] 0.86 [0.41, 1.83] 0.18 [0.02, 1.95]	Davis et al. pool multiple release year data. We use the data for each year separately.
Coffey (1983)	1.20 [0.71, 2.01]	0.69 [0.39, 1.22] 1.22 [0.70, 2.14]	Davis et al. pool multiple release year data. We use the data for each year separately.
Cronin (2011)	0.69 [0.65, 0.75]	0.73 [0.67, 0.79] 0.85 [0.77, 0.94]	Davis et al. pool participate (some progress and complete). We calculate comparisons separately.

Table A.1: Summary of estimate data reconciliation

Paper	Effect size Davis et al.	Effect size current study	Current study data and reconciliation
Davis (1986)	1.25 [1.12, 1.38]	1.13 [0.98, 1.29] 1.22 [1.10, 1.35] 1.23 [1.12, 1.35] 1.18 [1.07, 1.29] 1.15 [1.05, 1.26]	Davis et al. values are two years post release. Data is reported for multiple time periods and we use the data for each reporting period.
Dickman (1987)	0.66 [0.48, 0.92]	0.59 [0.40, 0.86] 0.66 [0.45, 0.97]	Davis et al. use data for the end of year two only. We use the data reported for both years.
Downes (1989)	1.24 [0.49, 3.13]	1.48 [0.47, 4.62]	We restrict the sample to the data on completed parole successfully or unsuccessfully, so have a slightly smaller sample relative to Davis et al., and hence greater uncertainty.
Gaither (1980)	0.18 [0.04, 0.78]	0.50 [0.11, 2.21]	There are two possible control groups. One defined by the author as a Non-Equivalent Control Group and one defined as the control group. We follow the primary study and use as the group defined as the Control in the paper. Values from Table viii and xi.
Gordon (2003)	0.03 [0.01, 0.05] 0.02 [0.01, 0.11]	0.28 [0.14, 0.56] 0.28 [0.06, 1.24]	The paper presents data in a confusing manner, but: (i) the values we use are consistent with the values used in the text of the paper; and (ii) the Davis et al. values look like outliers.
Harer (1995)	0.83 [0.70, 1.02]	0.83 [0.68, 1.02]	Both values are based on logistic regression model 2 results.
Holloway (1986)	0.72 [0.32, 1.60]	0.75 [0.33, 1.69]	Results are essentially the same.
Hopkins (1988)	0.38 [0.20 ,0.73]	0.33 [0.12, 0.95] 0.38 [0.18, 0.79] 0.37 [0.19, 0.72]	Davis et al. use data for the end of year two only. We use the data reported for all periods.
Hull (2000)	0.42 [0.35, 0.50] 0.40 [0.33, 0.49]	0.46 [0.33, 0.63] 0.38 [0.28, 0.52]	Very minor differences, due, we think, to using the participate but did not complete as the control group, or possibly rounding.

Table A.1: Summary of estimate data reconciliation

Paper	Effect size Davis et al.	Effect size current study	Current study data and reconciliation
Johnson (1984)	0.75 [0.57, 0.98]	0.88 [0.70, 1.12] 0.84 [0.66, 1.07]	Davis et al. use the total sample from Table 1 and Total recidivism rates from Table 16 to work out proportions. This gives stat. sig result which is in conflict with the discussion in the paper, eg Table 23 shows no effect from programs with controls. For both participate and complete vocational we use the change in R^2 from Table 23 to calculate an F-value, and then convert this to an odds ratio.
Kelso (1996)	0.42 [0.23, 0.74] 0.25 [0.11, 0.59]	0.54 [0.32, 0.90] 0.52 [0.31, 0.89] 0.44 [0.23, 0.87]	It has not been possible to identify exactly where the Davis et al. values are taken from, but here we use the values from Tables 2-5 to create total up to 5 year groups for those that complete v general population for high school, vocational, and post-secondary.
Langenbach (1990)	0.40 [0.25, 0.62]	0.40 [0.25, 0.62] 0.39 [0.28, 0.54] 0.38 [0.28, 0.52] 0.33 [0.24, 0.45] 0.31 [0.23, 0.43]	The difference is due to considering annual survival rates separately rather than as a pooled group.
Lattimore (1988)	0.58 [0.38, 0.89]	NA	Evaluation of a broader program. Data not used.
Lattimore (1990)	0.66 [0.40, 1.10]	0.52 [0.29, 0.95]	This is the same data as used in Lattimore (1988), however information on the pure education part of the program for complete vocation education is available and we use this data.

Table A.1: Summary of estimate data reconciliation

Paper	Effect size Davis et al.	Effect size current study	Current study data and reconciliation
Lichtenberger (2007)	0.63 [0.54, 0.74]	0.58 [0.42, 0.79] 0.63 [0.54, 0.74] 0.70 [0.63, 0.79] 0.73 [0.66, 0.82] 0.69 [0.61, 0.79] 0.61 [0.51, 0.72] 0.67 [0.50, 0.90]	Davis et al. use the year two data only. We use the data for each year.
Lichtenberger (2009)	0.79 [0.67, 0.93]	0.86 [0.80, 0.94]	Based on the discussion in the paper we use the total impact information from Table 10 not the logistic regression values from Table 4. This is consistent with the model presented in the paper.
Lichtenberger (2011)	0.80 [0.68, 0.95]	0.71 [0.56, 0.90] 0.77 [0.54, 1.11] 0.98 [0.73, 1.31]	Davis et al. use a pooled estimate, we separate out participate (two groups) and complete into separate categories.
Lockwood (1991)	0.68 [0.32, 1.42]	0.68 [0.32, 1.43]	Both values are based on Figure 2 in the paper.
Markley (1983)	1.00 [0.58, 1.74]	1.00 [0.57, 1.74]	Both values use the no difference ANOVA result to get pooled recidivism rates.
McGee (1997)	0.25 [0.19, 0.32]	0.25 [0.19, 0.32]	No difference. Both results are based on the information on p.vi.
Nally (2011)	0.27 [0.12, 0.59]	NA	The paper studies the impact of education level on recidivism, not the impact of prison education and vocational programs on recidivism. Data not used.
New York (1992)	0.80 [0.75, 0.86] 0.45 [0.34, 0.59]	0.80 [0.75, 0.86] 0.45 [0.34, 0.59]	Both values are based on the proportions data.
Nuttall (2003)	0.80 [0.73, 0.88]	0.81 [0.73, 0.89]	Both values are from the proportions data for the comparison No GED on entry and gain GED relative to did not gain GED.
O'Neil (1990)	0.31 [0.11, 0.89]	0.31 [0.11, 0.89]	Both values rely in the same proportions data.

Table A.1: Summary of estimate data reconciliation

Paper	Effect size Davis et al.	Effect size current study	Current study data and reconciliation
Piehl (1995)	0.66 [0.52, 0.84] 0.60 [0.37, 0.98]	0.76 [0.63, 0.92] 0.78 [0.65, 0.94]	It has not been possible to identify how the Davis et al. values were determined. We use the regression results from Table 6, p. 51 for education alone, and pooled education and vocational education.
Ryan (2000)	0.48 [0.34, 0.67]	0.48 [0.34, 0.67]	Both values rely in the same proportions data.
Saylor (1991)	0.67 [0.49, 0.93]	0.67 [0.49, 0.93]	Both values rely on the Table 1 proportions data.
Schumacker (1990)	0.79 [0.54, 1.14] 0.71 [0.43, 1.17] 0.63 [0.39, 1.04]	0.73 [0.49, 1.08] 0.68 [0.40, 1.17] 0.60 [0.36, 1.02]	Both sets of values appear to be based on sample size information from Table 1 and the proportions data from Table 3. There are minor differences in the values only.
Smith (2005)	1.33 [0.71, 2.52] 1.45 [0.66, 3.18] 1.64 [0.77, 3.45] 0.84 [0.55, 1.28]	0.70 [0.50, 0.97]	Davis et al. use the raw proportions data, and we use the logistic regression results from Table 12 for rearrest.
Steurer (2003)	0.74 [0.54, 1.01] 0.61 [0.44, 0.84] 0.70 [0.54, 0.90]	0.76 [0.57, 1.01] 0.61 [0.44, 0.85] 0.70 [0.52, 0.95]	Both sets of values use the re-incarceration data for the three locations.
Torre (2005)	0.20 [0.12, 0.31]	0.20 [0.12, 0.31]	Both values rely on the re-incarceration data.
Van Stelle (1998)	0.41 [0.18, 0.91]	NA	An evaluation of a broader rehabilitation program. Data not used.
Werholtz (2003)	0.93 [0.81, 1.06]	0.74 [0.64, 0.86] 0.93 [0.81, 1.06]	Davis et al. use a pooled participate and complete group. We include separate measures for each group.
Winterfield (2009	0.45 [0.21, 0.95] 0.44 [0.25, 0.78] 0.80 [0.63, 1.01]	0.46 [0.21, 1.00] 0.46 [0.24, 0.88] 0.37 [0.15, 0.91]	Davis et al. base the values on proportions data, and we use the values from the logistic regression models.

Table A.1: Summary of estimate data reconciliation

Paper	Effect size Davis et al.	Effect size current study	Current study data and reconciliation
Zgoba (2008)	0.59 [0.39, 0.88]	0.49 [0.34, 0.71]	Davis et al. use the raw proportions data, we use the change in R^2 from Table 3, to calculate F-value, and convert this to an odds ratio.

References

Adams, K., Bennett, K. J., Flanagan, T. J., Marquart, J. W., Cuvelier, S. J., Fritsch, E., . . . , Burton, V. S. (1994). A Large-Scale Multidimensional Test of the Effect of Prison Education Programs on Offenders' Behavior. *Prison Journal*, 74, 433-449.

Allen, R. (2006). An Economic Analysis of Prison Education Programs and Recidivism. Department of Economics, Emory University. Atlanta, GA.

Anderson, D. B. (1981). The Relationship Between Correctional Education and Parole Success. *Journal of Offender Counselling, Services and Rehabilitation*, 5(3/4), 13-25.

Anderson, D. B. (1991). Releasee Characteristics and Parole Success. *Journal of Offender Rehabilitation*, 17(1-2), 133-145.

Anderson, S. V. (1995). Evaluation of the Impact of Correctional Education Programs on Recidivism. Columbus, Ohio: Ohio Department of Rehabilitation and Correction.

Aos, S., Miller, M. G., & Drake, E. (2006). Evidence-based adult corrections programs: What works and what does not. Washington State Institute of Public Policy.

Batiuk, M. E., Lahm, K. F., McKeever, M., Wilcox, N., & Wilcox, P. (2005). Disentangling the Effects of Correctional Education: Are current policies misguided? An event history analysis. *Criminal Justice*, 5(1), 55-74.

Blackburn, F. S. (1979/1981). The Relationship Between Recidivism and Participation in a Community College Associate of Arts Degree Program for Incarcerated Offenders. Ann Arbor, MI: Virginia Polytechnic Institute and State University.

Blackhawk Technical College. (1996). RECAP (Rock County Education and Criminal Addictions Program) Program Manual Prepared to be of Assistance in Program Replication. Janesville, WI: Blackhawk Technical College.

Bozick, R., Steele, J., Davis, L. & Turner, S. (2018). Does providing inmates with education improve postrelease outcomes A meta-analysis of correctional education programs in the United States. *Journal of Experimental Criminology*, 1-40.

Brewster, D. R., & Sharp, S. F. (2002). Educational Programs and Recidivism in Oklahoma: Another Look. *Prison Journal*, 82, 314-334.

Burke, L. O., & Vivian, J. E. (2001). The Effect of College Programming on Recidivism Rates at the Hampden County House of Correction: A 5-Year Study. *Journal of Correctional Education*, 52(4), 160-162.

Cale, J., Day, A., Casey, S., Bright, D., Wodak, J., Giles, M., & Baldry, E. (2018). Australian prison vocational education and training and returns to custody among male and female ex-prisoners: A cross-jurisdictional study. *Australian & New Zealand Journal of Criminology*, 1-14.

Castellano, T. C. (1996). The Implementation and Impact of Illinois' Pre-Start Program: A Final Report. Carbondale, IL: Center for the Study of Crime, Delinquency and Corrections, Southern Illinois University.

Chappell, C. A. (2003). Post-Secondary Correctional Education and Recidivism: A Meta-Analysis of Research Conducted 1990-1999. University of Cincinnati.

Clark, D. D. (1991). Analysis of Return Rates of the Inmate College Program Participants. Albany, NY: New York State Department of Correctional Services.

Coffey, B. B. (1983). The Effectiveness of Vocational Education in Kentucky's Correctional Institutions: As Measured by Employment Status and Recidivism. (Doctor of Education), University of Kentucky, Lexington, KY.

Cronin, J. (2011). The Path to Successful Reentry: The Relationship Between Correctional Education, Employment and Recidivism (15-2011). Columbia, MO: Institute of Public Policy, University of Missouri.

Davis, L. M., Bozick, R., Steele, J. L., Saunders, J., & Miles, J. N. V. (2013). Evaluating the Effectiveness of Correctional Education: A Meta-Analysis of Programs That Provide Education to Incarcerated Adults.

Davis, S., & Chown, B. (1986). Recidivism Among Offenders Incarcerated by the Oklahoma Department of Corrections Who Received Vocational-Technical Training: A Survival Data Analysis of Offenders Released January 1982 Through July 1986. Oklahoma City, OK: Oklahoma State Department of Corrections.

Dickman, C. (1987). Academic Program Participation and Prisoner Outcomes. Lansing, MI: Facilities Research and Evaluation Unit, Michigan Department of Corrections Program Bureau.

Downes, E., Monaco, K., & Schreiber, S. (1989). Evaluating the Effects of Vocational Education on Inmates: A Research Model and Preliminary Results. In S. Duguid (Ed.), *The Yearbook of Correctional Education* (pp. 249-262). British Columbia, Canada: Simon Fraser University.

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). robumeta: Robust Variance Meta-Regression. R package version 2.0. https://CRAN.R-project.org/package=robumeta

Gaither, C. C. (1980). An Evaluation of the Texas Department of Corrections‘ Junior College Program. (Master of Arts), Northeast Louisiana University, Monroe, LA.

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, 9(6), 641-651.

Gerber, J., & Fritsch, E. J. (1995). Adult Academic and Vocational Correctional Education Programs: A Review of Recent Research. *Journal of Offender Rehabilitation*, 22(1-2), 119-142.

Gordon, H. R. D., & Weldon, B. (2003). The Impact of Career and Technical Education Programs on Adult Offenders: Learning Behind Bars. *Journal of Correctional Education*, 54(4), 200-209.

Greene, W. H. (2000). *Econometric Analysis*, (Fifth ed.) Boston: Pearson Education.

Harer, M. D. (1995). Prison Education Program Participation and Recidivism: A Test of the Normalization Hypothesis. Washington, DC: Federal Bureau of Prisons, Office of Research and Evaluation.

Hedges, L.V., Tipton, E. & Johnson, M.C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39-65.

Holloway, J., & Moke, P. (1986). Post Secondary Correctional Education: An Evaluation of Parolee Performance. Wilmington, OH: Wilmington College.

Hopkins, A. J. (1988). Offender Recidivism Report. Baltimore, MD: State of Maryland, Department of Public Safety and Correctional Services.

Hull, A., Forrester, S., Brown, J., Jobe, D., & McCullen, C. (2000). Analysis of Recidivism Rates for Participants of the Academic/Vocational/Transition Education Programs Offered by the Virginia Department of Correctional Education. *Journal of Correctional Education*, 51(2), 256-261.

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *Economic Journal*, 127(605), F236-F265.

Johnson, C. M. (1984). The Effects of Prison Labor Programs on Post-Release Employment and Recidivism. (Doctor of Philosophy), Florida State University, Tallahassee, FL.

Kansas Department of Corrections. (2003). Offender Programs Evaluation: Volume V. Topeka, KS: The Kansas Department of Corrections.

Kelso, C. (1996). A Study of the Recidivism of Garrett Heyns Education Center Graduates Released between 1985 and 1991. *Journal from the Northwest Center for the Study of Correctional Education*, 1(1), 25-44.

Laird, N. M., & Ware, J. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4), 963-974.

Langenbach, M., North, M. Y., Aagaard, L., & Chown, W. (1990). Televised Instruction in Oklahoma Prisons: A Study of Recidivism and Disciplinary Actions. *Journal of Correctional Education*, 41(2), 87-94.

Lattimore, P. K., Witte, A. D., & Baker, J. R. (1988). Sandhills Vocational Delivery System Experiment: An Examination of Correctional Program Implementation and Effectiveness. Washington, DC: National Institute of Justice.

Lattimore, P. K., Witte, A. D., & Baker, J. R. (1990). Experimental Assessment of the Effect of Vocational Training on Youthful Property Offenders. *Evaluation Review*, 14(2), 115-133.

Lichtenberger, E. (2007). The Impact of Vocational Programs on Post-Release Outcomes for Full Completers from the Fiscal Year 1999, 2000, 2001, and 2002 Release Cohorts.

Lichtenberger, E. (2009). Direct and Indirect Impacts of Career and Technical Education on Post-Release Outcomes.

Lichtenberger, E. (2011). Measuring the Effects of the Level of Participation in Prison-Based Career and Technical Education Programs on Recidivism.

Lipton, D., Martinson, R., & Wilks, J. (1975). *The Effectiveness of Correctional Treatment: A Survey of Treatment Evaluation Studies.* New York: Praeger.

Lockwood, D. (1991). Prison Higher Education and Recidivism: A Program Evaluation. In S. Duguid (Ed.), *The Yearbook of Correctional Education* (pp. 187-201). British Columbia, Canada: Simon Fraser University.

MacKenzie, D. (2006). *What Works in Corrections: Reducing the Criminal Activities of Offenders and Delinquents.* New York: Cambridge University Press.

MacKenzie, D. L., & Hickman, L. J. (1998). What Works in Corrections? An Examination of the Effectiveness of the Type of Rehabilitation Programs offered by Washington State Department of Corrections. College Park, MD.

Markley, H., Flynn, K., & Bercaw-Dooen, S. (1983). Offender Skills Training and Employment Success: An Evaluation of Outcomes. *Corrective and Social Psychiatry and Journal of Behavior Technology Methods and Therapy*, 29(1), 1-11.

McGee, C. J. (1997). The Positive Impact of Corrections Education on Recidivism and Employment. Springfield, IL: Illinois Department of Corrections and Illinois Council on Vocational Education.

Nally, J., Lockwood, S., Ho, T., & Knutson, K. (2011). The Effect of Correctional Education on Postrelease Employment and Recidivism: A 5-Year Follow-Up Study in the State of Indiana. Working Paper.

Nuttall, J., Hollmen, L., & Staley, E. M. (2003). The Effect of Earning a GED on recidivism rates. *Journal of Correctional Education*, 54(3), 90-94.

O'Neil, M. (1990). Correctional Higher Education: Reduced Recidivism? *Journal of Correctional Education*, 41(1), 28-31.

Piehl, A. M. (1994). Learning While Doing Time. Working Paper. Malcom Wiener Center for Social Policy. Cambridge, MA: John F. Kennedy School of Government, Harvard University.

Re, D.A.C. (2013). compute.es: Compute Effect Sizes. R package version 0.2-2. URL http://cran.r-project.org/web/packages/compute.es

Ryan, T. P., & Desuta, J. F. (2000). A Comparison of Recidivism Rates for Operation Outward Reach (OOR) Participants and Control Groups of Non-Participants for the Years 1990 Through 1994. *Journal of Correctional Education*, 51(4), 316-319.

Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, 2(6), 110-114.

Saylor, W. G., & Gaes, G. G. (1996). The effect of prison employment and vocational/apprenticeship training on long-term recidivism. *Forum on Corrections Research*, 8(1), 12-14.

Schumaker, R. E., Anderson, D. B., & Sara, L. (1990). Vocational and Academic Indicators of Parole Success. *Journal of Correctional Education*, 41(1), 8-13.

Smith, L. G. (2005). Pennsylvania Department of Corrections Education Outcome Study. Lanham, MD: Correctional Education Association. State of New York Department of Correctional Services. (1992). Overview of Department Follow-up Research on Return Rates of Participant in Major Programs 1992. Albany, NY: Division of Program Planning, Research and Evaluation.

Stanley, T.D. & Doucouliagos, C.(H.). (2014). Meta-regression approximations to reduce publication selection bias, *Research Synthesis Methods*, 5(1), 60–78

State of Washington Department of Corrections. (1998). Evaluation of the Work Ethic Camp. Olympia, WA: Washington State Department of Corrections, Planning and Research Section.

Steurer, S. J., Smith, L., & Tracy, A. (2001). Three State Recidivism Study (report submitted to US Office of Correctional Education). Lanham, MD: Correctional Education Association.

Tipton, E. (2015). Small Sample Adjustments for Robust Variance Estimation with Meta-regression. *Psychological Methods*, 20(3), 375-393.

Torre, A., & Fine, M. (2005). Bar None: Extending Affirmative Action to Higher Education in Prison. *Journal of Social Issues*, 61(3), 569-594.

Van Stelle, K. R., Lidbury, J. R., & Moberg, D. P. (1995). Specialized Training and Employment Project (STEP), Wisconsin Department of Corrections, October 1, 1993 Through September 30, 1995. Madison, WI: University of Wisconsin-Madison Medical School, Department of Preventive Medicine, Center for Health Policy and Program Evaluation.

Wells, R. E. (2000). Education as Prison Reform: A Meta-Analysis. Louisiana State University, Baton Route, La.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4), 817-838.

Wilson, D. B., Gallagher, C. A., & MacKenzie, D. L. (2000). A meta-analysis of corrections-based education, vocation, and work programs for adult offenders. *Journal of Research in Crime and Delinquency*, 37(4), 347-368.

Winterfield, L., Coggeshall, M., Burke-Storer, M., Correa, V., & Tidd, S. (2009). The Effects of Postsecondary Correctional Education Final Report. Washington, DC: Urban Institute, Justice Policy Center.

Zgoba, K. M., Haugebrook, S., & Jenkins, K. (2008). The Influence of GED Obtainment on Inmate Release Outcome. *Criminal Justice and Behaviour*, 35, 375-387.