

A Rapid Evidence Assessment of the impact of probation caseloads on reducing recidivism and other probation outcomes

Probation Journal

1–21

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/02645505211025595

journals.sagepub.com/home/prb

Chris Fox , **Jordan Harrison**, **Grace Hothersall** 
and **Andrew Smith** 

Manchester Metropolitan University, UK

Russell Webster

Independent Consultant, UK

Abstract

We undertook a Rapid Evidence Assessment to explore the existing empirical evidence relating to the impact of probation caseloads on recidivism. Over 3,000 potentially relevant papers were sifted from which five were deemed robust enough to be analysed in detail. All five were US studies which examined the impact of particular initiatives to reduce caseloads and were delivered by mainstream community-based probation officers. All recorded reductions in measured outcomes compared to comparators. Overall, although the number of robust studies remains quite small for such a key area of consideration, there appears to be a growing body of evidence that lower probation caseloads have a positive impact in terms of reducing reoffending in the USA. All five studies looked at a range of criminal justice outcomes including technical probation violations, violations for new arrests and reconvictions. Interestingly, although researchers were expecting to find a higher rate of probation violations among the cohorts supervised by probation officers with lower caseloads (due to the increased intensity of supervision), this did not turn out to be the case.

Corresponding Author:

Chris Fox, Manchester Metropolitan University, Geoffrey Manton Building, Rosamond St West, Manchester M15 6LL, UK.

Email: c.fox@mmu.ac.uk

Keywords

caseload, probation, recidivism, engagement, breach, Rapid Evidence Assessment

Introduction

In this paper we report on a Rapid Evidence Assessment (REA) of the impact of probation caseloads on reducing recidivism and a number of 'intermediate outcomes' such as improving engagement, improving completions, reducing breaches and reducing staff absence. It is important to distinguish between (i) caseload, which equates to the number of cases handled in a given period by an organisation or by an individual, and (ii) workload, which is the amount of work allocated to an individual or a team. The focus of this REA was on the former and aspects of probation caseload we were interested in included managing caseloads, managing change in caseload volume and type and supporting probation staff with complex and/or high workloads. In particular we will look for evidence of how changes to caseloads and workload effect outcomes.

REAs are a form of Systematic Review, but are undertaken over a shorter period than a full review – approximately 3 months, rather than 12 months. REAs and Systematic Reviews systematically search for, evaluate and synthesise evidence about a specific intervention. Where possible, they include a statistical meta-analysis of individual studies, in order to provide a clear indication of the likely impact (effect size) of the intervention. The evidence which is eligible for synthesis in an REA is that which is trustworthy, its design being capable of supporting logical and, if possible, statistical inference about the causes of observed effects.

We first set out the context for the REA before describing the methodology adopted. We then set out findings from the REA and conclude with a discussion of these findings, including reference to wider debates about caseloads in comparable public services.

Context

The question of identifying optimum caseloads and workloads for probation staff has, of course, always been a thorny one as governments have consistently sought to reconcile the competing aims of maximum effectiveness and value for public money. McWilliams (1966) sought to attribute the decreasing proportion of probation orders which were successfully completed in London in the 1950s and 60s to an increase in average caseloads. Regrettably, the study is methodologically weak and seeks to correlate a 10 percentage point improvement in completion rate between 1959 and 1960 with a fall in average male officer caseloads from 60 to 50 cases in the same year. However, there is insufficient information provided to assess the validity of this claim which could be attributable to a number of other reasons or unreliable data. More methodologically rigorous was the Intensive Matched Probation and After-Care Treatment programme known as IMPACT (Folkard et al., 1976) probationers were randomly assigned to normal or 'intensive'

caseloads, and subsequent reconviction rates were compared. There was no statistically significant difference between the intervention and control group, although, as Raynor (2018) notes the only group which appeared to benefit from smaller caseloads was a fairly small number of offenders with high self-reported problems and low 'criminal tendencies', but these were not very representative of offenders in general. Raynor speculates that this might have been because officers were using in the extra time made available by lower caseloads to offer some form of counselling, which might have helped this group more than others.

Burrell (2006) provides an interesting overview of US researchers' attempts to link probation caseloads to offender outcomes. Writing for the American Probation and Parole Association, he notes that numerous evaluations of Intensive Supervision Probation (ISP) programmes, one of whose defining characteristics was small caseloads and which were widespread in the US in the 1980s, were 'uniformly dismal'. While the caseloads were small, and the officers had much more time to devote to supervision, the ISPs did not reduce reoffending. In many instances, the aggressive and rigid enforcement policies employed in these programmes resulted in more offenders being returned to court and sentenced to custody. Burrell concludes that the 'get tough' approach to community supervision which had no grounding in the evidence base was responsible for these poor results and demonstrates that reducing caseloads alone will not improve outcomes.

Different jurisdictions have very different models of probation. The most recent Council of Europe Annual Penal Statistics (Council of Europe, 2019) found that the ratio of probationers per individual staff member varied from 4.7 in Norway to 240.1 in Greece with an average (median) ratio of 32.8.

Probation staff have long been convinced that high caseloads affect both the reoffending rates of the probationers they supervise and their own well-being. A number of surveys in Canada and the US have covered the issue of caseloads. For example, a Canadian survey of 541 parole officers (Union of Safety and Justice Employees, 2019), including staff working both in custody and the community, is typical. Almost all those surveyed (94.2%) disagreed with the statement that: Caseload size or frequency of contact has no impact on public safety. Similarly, 86 percent of institutional parole officers and 87 percent of community parole officers in the same survey believed that their workload was affecting their psychological or physical health.

Gayman and colleagues (2018) surveyed 798 probation and parole officers working in North Carolina as part of a cross-sectional survey concerning job characteristics and well-being. The survey investigated the impact on those probation officers who had the highest number of probationers with mental health problems on their caseload. Depressive symptoms were measured using the Centre for Epidemiologic Studies Depression Scale and the level of emotional exhaustion using an established five item scale. Interestingly, the authors found that total caseload numbers were not as important as the number of supervisees with serious mental health problems for predicting depressive symptoms and emotional exhaustion amongst probation and parole officers. Neither the number of other mental health services received by supervisees nor probation officers'

training in mental health mitigated the link between the number of supervisees on an officer's caseload with mental health problems and that officer's emotional exhaustion.

Returning to England and Wales, HM Inspectorate of Probation (2021a) note that *Transforming Rehabilitation* has been a key driver in relation to changes in caseloads in recent years with probation trusts replaced by a National Probation Service (NPS) managing people at high risk of serious harm and Community Rehabilitation Companies (CRCs) managing low to medium risk service users. The result has been that staff no longer supervised 'mixed' caseloads in terms of risk levels. *Transforming Rehabilitation* also introduced (through the *Offender Rehabilitation Act (ORA) 2014*) a new duty on probation services to supervise prisoners released from short prison sentences of less than 12 months. This increased the number of post-release service users managed by probation in the community, climbing from just under 40,000 on 31st March 2015, to 68,863 on 31st March 2020, a rise of 74% (HM Inspectorate of Probation, 2021a). Prior to *Transforming Rehabilitation*, the caseload for the probation trusts had been falling year on year since 2000 (HM Inspectorate of Probation, 2021a).

HM Inspectorate of Probation (2021a) also identify the introduction of the suspended sentence order (SSO) in the *Criminal Justice Act 2003* (with the *Legal Aid, Sentencing and Punishment of Offenders Act 2012* introducing SSOs without requirements) as impacting on probation caseloads. They find that these orders have proved popular at the expense of community orders (COs) – those currently supervised under COs have declined by 31% since 2008, whereas SSOs have remained relatively stable (-2%). They conclude that:

The result of the *Transforming Rehabilitation* reforms combined with offending and sentencing trends have left probation services with a caseload of service users who have more complex needs, more entrenched offending attitudes, behaviours and lifestyles, and often higher levels of risk than before the ORA caseload commenced. (HM Inspectorate of Probation: 7)

The NPS use the online Workload Measurement Tool to monitor staff capacity. The Workload Measurement Tool takes account of attributable time (time spent managing offenders), non-attributable time (travel, IT problems, supervision, comfort breaks – 16 percent is assumed) and non-effective time (holidays, sickness, training – 20 percent is assumed). The current measure of excessive workload is where an officer has a WMT capacity of over 110 percent over a consecutive 4-week period. CRCs use different systems. In a recent annual report, HM Inspectorate of Probation (2019: 74) noted that 'pressures in the NPS and in CRCs are felt most keenly at probation officer level, where shortages are greatest'. They found probation officer workloads at 120–160% of capacity (as measured by the workload tool) for the NPS and wide variation among CRCs. At one end of the spectrum, Durham Tees Valley CRC had a stable and experienced workforce with manageable caseloads, whereas in Dorset, Devon and Cornwall CRC caseloads ranged

from 18 to 102 for probation officers and from 14 to 168 for probation service officers (HM Inspectorate of Probation, 2019).

The human price of high workloads was also emphasised by the Chief inspector in a recent speech when he reported that that:

the impact of [high caseloads] on some of the staff we spoke to was clear. Some were in tears as we spoke to them. Others spoke of being burnt out and of having to work evenings and weekends to keep their head above water. (Russell, 2020)

Some of these issues will be superseded by new reforms to the probation sector. The Ministry of Justice (2018) has announced that all offender management will move to the NPS to take effect in June 2021. However, in its report assessing the readiness of planning for the reunification of the probation service, HM Inspectorate of Probation (2021b) highlighted continuing concerns about the number of qualified probation officers in post, noting that despite a recruitment initiative, it would take several years for sufficient fully qualified probation officers to be in post.

Methodology

Scope

The aim of the Rapid Evidence Assessment was to synthesise robust UK and international evidence on probation caseloads including managing caseloads, managing change in caseload volume and type in probation settings, and supporting probation staff with complex and/or high workloads. In particular we looked for evidence of how changes to caseloads and workload effected outcomes. The ultimate outcome of interest was reducing recidivism, but a range of 'intermediate outcomes' were also considered including, improving engagement, improving completions, reducing breaches and staff absence. Based on this aim we defined the scope of the REA using the PICOS acronym (Campbell Collaboration, 2014).

Population: Only studies involving participants who are under probation supervision were included. Offenders released from prison under probation supervision and offenders given a community sentence where probation supervision was a component were eligible for inclusion. Since offenders in England and Wales under the age of 18 are in the care of youth offending services, only studies where some or all of the participants were aged 18 or over were included.

Intervention: Studies about understanding and managing caseloads, managing change in caseload volume and type in probation settings and supporting staff with complex and/or high probation workloads were included.

Comparison involved: We sought studies where changes in the volume, complexity, type or management of caseloads were compared to a control, most likely 'business as usual'. However, to draw in as wide a range of studies as possible we took a broad view of possible comparisons.

Outcomes: Ideally we sought studies where the primary outcome was a measure of recidivism such as arrests, convictions (binary, frequency, severity), or breaches of condition (eg recalls to custody or return to court). However, anticipating that studies

Table 1. Maryland Scientific Methods Scale adapted for reconviction studies.

Level	Comparison	Description	Methods
Level 1	No comparison	Reconviction rate is reported for intervention group only	Before and after study
Level 2	Comparison with predicted rate	Actual and expected reconviction rates of intervention group are compared	Expected reconviction rates generated by Offender Group Reconviction Scale (OGRS)
Level 3	Unmatched comparison group	Reconviction rate of intervention group is compared with reconviction rate of an unmatched comparison group	Comparison of mean levels of reoffending
Level 4	Well-matched comparison group	Reconviction rate of intervention group is compared with reconviction rate of a comparison group matched on static (and dynamic) risk factors e.g. criminal history, gender	Propensity score matching; regression discontinuity
Level 5	Randomised control trial (RCT)	Reconviction rates are compared between intervention and control groups that have been created through random assignment	Randomisation

such as these would be relatively rare we also looked for a range of 'intermediate outcomes' including, engagement in probation supervision, completions of community sentences or license conditions, breaches and staff absence, sickness or turnover.

Study designs: Traditionally an REA focuses on counterfactual impact evaluations and where these were found they were preferred. To ensure reasonable levels of internal validity we preferred studies of levels 3 to 5 on the Maryland Scale adapted for reconviction studies (Friendship et al., 2005: 7) as set out in Table 1.

Studies in scope were published in English since 2000.

Search strategy

We used a 4-step search strategy resulting in the identification of 22 unique papers.

First, we developed the following Boolean search string:

(TITLE-ABS-KEY ((offender* OR probation* OR licen* OR 'service user*' OR parol* OR supervis* OR 'case manage*' OR practition* OR 'corrections officer' OR correct*) AND (staff OR 'case work*' OR caseload OR workload OR 'case type' OR capacity OR stress OR 'human resources' OR 'organisational culture') AND (reoffen* OR offend* OR recidiv* OR rearrest* OR reconvict* OR incarceration OR desist*) AND (evaluation OR experiment* OR trial OR impact OR effect*)) AND PUBYEAR > 2000)

Using this search string we searched the following electronic databases: ASSIA (Applied Social Sciences Index and Abstracts); Criminal Justice Database; PsycINFO; Scopus; Sociological Abstracts; and Web of Science. This process identified

5,633 papers, which, after being put into EPPI Reviewer where 2,444 duplicates were removed.

Secondly, the websites of the following governmental agencies and organisations associated with criminal justice research were searched for reports and other grey literature:

- UK Ministry of Justice
- The Scottish Government
- Correctional Services Canada
- Australian Institute of Criminology
- US National Institute of Corrections
- The Nuffield Foundation (UK)
- Vera Institute for Justice (US)
- Washington State Institute for Public Policy (US)
- The Urban Institute (US).

Although a considerable number of initial 'hits' were generated, closer scrutiny only identified 5 papers that appeared, on initial inspection to be potentially within scope and that were taken forward to the screening stage.

Thirdly, we hand searched a number of key journals back to 2000:

- *Probation Journal*
- *European Probation Journal*
- *International Journal of Offender Therapy and Comparative Criminology*
- *Journal of Forensic Practice*.

One additional paper was identified that was taken forward for initial screening.

Fourthly, via www.russellwebster.com we put out a call¹ for relevant papers, with a focus on studies that had not been published in the academic or grey literature. The call generated a number of responses and suggestions of papers. Seven papers identified were taken forward for initial screening. Most suggested papers were either individual or organisational contributions to policy debate or were studies that fell outside of the PICOS criteria, most commonly because although they touched on issues of workload, they did not examine the impact of caseloads through empirical study.

Screening studies

The titles and abstracts of studies identified during the search were downloaded into EPPI Reviewer where duplicates were removed. This resulted in a total of 3,202 papers retained for screening.

The title and abstract of each identified paper was then screened for relevance using the PICOS criteria (above). Titles and abstracts were screened by one reviewer and a second reviewer screened 25% at random. Disagreements between reviewers were resolved through discussion and the involvement of a third reviewer where necessary.

A total of 3,180 papers were excluded. Some papers were excluded for multiple reasons. We recorded the primary reason for exclusion and initially focused on whether the paper focused on the target population and included a relevant intervention. On this basis 1,087 papers did not focus on the relevant target group and 2,062 did not focus on a relevant intervention. A further 31 were excluded because they did not report on a relevant research design.

At the end of the title and abstract screening process 22 papers were assessed as potentially eligible and full versions of all these papers were retrieved. Full text papers were then screened again for relevance using PICOS. Five papers were retained for analysis. The main reasons for discarding papers were that they were not evaluations. For example, several papers were based on surveys or interview programmes that did not have an evaluative dimension (eg Council of Europe, 2019; DeMichele, 2007; Phillips et al., 2016). The five retained papers are described in more detail in Table 2.

Retained papers were then screened for methodological rigour. The five retained papers were all evaluations involving a counterfactual and at least some quantitative outcome measures.

The overall process of identifying, screening and assessing papers for relevance is summarised in the PRISMA flow diagram in Figure 1.

Findings

In this section we describe the studies retained for analysis, which are summarised in Table 2.

Description of studies

All five studies were evaluations of interventions where reduced caseload was a feature of the intervention being tested.

Taxman and colleagues (2006) evaluated Maryland's Proactive Community Supervision (PCS) programme in which moderate and high risk probationers and parolees were supervised in reduced caseloads of 55 (compared with the usual caseload of 100), using an evidence-based model of intervention. The evaluation, conducted with a high degree of methodological rigour, included 274 randomly selected cases for PCS, matched with 274 cases supervised under the traditional model (non-PCS). The PCS cases had significantly lower re-arrest rates (32.1% for PCS vs. 40.9% for non-PCS) and significantly lower technical violation rates (20.1% for PCS vs. 29.2% for non-PCS). The PCS offenders had a 38 percent lower chance of being rearrested or being charged with a technical violation, as compared with the non-PCS offenders. These findings were found to be true regardless of the criminal history of the offender. The researchers note that these positive outcomes were found in Baltimore, one of the jurisdictions covered by the initiative, even though the city had heightened law enforcement activity during the study period, which should have meant that probationers/parolees under supervision who were involved in criminal activities would have had an increased likelihood of arrest.

Table 2. Summary of retained papers.

Study	What was done	Outcome
Cox et al. (2005)	Compared two intensive probation programmes in Connecticut: the Probation Transition Programme (PTP) and the Technical Violation Unit (TVU). Both involved reduced caseloads compared against mainstream probation supervision. Both programmes designed to increase supervision completion rates and reduce reoffending via evidence-based models of intervention with low caseloads.	Probation violation rates lower for intervention group
Jalbert et al. (2011)	Evaluation of intensive probation programmes in Polk County, Iowa and Oklahoma City. Jurisdictions specifically chosen because they applied Evidence Based Practice (EBP). Intervention programmes involved deliberately reduced caseloads and were compared to mainstream probation supervision to test whether reduced caseloads improved probation outcomes.	Lower rate of reoffending for intervention group
Manchak et al. (2014)	Compared specialist mental health probation teams to mainstream probation provision. Reduced caseload was a critical component of the specialist provision, alongside other differences including specialist mental health probation officers having additional training and adopting a different approach to the people they supervised.	Probation violation rates lower for intervention group
Taxman et al. (2006)	Compared Maryland's Proactive Community Supervision (PCS) programme, a form of intensive probation involving reduced caseloads against mainstream probation supervision to test whether reduced caseloads improved probation outcomes in jurisdictions specifically chosen because they applied Evidence Based Practice (EBP).	Lower rate of re-arrests & probation violation rates for intervention group
Wolff et al. (2014)	Compared specialist mental health probation teams to mainstream probation provision. Reduced caseload was a critical component of the specialist provision, alongside other differences including specialist mental health probation officers having additional training and adopting a different approach to the people they supervised.	The average number of jail days reduced by a greater proportion for the intervention group

A study by Cox and colleagues (2005) examined two programmes in Connecticut which were designed to reduce probation violations and subsequent incarceration,. The evidence-based model of intervention included low caseloads – a prescribed maximum of 25 compared to the local average of 100. The Probation Transition Programme (PTP) targeted inmates who had terms of probation upon discharge from a correctional facility, halfway house, parole, transitional

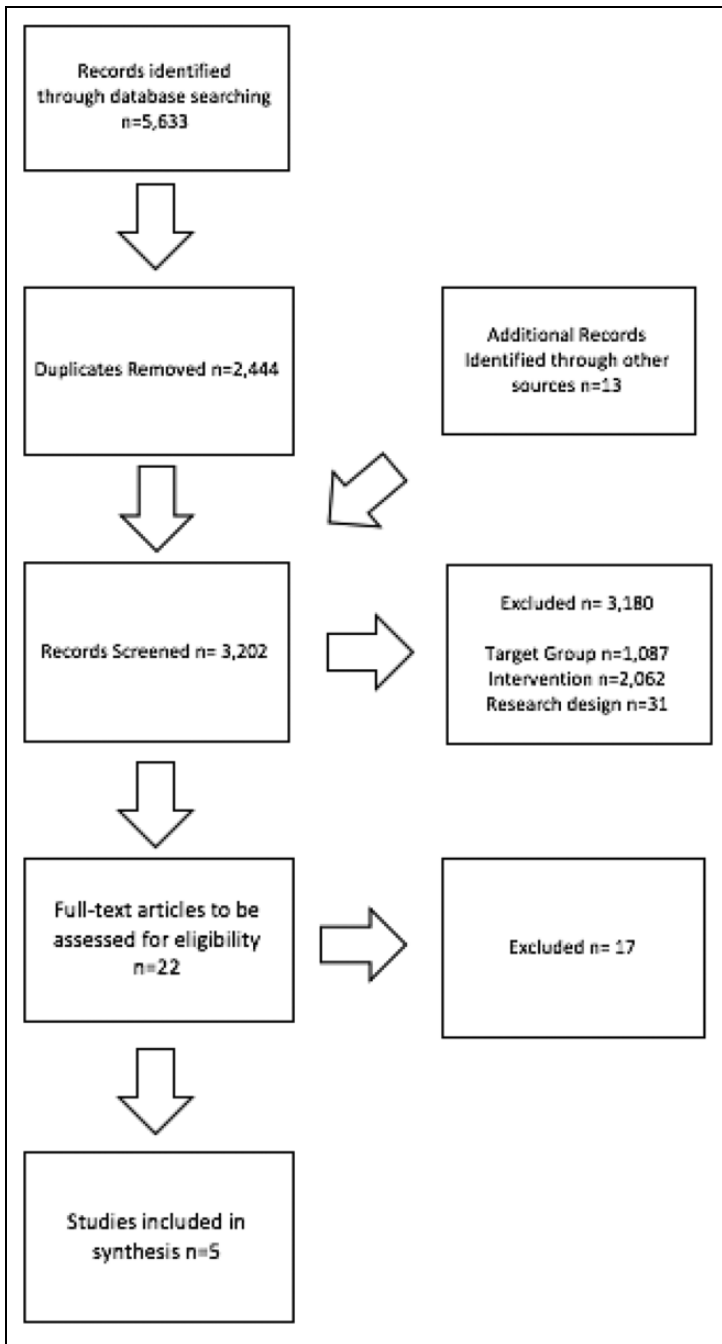


Figure 1. PRISMA flow diagram.

supervision or a furlough and aimed to increase the likelihood of a successful probation period for split sentence probationers by reducing the number and intensity of technical violations during the initial period of probation. The Technical Violation Unit (TVU) also focused on probationers about to be violated for technical reasons (e.g., deliberate or repeated non-compliance with court ordered conditions, reporting requirements, and service treatment requirements). However, it addressed all probationers regardless of whether they had been incarcerated or not. Although the evaluation findings for the two programmes were positive, they were restricted to a short time period – just the first four months on probation and did not extend to analysis of formal reconviction rates. The PTP probation violation rates were lower than the PTP comparison group during this four month period (8% for PTP and 13% for the PTP comparison group). PTP probationers were violated at similar rates for new arrests (3%), technical violations (3%), and both new arrests and technical violations (2%) while the PTP comparison group had a slightly higher rate of technical violations (5%) and both new arrests and technical violations (5%). The TVU group had a violation rate of 30 percent. However, since these offenders were only referred to the programme because they had already demonstrated poor compliance and were on the point of being violated, the researchers argue that this 30 percent should be compared with an expected violation rate of 100 percent.

Jalbert and colleagues (2011) set out to test whether reduced caseloads improve probation outcomes in areas where Evidence-Based Practice (EBP) applied. The researchers purposefully selected probation areas where EBP was not only claimed to be in operation but where there was clear evidence that more resources were allocated to the supervision of high-risk offenders. Polk County, Iowa was an area where probation officers who supervise intensive supervision probation (ISP) programmes had smaller caseloads (but equivalent workloads) to probation staff who supervised offenders under high-normal supervision. The research team implemented a regression discontinuity design study (RDD) (also reported separately in Jalbert et al., 2011). They estimated that ISP caseloads allowed probation officers to spend about 1.7 hours on offenders supervised on ISP for every one hour spent on offenders supervised under high-normal caseloads. However, ISP was also demonstrably different from high-normal supervision with respect to amount of contact and rehabilitative interventions, so size of caseload was not the only difference. In Polk County, the researchers found that probationers supervised by officers with reduced caseloads had a lower rate of arrests for new crimes. When the follow-up period was limited to 6 months, ISP reduced the likelihood of criminal recidivism by 25.5 percent for all offences; 39.4 percent for drugs, property, and violent offences; and 45 percent for property and violent offenses (drug offenses excluded). They concluded that ISP therefore reduced recidivism when compared to normal supervision. For longer periods of time, they also found that recidivism was reduced significantly for property and violent crimes (37 percent at 18 months and 30 months, respectively). However, the researchers were concerned that the analysis was only able to identify very large treatment effects, despite the fact that the samples being analysed were large. This suggests that results should be treated with caution. It is possible that more intensive supervision resulting from lower caseloads

leads to increased revocations for technical violations, but the study found no strong evidence of this. The researchers therefore concluded that reduced caseloads in this context probably reduced criminal recidivism and probably did not increase revocations for technical violations.

The research team also worked with probation in Oklahoma City (a locality that implemented evidence-based supervision practices in probation) to manipulate work assignments so that some probation officers had caseloads averaging 54 probationers per officer while other probation officers maintained caseloads averaging 106 probationers per officer. The team found (also published separately in Jalbert and Rhodes, 2012) that there were few significant differences between the characteristics of probationers supervised by probation officers with reduced caseloads and probationers supervised by probation officers with regular caseloads. The study, which was originally designed as an RCT, degenerated as many of the control probation officers moved into other roles and the study team turned to a difference in differences (DD) design. The study team found that probation officers with smaller caseloads made more frequent supervision contact and the probationers supervised by these officers were more likely to receive correctional interventions. They used survival analysis to estimate that the smaller caseload reduced the rate of recidivism by roughly 30 percent, while technical violations increased by 4 percent. The team concluded that reduced caseloads in agencies using modern supervision practices reduce recidivism.

Two US studies compare specialist mental health probation teams to mainstream supervision. Five key features distinguish specialty mental health supervision from mainstream supervision:

- Relatively small caseloads (an average of less than 50 compared to usual caseloads of more than 100)
- Specialty officers have mental health training
- Specialty officers actively coordinating and integrating probation work and treatment resources (as opposed to brokering them)
- Traditional offices tend to rely on threats and sanctions, specialty officers emphasise problem-solving.
- Specialty officers try to establish firm, fair and caring relationships (authoritative, not authoritarian).

Manchak et al.'s (2014) study employed a longitudinal multi-method and multi-measure design in which 176 probationers on traditional probation supervision were matched with 183 probationers on specialty mental health supervision. Probationers were interviewed at three time points over the course of one year and the supervising officers completed a brief survey within the same time schedule. Researchers analysed probation and court records for information about violations. The speciality programme site was selected from an earlier national survey of these mental health programmes by Skeem and Eno (2006). The traditional agency (mainstream probation) site was selected because it matched the speciality site in terms of jurisdiction size, urban location, probation demographic characteristics

(gender, age and race) and county mental health expenditure. Importantly, at the midpoint of data collection, average caseload sizes for the specialty and traditional offices were approximately 50 and 100 probationers, respectively. The researchers found a large effect (OR = 2.19) was observed favouring specialty probation in probation violations. Probationers on specialty probation were approximately two times less likely than those on traditional probation to have a formal violation report filed against them.

In a similar study, Wolff and her colleagues (2014) set out to measure whether specialised mental health caseloads were effective in terms of criminal justice, mental health and community engagement outcomes. The researchers assessed the impact of three different teams with caseload the main differentiator between the teams. In addition to analysing official probation records, researchers undertook 103 interviews with service users from one of these teams. The three teams were:

- A newly established Specialised Mental Health Caseload (SMHC) supervising no more than 30 clients per officer (n = 1367) labelled 'Grant'. A sub-group of this sample, comprised the 103 probationers who were interviewed – 'Research';
- An established SMHC supervising roughly 50 clients per officer (n = 495) labelled 'Pilot'; and
- A traditional caseload of clients receiving mental health treatment and supervised by officers with average caseloads of over 130 clients (n = 5453) 'Traditional'.

Findings supported the effectiveness of the specialty teams:

- Clients with mental illnesses assigned to the SMHC had significantly fewer violations of probation resulting in arrests and fewer jail days in the six months post-assignment to the SMHC.
- The average number of jail days for the grant and pilot caseloads, as a whole, significantly decreased in the six months post-assignment, from 4 days to 1 day and from 6 days to 2 days respectively. Similar but smaller declines were observed for the traditional caseload.
- Clients with mental illnesses assigned to the SMHC were found to have improved mental health outcomes six months post-assignment to the SMHC.

Although both Manchak et al. (2014) and Wolff et al. (2014) believed that the significantly smaller caseloads were a key element of the successful operation of these mental health specialist teams, we cannot of course be sure how much this improvement can be attributed to caseload size and how much to the additional training and skills of the specialist officers.

Sources of bias within the papers

Based on the ROBINS-I tool (Sterne et al., 2016) for assessing methodological rigour we identified areas of bias common across most studies. Bias due to

confounding was the most common; there was generally a high likelihood of potential for confounding of the effect of intervention in the studies. For example, participants in Cox et al. (2005) were not randomly assigned to the two interventions – TVU (Technical Violation Unit) and PTP (Probation Transition Programme). Methods used to screen participants for inclusion were different for the two interventions, as were requirements for inclusion. There was no information given for specific analysis used to correct this bias (leading to classification as ‘serious risk’ of bias). Similarly, selection of participants to interventions in Manchak et al. (2014) was not random. Bias occurred due to (1) probationer effects; site recruitment required inmates to have been supervised on speciality caseload supervision, and on traditional probation to begin with, and (2) officer effects; some officers supervised multiple probationers, leading to potential ‘nesting’ of effects within officers. Therefore outcomes differences may reflect differences between officers who are consistent in their practices across multiple cases. However, propensity scores using a binary logistic regression and a mixed multilevel modelling strategy controlled for this bias. Bias due to confounding arose in Jalbert et al. (2011) as a result of the limited ability to match offenders to their criminal history records to measure arrest frequency, which led to only those with available data being included in the analysis. This could bias the data if availability of criminal histories was linked to offenders being supervised by officers with reduced caseloads versus officers with regular caseloads. However, researchers used a logistic regression model to test for systematic differences. There was also a lack of diagnostic information available in Wolff et al. (2014) for participants for two interventions. Wolff et al. (2014) also encountered bias due to deviations from intended interventions, 81 participants were excluded from the analysis due to switching interventions, and 545 participants were excluded from HLM regression due to missing independent variable. There is no information on how this was controlled for, leading to a classification of moderate bias. Jalbert et al. (2011) suffered bias due to deviation from intended interventions. A randomised controlled trial (RCT) was originally proposed for all three sites involved. However, for two sites the RCT was disbanded in favour of a regression discontinuity design (RDD), and the third site was replaced with a difference in differences (DD) estimator. Additionally, Jalbert et al. (2011) suffered from missing data – some officers from the control group switched administrative assignments and there was attrition among officers between follow-up times. Level of Service Inventory (LSI) scores were missing for approximately 20 percent of probationers, although bias was corrected for using a regression-based imputation procedure. Similarly, Taxman et al. (2006) experienced attrition and both non-PCS cases with missing data and their matched PCS counterparts were excluded from analysis.

Synthesis of findings

The five studies which were assessed to be of sufficient methodological rigour to be retained for analysis have a number of similarities. All five were US studies which examine the impact of particular initiatives delivered by mainstream community-

Table 3. Reductions in average caseload size for retained studies.

Study	Average intervention officer caseload	Average comparison group officer caseload	Ratio between intervention & comparison group
Cox et al. (2005) (PTP)	25	100	1: 4
Cox et al. (2005) (TVU)	25	100	1: 4
Jalbert et al. (2011)	54	106	1: 2
Manchak et al. (2014)	50	100	1: 2
Taxman et al. (2006)	55	100	1: 1.8
Wolff et al. (2014)	30	130	1: 4.3

based probation officers working for the local City or State. Further, all five are comparative studies seeking to examine the differential impact of these initiatives compared with mainstream probation practice.

However, there were also substantial differences between the programmes. Three of these studies (Cox et al., 2005; Jalbert et al., 2011; Taxman et al., 2006) compared intensive probation programmes with deliberately reduced caseloads against mainstream probation supervision. The Cox et al. study compared two different programmes² in Connecticut, both of which were designed to increase supervision completion rates and reduce reoffending via evidence-based models of intervention with low caseloads. Jalbert et al. and Taxman et al both set out to test whether reduced caseloads improved probation outcomes in jurisdictions specifically chosen because they applied Evidence Based Practice (EBP), an approach analogous to the Risk, Needs, Responsivity (RNR) paradigm implemented widely in England and Wales. The two other studies (Manchak et al., 2014; Wolff et al., 2014) compared specialist mental health probation teams to mainstream probation provision. In these cases, although a reduced caseload was a critical component of the specialist provision, there were other differences in terms of specialist mental health probation officers having additional training and adopting a different approach to the people they supervised. These differences in the interventions studied led us to decide not to undertake a statistical meta-analysis.

All five studies were rigorous about measuring the caseload size for both the intervention and comparison groups. As we can see from Table 3 below, the reductions in caseload size were considerable:

All five studies recorded reductions in measured outcomes compared to comparators. As can be see from Table 4, both Jalbert et al. (2011) and Taxman et al. (2006) found lower rearrest rates amongst the probationers supervised by officers with smaller caseloads while Cox et al. (2005) (in both the programmes evaluated) and Manchak et al. (2014) found lower probation violation rates for these groups. Wolff recorded a lower number of average jail days for the intervention group in her study, although she does not specify whether these jail days are related to probation violations or rearrests.

Table 4. Criminal justice outcomes for retained studies.

Study	Outcome Verdict	Outcome detail
Cox et al. (2005) (PTP)	Probation violation rates lower for intervention group	8% violation rate compared to 13% for comparison group
Cox et al. (2005) (TVU)	Probation violation rates lower for intervention group	30% violation rate compared to an expected 100% violation rate ³ for comparison group
Jalbert et al. (2011)	Lower rate of reoffending for intervention group	25.5% reduction in recidivism at 6 months and 36% reduction in both property & violent crime reoffending at 18 and 30 months.
Manchak et al. (2014)	Probation violation rates lower for intervention group	Violation rate approximately half that of comparison group [Odds Ratio = 2.19]
Taxman et al. (2006)	Lower rate of re-arrests & probation violation rates for intervention group	32.1% rearrest rate compared to 40.9% for comparison group 20.1% violation rate compared to 29.2% for comparison group
Wolff et al. (2014)	The average number of jail days reduced by a greater proportion for the intervention group	19% and 24% violation rate for two intervention groups compared to 32% for comparison group

Wolff et al. (2014) also reported improved mental health outcomes for the intervention group with probationers self reporting improved mental health symptoms, less loneliness or boredom, better work performance, and the fact that their emotional problems interfere less in their lives. However, the researchers were not able to access comparative data from those probationers with mental health problems receiving mainstream supervision.

Overall, although the number of robust studies remains quite small for such a key area of consideration, there appears to be a growing body of evidence that lower probation caseloads have a positive impact in terms of reducing reoffending in the USA. Clearly, there are numerous differences between the US and English and Welsh probation systems which means that caution should be exercised in assuming that lower caseloads in this country would also result in reduced reoffending.

One key issue is that smaller caseloads naturally lead to more contact with probationers and this might increase the probability of more detection of any violations or breaches of the conditions of supervision. This in turn can lead to the termination of more orders, and an increase in the imposition of custodial sentences because of failure to comply with community sentences. All of these research teams were aware that previous evaluations had found that intensive supervision programmes could result in a higher rate of what are known in the US as probation violations (the equivalent of breaching the conditions of supervision in an English and Welsh context). They attributed this finding to the fact that supervising officers in intensive programmes have more contact with their probationers and were therefore more likely to detect when such violations/breaches had occurred.

In these five evaluations, the research teams looked at a range of criminal justice outcomes including technical probation violations, violations for new arrests and re-convictions. The studies looking at mental health probation teams (Manchak et al., 2014; Wolff et al., 2014) also looked at the extent to which probationers engaged with community treatment services. These studies did not find any large increases in violations or breaches of the conditions of supervision.

Interestingly, although researchers were expecting to find a higher rate of probation violations among the cohorts supervised by probation officers with lower caseloads, this did not turn out to be the case.

In the five studies which examined violation rates, the violation rate for the intervention group was consistently lower than for the comparison cohorts receiving mainstream supervision from probation officers with higher caseloads.

Discussion and conclusion

This REA used a 4-step search strategy to identify potentially relevant papers, which, through a structured sifting process considering both their relevance and methodological rigour resulted in the identification of 5 papers that were analysed in detail.

All five were US studies which examine the impact of particular initiatives delivered by mainstream community-based probation officers. They were all comparative studies seeking to examine the differential impact of these initiatives compared with mainstream probation practice. However, there were also substantial differences between the programmes. Three of the studies compared intensive probation programmes with deliberately reduced caseloads against mainstream probation supervision. Two studies compared specialist mental health probation teams operating reduced caseloads to mainstream probation provision. These differences in the interventions studied led us to decide not to undertake a statistical meta-analysis.

All five studies recorded reductions in measured outcomes compared to comparators. Outcomes measured included lower rearrest rates, lower probation violation rates and a lower number of average jail days. We conclude that there is some evidence from methodologically robust studies that lower probation caseloads can reduce recidivism, although there is no robust evidence that reduced caseloads reduce re-offending where re-offending is understood as conviction for new offences as oppose to technical violations of an existing probation order. One key issue is that smaller caseloads naturally lead to more contact with probationers and this might increase the probability of more detection of any violations or breaches of the conditions of supervision. These studies did not find any large increases in violations or breaches of the conditions of supervision.

All of these are US studies so any attempt to draw conclusions about UK practice must be extremely cautious. As can be seen from Table 3 caseloads in all of these studies were reduced from rates of a hundred or more. While there are probation officers in England and Wales with caseloads at this level, this would represent a high probation caseload. There are other significant differences between UK and

US probation practice that also suggest caution in assuming that effective studies in the US would be replicated in the UK.

To supplement the findings from the REA we also examined evidence from other, comparable sectors. There are many studies evaluating the impact of caseload size on the outcomes of clinical interventions in a health setting. However, this setting seemed less relevant to probation practice. Instead, we concentrated on sectors where professionals maintain a sustained engagement with people on their case-loads undertaking a mixture of assessment, referral and intervention through a series of interactions. These settings included: youth justice, social work, substance misuse, mental health, education welfare and complex needs. Of particular interest were two substantial reviews of interventions in mental health where size of caseload was a component of the intervention (Dieterich et al., 2017; Happell et al., 2012).

Dieterich et al. (2017) is a systematic review of effects of Intensive Case Management (ICM), which is characterised as involving lower caseloads, as a means of caring for severely mentally ill people in the community. The review looked at the findings from 40 randomised controlled trials with a total of 7524 participants. Dieterich et al. found that ICM is effective in ameliorating many outcomes relevant to people with severe mental illness compared to standard care. However, when the researchers compared ICM with what they term non-ICM – a similar approach but with a larger caseload – they found that there was moderate-quality evidence that ICM probably makes little or no difference in the average number of days in hospital per month or in the average number of hospital admissions. However, Dieterich et al. described the quality of the evidence as ‘at best... of moderate quality’ (Dieterich et al., 2017: 3).

Happell and colleagues (2012) in their synthesis of research and policy on the contribution mental health nurses to community case management work reflects many of our findings in the probation sector:

- Determining caseloads using ratios of clients per case manager might be overly simplistic. Happell identifies seven factors to be considered when developing caseload indices: contact frequency, response difficulty, intervention type, competence/seniority of the case manager, caseload maturity, location of clients, and roles other than case management;
- Heavy workloads can be counter-productive to mental health nurses providing optimum care for patients;
- The size of caseloads is associated with case managers’ perceptions of their own clinical effectiveness;
- Overall there is not clear evidence that smaller caseloads lead to better patient outcomes; and
- The evidence-based from which policymakers can draw when making decisions relating to caseloads is limited.

Our brief review of research on caseload size in other sectors suggests that the composition of caseload and support to deliver effective practice is at least equally and probably more important than overall caseload size in determining individual

worker caseload levels. There is an emerging theme across sectors that increased administrative burdens on workers mean that even less demanding cases generate considerable workload and that the additional time spent with service users who receive a more intensive service is rarely equivalent to the amount of time dedicated to administrative tasks.

Looking at other sectors gives us additional confidence in the findings of this review of the impact of caseloads in the probation sector. There is some evidence from methodologically robust studies that lower probation caseloads can reduce recidivism, although there is no robust evidence that reduced caseloads reduce re-offending where re-offending is understood as conviction for new offences as oppose to technical violations of an existing probation order. It is also probable that caseload reductions need to be combined with training in more effective methods and support for probation staff if improved outcomes are to be realised.

Finally, there is clearly a need for more methodologically robust studies on caseload reduction in probation practice to be undertaken in the UK. The US studies we have identified show how such studies might be designed. They should include an outcome that measures recidivism as well as 'intermediate outcomes' and a comparative element such that they are classed as level 4 or 5 on the scientific methods scale set out in Table 1. The design should allow unintended outcomes such as smaller caseloads leading to an increase in the rate of technical breaches to be evaluated and accompanied by in-depth, qualitative research to explore how practice changes when caseloads are reduced. Indeed, any future study which explored the relationship between the quality of the supervisory relationship and any impact on reconviction outcomes would be particularly valuable.

Authors' note

Grace Hothersall was affiliated to Manchester Metropolitan University when research undertaken.

Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by HM Inspectorate of Probation.

ORCID iDs

Chris Fox  <https://orcid.org/0000-0003-1931-2058>

Grace Hothersall  <https://orcid.org/0000-0003-1891-4225>

Andrew Smith  <https://orcid.org/0000-0003-0344-913X>

Notes

1. For the text of the call visit <http://www.russellwebster.com/caseloadarea/>
2. The Probation Transition Programme and the Technical Violation Unit
3. Probationers were only referred to the TVU programme if they had already demonstrated poor compliance and were on the point of being violated.

References

References marked with asterisk in final selection of papers included in the analysis.

- Burrell W (2006) *Caseload Standards for Probation and Parole*. Lexington, KY: American Probation & Parole Association. Available at: https://www.appa-net.org/eweb/docs/APPA/stances/ip_CSPP.pdf (accessed 1 March 2020).
- Campbell Collaboration (2014) Campbell Collaboration Systematic Reviews: Policies and Guidelines. DOI: 10.4073/csrs.2014.1.
- Council of Europe (2019) Persons under the supervision of probation agencies. SPACE II-2018.
- *Cox S, Bantley K and Roscoe T (2005) *Evaluation of the Court Support Services Division's Probation Transition Program and Technical Violation Unit: Final Report*. Central Connecticut State University.
- DeMichele M (2007) *Probation and Parole's Growing Caseloads and Workloads Allocation: Strategies for Managerial Decision-Making*. Lexington, KY: The American Probation & Parole Association.
- Dieterich M, Irving C, Bergman H, et al. (2017) Intensive case management for severe mental illness. *Cochrane Database of Systematic Reviews* 2017(1): CD007906.
- Folkard M, Smith D and Smith D (1976) *IMPACT Volume II: The Results of The experiment. Research Study 36*. London: HMSO.
- Friendship C, Street R, Cann J, et al. (2005) Introduction: the policy context and assessing the evidence. In: Harper G and Chitty C (eds) (2005), *Home Office Research Study No. 291. The impact of corrections on re-offending: a review of 'what works'*. London: Home Office.
- Gayman M, Powell N and Bradley M (2018) Probation/parole officer psychological wellbeing: the impact of supervising persons with mental health needs. *American Journal of Criminal Justice* 43(3): 509–529.
- Happell B, Hoey W and Gaskin C (2012) Community mental health nurses, caseloads, and practices: a literature review. *International Journal of Mental Health Nursing* 21: 131–137.
- HM Inspectorate of Probation (2019) *Report of the Chief Inspector of Probation*. Available at: <https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2019/03/HMI-Probation-Chief-Inspectors-Report.pdf> (accessed 2 April 2021).
- HM Inspectorate of Probation (2021a) *Caseloads, Workloads and Staffing Levels in Probation Services*. Research and Analysis Bulletin 2021/02. Manchester: HMIP.
- HM Inspectorate of Probation (2021b) *A Thematic Review of Work to Prepare for the Unification of Probation Services*. Manchester: HMIP.
- Jalbert S and Rhodes W (2012) Reduced caseloads improve probation outcomes. *Journal of Crime and Justice* 35(2): 221–238.

- *Jalbert S, Rhodes W, Kane M, et al. (2011) *A Multi-Sited Valuation of Reduce Probation Caseload Size and Evidence-Based Practice Setting*. Washington, DC: US Department of Justice.
- *Manchak S, Skeen J, Kennealy P, et al. (2014) High-fidelity specialty mental health probation improves officer practices, treatment access and rule compliance. *Journal of Law and Human Behaviour* 38(5): 450–461.
- McWilliams W (1966) Probation failure. *Case Conference* 13(3): 89–92.
- Ministry of Justice (2018) *Justice Secretary Outlines Future Vision for Probation*. London: MoJ. Available at: <https://www.gov.uk/government/news/justice-secretary-outlines-future-vision-for-probation>.
- Phillips J, Westaby C and Fowler A (2016) 'Its relentless': the impact of working primarily with high-risk offenders. *Probation Journal* 63(2): 182–192.
- Raynor P (2018) From 'nothing works' to 'post-truth': the rise and fall of evidence in British probation. *European Probation Journal* 10(1): 59–75.
- Russell J (2020) *Probation – In Crisis or on the Road to Recovery?* Academy for Social Justice Lecture, Tuesday, 30 June 2020.
- Skeem JL and Eno Loudon J (2006) Toward evidence-based practice for probationers and parolees mandated to mental health treatment. *Psychiatric Services* 57(3): 333–342
- Sterne JA, Hernán MA, Reeves BC, et al. (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Online)* 355: i4919.
- *Taxman F, Yancey C and Bilanin J (2006) *Proactive Community Supervision in Maryland*. Baltimore, MD: Maryland Division of Parole and Probation.
- Union of Safety and Justice Employees (2019) Protecting public safety: the challenges facing federal parole officers in Canada's highly stressed criminal justice system. Available at: <http://www.usje-sesj.com/en/reports> (accessed 1 March 2020).
- *Wolff N, Epperson M, Shi J, et al. (2014) Mental health specialist probation caseloads: Are they effective? *International Journal of Law and Psychiatry* 37: 464–472.