

ARE WE BARKING UP THE RIGHT TREE?

A Meta-Analysis on the Effectiveness of Prison-Based Dog Programs

HANNE M. DUINDAM 

University of Amsterdam

JESSICA J. ASSCHER

University of Amsterdam

Utrecht University

MACHTELD HOEVE

GEERT JAN J. M. STAMS

HANNEKE E. CREEMERS

University of Amsterdam

Prison-based Dog Programs (PBDPs) are used in correctional facilities to decrease recidivism and improve social-emotional functioning. The aim of this meta-analysis was to provide an overview of the effectiveness of PBDPs, accounting for the potential influence of study, program, and sample characteristics through moderator analyses. We included 11 manuscripts, seven published and four unpublished, yielding 93 effect sizes ($N = 3,013$). Eight studies were quasi-experimental and three were randomized controlled trials. The overall effect of PBDPs was significant and small ($d = 0.153$, 95% confidence interval [CI] = [0.026, 0.281]), and may have been somewhat inflated by possible publication bias, while study quality was generally low. Moderator analyses showed that the overall effect was largely driven by the small-to-medium effect of PBDPs on recidivism ($d = 0.414$, 95% CI = [0.153, 0.676]). It is therefore concluded that PBDPs may be a promising intervention to reduce recidivism, although more (robust) research is needed.

Keywords: Animal-Assisted Intervention; correctional facilities; multi-level meta-analysis; Prison-based Dog Programs; primary and secondary outcomes

INTRODUCTION

A variety of intervention and diversion programs exist in correctional facilities that aim to reduce criminal recidivism, such as vocational training, education programs, substance abuse interventions, and treatment of mental health problems (Tripodi et al.,

AUTHORS' NOTE: *The authors thank Janneke Staaks and Emmeke Kooistra for their assistance. Correspondence concerning this article should be addressed to Hanne M. Duindam, Forensic Child and Youth Care Sciences, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, the Netherlands; e-mail: h.m.duindam@uva.nl.*

CRIMINAL JUSTICE AND BEHAVIOR, 2020, Vol. 47, No. 6, June 2020, 749–767.

DOI: 10.1177/0093854820909875

Article reuse guidelines: sagepub.com/journals-permissions



© 2020 International Association for Correctional and Forensic Psychology

2011; Wilson et al., 2000). One popular relatively new and underexplored intervention that might positively contribute to rehabilitation is the Prison-based Animal Program (PAP). The most commonly used animals in PAPs are dogs, that is, Prison-based Dog Programs (PBDPs; Furst, 2006; Jasperson, 2013). The aim of the present meta-analysis was to examine the effectiveness of PBDPs in reducing criminal recidivism and improving social-emotional functioning of people convicted of a crime, while testing to what extent study, program, and sample characteristics moderate the outcome of the program.

Even though there is not one prevailing theory that explains the working mechanisms of PBDPs, several have been proposed building on the abundance of research supporting (therapeutic) benefits of the human–animal interaction (Mercer et al., 2015). Human–animal contact has psychosocial as well as physiological benefits, such as reducing stress and providing social support (Barker & Dawson, 1998; Barak et al., 2001; Berget et al., 2008; Brooks et al., 2013; Friedmann et al., 1980; Nimer & Lundahl, 2007; Prothmann et al., 2006). Therefore, companion animals have a long history of being incorporated in the care for those with physical and psychiatric illnesses (Furst, 2006; Leonardi et al., 2017). When focusing on the working mechanisms of PBDPs, dogs in particular are viewed as catalysts for change because they respond directly to what they observe, either positive or negative, providing people with instant behavioral feedback that might motivate them to critically self-assess and learn new (emotional regulation) skills (Kruger et al., 2004). Effective emotion (e.g., anger) regulation has been identified as a dynamic criminogenic need, which is therefore considered to be a promising intermediate target to help reduce criminal recidivism (Andrews et al., 2006).

In addition, individuals in correctional facilities may become attached to the dog, which may be a new and healing experience that may help them cope with stress. Using an animal as a secure base may also make those who are incarcerated more receptive to the therapeutic impact of PBDPs (Jasperson, 2010). Moreover, attachment to the dog may generalize to relationships with humans; which may ultimately help in the process of developing more positive relationships with others who have parted from a criminal lifestyle (Holbrook et al., 2001; Jasperson, 2010), thereby potentially contributing to a protective factor for criminal recidivism: relationships with prosocial peers (Andrews et al., 2006).

Several types of PBDPs exist (e.g., Grommon et al., 2018; Jasperson, 2013). One of these is the dog-training program (DTP), in which individuals who are incarcerated train asylum dogs, equipping them for adoption (i.e., community service design), or train dogs to become assistance dogs for people with disabilities or mental health problems (i.e., service animal socialization program; Furst, 2006). Assisting the dog or the community may help people perceive themselves as someone who can take responsibility and do good for others, thereby conforming to the rules and expectations of society, which has been shown to be associated with desistance, because it helps build an alternative, “anticriminal” identity (Andrews et al., 2006; Hill, 2016). Moreover, taking responsibility and caring for the dog are skills believed to generalize to other life domains (e.g., work) and relationships with humans, thereby improving overall social-emotional functioning and aiding the rehabilitation process (Humby & Barclay, 2018).

DTPs are the most common form of PBDP in the United States and Australia (Furst, 2006; Humby & Barclay, 2018). However, dogs are also incorporated into therapeutic

interventions in correctional facilities in the form of Animal-Assisted Interventions (i.e., AAI) or Animal Assisted Therapy (i.e., AAT) to facilitate the achievement of therapeutic or educational outcomes (Contalbrigo et al., 2017; Furst, 2006; Jaspersen, 2013; Nimer & Lundahl, 2007). In AAI/T's, the interaction between dog and patient is controlled to serve a therapeutic purpose best achieved through exposure to the animal (Leonardi et al., 2017; Mercer et al., 2015; Nimer & Lundahl, 2007). The contact between dogs and participants in AAI/Ts is generally guided by a professional to facilitate the achievement of preset therapeutic goals. For example, a dog might facilitate a pleasant and safe atmosphere in stressful environments, such as correctional facilities, to make attendants more receptive to group therapy. DTPs differ from AAI/Ts in that their purpose is not to exclusively serve a therapeutic aim. Rather a substantial focus of DTPs is directed toward the training and future well-being of the dogs, without employing the therapeutic techniques that are used in AAI/Ts. Nevertheless, DTPs may contribute to rehabilitation by strengthening the bond to society (adherence to social norms and expectations) and improving psychological functioning (Hill, 2016; Leonardi et al., 2017).

In the last decades, the popularity of PBDPs—and community service DTPs in particular—has been rising worldwide (Britton & Button, 2005; Mulcahy & McLaughlin, 2013). Not only have these programs become popular in correctional facilities, they also generally receive positive attention from the public and media because of their outward appeal of connecting two groups that are both isolated from society (i.e., shelter dogs and people in prison; Britton & Button, 2005; Mulcahy & McLaughlin, 2013). Findings of studies that have examined the effectiveness of PBDPs highlight positive effects, such as improved physical and emotional health, reduced prison misconduct, and an enhanced sense of responsibility and self-control (e.g., Leonardi et al., 2017; Mercer et al., 2015). However, many studies have relied solely on qualitative data and anecdotal reports of the effectiveness of PBDPs.

Cooke and Farrington (2016) conducted a first meta-analysis on DTPs exclusively, which synthesized the (quantitative) research findings on this type of PBDP up until 2014. They found a large and significant effect (Cohen's $d = 0.78$) of DTPs on externalizing outcomes (including recidivism, aggression, self-control, and institutional infractions), and a small positive effect (Cohen's $d = 0.24$) on internalizing outcomes (Cooke & Farrington, 2016). Unfortunately, studies with a quasi-experimental or randomized controlled trial (RCT) design were generally lacking, compelling Cooke and Farrington to also include studies without a control group.

Recently, efforts have been made to evaluate PBDPs more rigorously. For example, Hill (2018) conducted a large, retrospective study on the effectiveness of DTPs on postrelease recidivism in Florida, and found that the likelihood of re-arrest within 1 year decreased. In addition, an RCT by Seivert et al. (2016) of an AAI for juveniles failed to show an improvement in behavioral functioning. These and similar research advances in the field warrant an updated, more stringent overview of the effectiveness of PBDPs, including both DTP and AAI/T, comparing their effectiveness for improving social-emotional and behavioral functioning. The current meta-analysis therefore assessed the effectiveness of all types of PBDPs that build on the benefits of dog-human interaction, including only quasi-experimental and RCT studies. We formulated no a priori hypotheses about the overall effectiveness of PBDPs because previous research showed conflicting findings.

A three-level meta-analytic model was chosen for this study over more traditional meta-analytical approaches used previously because it allows for the inclusion of multiple effect sizes from the same study (Lipsey & Wilson, 2001; Van den Noortgate et al., 2013). Therefore, all relevant information about the effectiveness of PBDPs could be derived from research (Assink et al., 2015). Moreover, this type of analysis enables the examination of both within and between study, sample, and intervention characteristics as potential moderators, thereby gaining new insights in what works for whom under the studied conditions. For example, it can be examined whether the effect of PBDPs is larger for programs with a different approach (e.g., AAI vs. DTP) or longer duration, and to what extent study quality and participants' age moderate the effects. To our knowledge, previous research on PBDPs has not examined the potential role of sample and intervention characteristics on the program's effectiveness. Therefore, moderator analyses were exploratory.

Overall, the present meta-analysis provided new insights into the effectiveness of PBDPs, aiming to create a more robust scientific foundation for future research. The current meta-analysis added to the previous meta-analysis by Cooke and Farrington (2016), increasing knowledge on the effectiveness of PBDPs by (a) only including studies with a control group (i.e., reducing extraneous and temporal threats to validity); (b) examining the effects of a broader scope of PBDPs that have been implemented in correctional facilities (as opposed to DTPs alone); (c) applying a three-level random-effects model that allows for the inclusion of multiple effect sizes from the same study; and (d) examining to what extent study, sample, and intervention characteristics influenced the effectiveness of PBDPs. In sum, we assessed the overall effectiveness of PBDPs in terms of improving social-emotional functioning and decreasing (risk for) criminal recidivism. To our knowledge, this is the first meta-analysis on (all types of) PBDPs that included an examination of potential moderating factors on the overall effectiveness of the programs, and which took into account both within and between study variability.

METHOD

STUDY SAMPLE

A systematic literature search was conducted to select studies for the current meta-analysis. Studies were included up until March 2019. The following eight databases were selected for the literature search: PubMed, PsycInfo, Google Scholar, Criminal Justice Abstracts, Web of Science, Social Services Abstracts, Medline, and Scopus. Search strings included "animal assisted therapy/dog/pet/companion/animal/canine/puppy/treatment/training/therapy/intervention/workshop/project/counsel/course" in combination with "correctional institutions/incarceration/maximum security facility/prison/forensic psychiatric/remand centre/penitentiary/inmate." To maximize the chance of finding additional (nonpublished) studies conducted on PBDPs, the search was extended by conducting a Google search examining the first 200 hits. These hits included gray literature, such as reports, conference proceedings, posters, master's theses, and dissertations. Furthermore, reference lists of a review (Mulcahy & McLaughlin, 2013) and a meta-analysis (Cooke & Farrington, 2016) on DTPs were checked for additional studies.

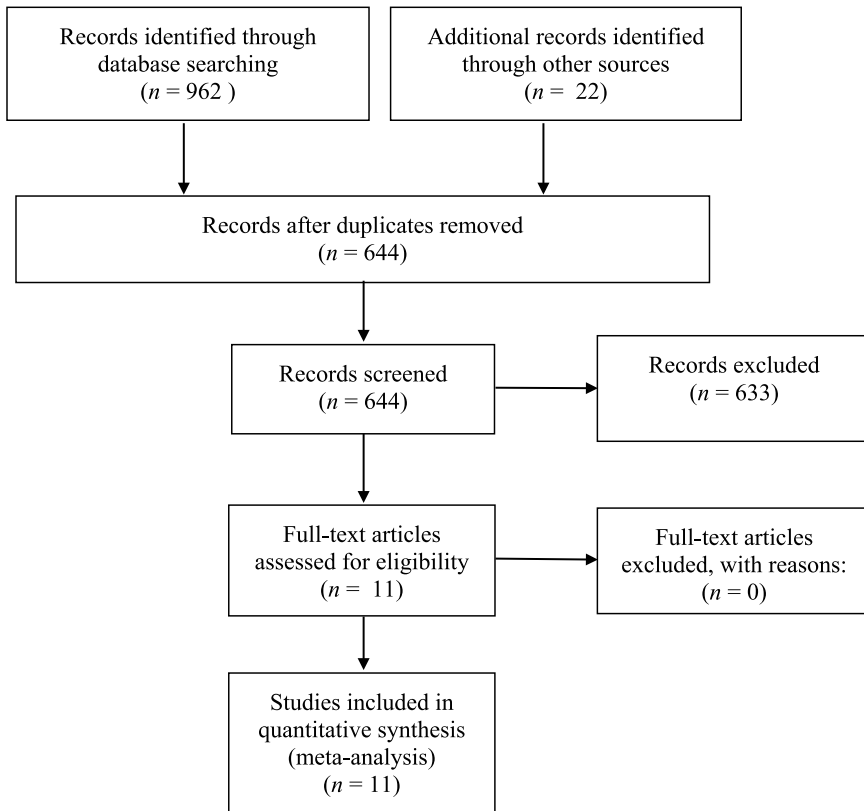


Figure 1: Flowchart of Literature Search and Screening Process

Source: Adapted from Moher et al. (2009).

Note. n = amount of records.

To be selected, studies had to meet several inclusion criteria. First, studies were selected when they examined PBDPs. No criteria were formulated based on the type of PBDP (e.g., DTP, AAI), which means that studies were included as long as the animal was a dog and the intervention was carried out in a correctional facility. The dog (vs. other animals) was chosen as this type of animal is most frequently used in interventions in correctional facilities (Furst, 2006). Second, studies had to include a control group consisting of participants who were not receiving the PBDP under study. Third, studies had to report quantitative outcomes that permitted the calculation of Cohen's d . If studies did not provide the needed information, authors were contacted and asked for the missing data.

In total, the searches yielded 644 studies (see Figure 1). Studies were screened for relevance based on titles, abstracts, and full texts (if necessary). Eleven studies were included, yielding 93 effect sizes, reporting on 11 independent samples consisting of $N = 3,013$ participants in total. Study characteristics are displayed in Table 1. Two other studies (i.e., Duncan, 2011; Gilger, 2007) were deemed appropriate for the current meta-analysis, but despite extensive efforts of the authors to retrieve the required information from the

TABLE 1: Study Characteristics

First author	N	Gender	Sample	No. of ES		Year	Quality (0–12)	Program type	Outcome	Study	
				P	NP					Design	Type
Chianese	144	Female	Youth	—	1	2009	2	DTP	P	Quasi	Retro
Contalbrigo	18	Male	Adult	17	—	2017	4	AAI	S	Quasi	Pro
Cooke	77	Male	Adult	—	6	2014	2	DTP	P & S	Quasi	Pro/Retro
Fournier	48	Male	Adult	7	16	2007	4	DTP	S	Quasi	Pro
Grommon	310	Male	Youth	6	—	2018	5	DTP	S	RCT	Pro
Hill	1,170	Male	Adult	4	1	2018	5	DTP	P & S	Quasi	Retro
Jaspersen	74	Female	Adult	3	—	2013	8	AAI	S	RCT	Pro
Mulcahy	10	Male	Adult	—	17	2011	2	DTP	S	Quasi	Pro
Richardson-Taylor	23	Female	Adult	—	5	2001	0	DTP	S	Quasi	Pro
Seivert	138	Mix	Youth	6	—	2016	7	DTP	S	RCT	Pro
Van Wormer	1,001	Male	Adult	4	—	2017	4	DTP	S	Quasi	Retro

Note. N = sample size in analysis; No. of ES = number of effect sizes per manuscript; P = published effect sizes extracted from journal article; NP = nonpublished effect sizes extracted from manuscript (i.e., dissertation, master's thesis); AAI = Animal-Assisted Intervention, defined as a therapeutic intervention that incorporates one or more dog(s); DTP = Dog Training Program, defined as a program where individuals train dogs new skills for different purposes (e.g., rehoming, to become service animal); (Outcome) P = primary outcomes variables (i.e., criminal recidivism); (Outcome) S = secondary outcome variables (i.e., social-emotional functioning); Quasi = quasi-experimental design; RCT = randomized controlled trial; Pro = prospective; Retro = retrospective.

researchers, the full texts articles remained unavailable. In total, four unpublished manuscripts (i.e., master's theses, dissertations) and seven published manuscripts were included. It is important to note that for two of the published manuscripts, we also included effect sizes from the unpublished (i.e., dissertation) manuscripts of the same study because this information was not available in the published article (see Table 1 for an overview of the published vs. unpublished effect sizes). To be as inclusive as possible, all relevant effect sizes were included, regardless of publication status of the manuscript. No protocol for the current meta-analysis has been submitted. However, research methods and reporting were applied in line with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-analyses) statement, which consists of a 27-item checklist and a flow diagram of different phases to help researchers improve reporting their meta-analytical findings (Moher et al., 2009).

CODING OF STUDIES

To extract information from the studies in a cohesive and reliable manner, a coding form was developed that included several descriptors (i.e., title, author names, publication year, type of publication, research country). In addition, all the outcome variables and several study, sample, and intervention characteristics were listed on the coding form. Study characteristics included the outcome type. Outcome variables were categorized as primary outcome (i.e., criminal recidivism) or secondary outcomes in terms of social-emotional functioning (e.g., self-esteem, depression, anxiety, aggression), associated with well-being. In addition, the number of assessment points (i.e., one vs. multiple), participation and response rates, research design (i.e., RCT vs. quasi-experimental), study type (i.e.,

retrospective vs. prospective), and report type (i.e., self, staff, or registered data) were also recorded. Furthermore, information about the program enrollment of the control group (i.e., treatment as usual [TAU], waiting list, or alternative intervention) and whether the control group also had contact with dogs (e.g., by staying on the same floor as a participant with a dog) was also recorded.

The following sample characteristics were registered on the coding form: sample size, inclusion and exclusion criteria of program participation, population type (i.e., general, psychiatric, addicted, or different), mean age, gender, offense type (i.e., violence, sex delinquency, property crime), and cultural background (i.e., European American/Australian vs. other). Moreover, the intervention characteristics of program type (DTP vs. AAI) and participants' amount of contact with the dogs (i.e., full-time: dog resides with person in correctional facility 24/7, vs. part-time: dog visits correctional facility during intervention) were also recorded. In addition, the intervention characteristics of program duration (i.e., in weeks) and program intensity (i.e., amount of hours in therapy or training) were also coded. Unfortunately, only half of the studies reported (mostly limited) information on program duration and intensity. As these moderators could not be reliably assessed, they were excluded.

Two researchers (i.e., a trained PhD and master's student of educational sciences) coded the studies independently by filling out the form. Coding forms were compared in the presence of a third researcher (i.e., professor with extensive experience in conducting meta-analyses, last author), and inconsistencies were discussed until a consensus was reached. Most studies did not provide enough information to code all the characteristics. One characteristic—offense type—was excluded because it was not possible to create mutually exclusive categories (i.e., most studies included individuals convicted for multiple offenses of various nature or did not distinguish groups based on type of offense). Mean age (15.7–39.1 years), publication year, sample size, and proportion of “other” cultural background (20%–72%) were coded as continuous variables, whereas the other characteristics were coded as dichotomous variables.

Finally, the quality of the studies was coded using the Quality Assessment Tool Quantitative Studies (Thomas et al., 2004). Again, the two researchers independently scored the Quality Assessment Tool and inconsistencies were discussed until a consensus was reached under supervision of the third researcher. The following quality assessment components were scored: (presence of) selection bias, study design, (control of) confounders, blinding, validity/reliability of data collection methods, and withdrawals and dropout rates. Items were scored on a 3-point scale ranging from 0 (*weak*) to 2 (*strong*). The total scores were calculated by adding up the item scores, resulting in a quality score per study on a scale of 0 to 12 (mean score = 3.9; Thomas et al., 2004).

STATISTICAL ANALYSIS

Computation of Effect Sizes

For the 11 studies, effect sizes (Cohen's *d*) were calculated for each outcome variable to generate a statistic that summarizes the effects of PBDPs on participants' social-emotional and behavioral functioning in terms of criminal recidivism. Mostly, mean values, standard deviations, and proportions were extracted from studies to compute Cohen's *ds*. These

descriptive statistics were transformed into Cohen's *ds* using the formulas of Lipsey and Wilson (2001). Potential differences in scores between the intervention and control group were calculated at pre- and posttreatment using the Practical Meta-Analysis Effect Size Calculator (Lipsey & Wilson, 2001). To control for the influence of any pretreatment differences on the posttreatment effects, pretreatment differences were subtracted from posttreatment effects. Subsequently, positive *d* values were assigned when the PBDP resulted in positive outcomes for the treatment group versus the control condition. Negative *d* values were assigned when there was less improvement on outcome variables for the treatment group, in comparison with the control condition.

A Three-Level Random Effects Meta-Analytic Model

In conducting a meta-analysis, either a fixed or random effects model can be used. For the current study, a three-level random effects model was chosen to compute the overall effect size (including all outcome variables) and conduct moderator analyses. This approach was preferred over the fixed-effects model, which assumes that there is no meaningful variation between studies, indicating that effect sizes of each individual study are considered to be approximates of the same, true population effect size (Van den Noortgate & Onghena, 2003). When several (nonidentical) studies are included in a meta-analysis, each with their own methodologies and samples, it is rarely the case that there is no variation between studies. The random-effects model does account for within- and between-study variation; therefore, it was considered more appropriate for the current meta-analysis. The effect sizes included in a random-effects model are seen as a random sample of the population of effect sizes, which allows generalization of the meta-analytical findings to the population of interest (Kelley & Kelley, 2012).

An advantage of the three-level random-effects model (over a two-level random-effects model) is that it allows for the inclusion of multiple effect sizes from the same study by modeling the different levels of variance (Assink & Wibbelink, 2016). An assumption of traditional meta-analysis is that effect sizes should be independent, which has been dealt with by, for example, averaging effect sizes from the same study or "hand-picking" the effects that seem most important (Assink & Wibbelink, 2016; Lipsey & Wilson, 2001). These methods result in a loss of relevant information and low(er) statistical power (Assink et al., 2019). Therefore, we chose to use the three-level random effects model, which accounts for the interdependency of effect sizes by modeling the sampling variance of effect sizes (Level 1), variance between effect sizes from the same study (Level 2), and the variance between studies (Level 3; Assink & Wibbelink, 2016).

First, an intercept only-model was used to calculate the overall effect of PBDPs on the outcome variables. Therefore, all outcome variables (i.e., primary and secondary outcomes) were taken together to assess the effectiveness of PBDPs as a whole. Study results were weighted based on sample size; no corrections were made to the raw data (e.g., no outlier removal; Hunter & Schmidt, 2004). Second, log-likelihood ratio tests were run to determine whether there was significant heterogeneity between effect sizes from the same studies (Level 2) and between studies (Level 3). If significant variance was present, potential moderators were included in the model, and moderator tests were run to examine whether they (significantly) influenced the overall effect. Before

TABLE 2: Overall Effect Sizes

Overall effect	No. of studies	No. of ES	Mean <i>d</i> (SE)	95% CI	<i>p</i>
Prison-based Dog Programs	11	93	0.153 (0.064)	[0.026, 0.281]	.019*
- Excluding Chianese	10	92	0.138 (0.064)	[0.010, 0.266]	.034*
- Excluding Contalbrigo	10	76	0.110 (0.055)	[0.001, 0.219]	.048*
- Excluding Cooke	10	87	0.153 (0.072)	[0.010, 0.296]	.036*
- Excluding Fournier	10	70	0.159 (0.074)	[0.011, 0.308]	.036*
- Excluding Grommon	10	87	0.179 (0.068)	[0.045, 0.314]	.009**
- Excluding Hill	10	88	0.145 (0.071)	[0.005, 0.286]	.043*
- Excluding Jasperson	10	90	0.155 (0.070)	[0.016, 0.294]	.030*
- Excluding Mulcahy	10	76	0.174 (0.066)	[0.043, 0.306]	.010*
- Excluding Richardson-Taylor	10	88	0.130 (0.063)	[0.005, 0.254]	.041*
- Excluding Seivert	10	87	0.183 (0.063)	[0.058, 0.308]	.004**
- Excluding Van Wormer	10	89	0.164 (0.073)	[0.019, 0.309]	.027*

Note. No. of ES = number of effect sizes; mean *d* = mean effect sizes; SE = standard error; CI = confidence interval; *p* = *p*-value of omnibus test.

p* < .05. *p* < .01.

conducting these moderator analyses, continuous (moderator) variables were centered around their mean, and dummy variables were created for the categorical moderators (Assink & Wibbelink, 2016). To investigate the robustness of the overall results, sensitivity analyses were conducted. The effect sizes were recalculated 11 times, each time removing a different study, to examine the influence of the individual studies on the overall effect (Viechtbauer & Cheung, 2010).

The File Drawer Problem

An advantage of a meta-analysis is that it allows researchers to statistically summarize a large amount of data on one topic (Lipsey & Wilson, 2001), thereby providing a meaningful overview of the research conducted so far. However, the accuracy and representativeness of this overview depend on the extent to which all studies relevant to the research area are included (Duval & Tweedie, 2000b). Inclusion is less likely for studies that did not find significant or positive effects, because they are less likely to be published, which is commonly referred to as the “file drawer problem” or “publication bias” (Rosenthal, 1979). For the present meta-analysis, we therefore made an effort to also include nonpublished studies, resulting in the incorporation of effect sizes of six unpublished dissertations and/or master theses.

Despite the best search methods, a meta-analysis that does not take potential missing data (i.e., effect sizes) into consideration can result in positively (or negatively) biased findings (Duval & Tweedie, 2000b). One way to examine the file drawer problem is by using the funnel-plot-based trim and fill method (Duval & Tweedie, 2000a, 2000b). This method is available through the metaphor package in R (R Development Core Team, 2018; Viechtbauer, 2010) and tests to what extent the effect sizes are normally distributed. An asymmetric funnel plot indicates bias; consequently, effect sizes are imputed to make the funnel plot symmetrical. Analyses are rerun with imputed effect sizes to examine the potential influence of publication bias on the overall effect. It is important to note that the imputed effect sizes can provide meaningful insight into the possible influence of publication bias, but should not be

regarded as “real” effect sizes of which the inclusion leads to a more “accurate” mean effect size estimation (Sutton et al., 2000).

RESULTS

OVERALL EFFECT SIZES

Results of the analysis indicated that there was a significant ($t = 2.392, p = .019$) and small overall mean effect ($d = 0.153, 95\% \text{ CI} = [0.026, 0.281]$) of PBDPs on primary and secondary outcomes in terms of criminal recidivism and social-emotional functioning, respectively. To assess each study’s contribution to the overall effect, analyses were rerun 11 times, each time removing a different study, see Table 2 (Viechtbauer & Cheung, 2010). Findings from these sensitivity analyses indicated that the overall effect remained significant after each rerun; therefore, none of the studies had an individual, disproportionate, impact on the overall findings. Log-likelihood-ratio test results indicated that there was significant heterogeneity between effect sizes within studies, Level 2: $\chi^2(1) = 133, 9541, p < .001$, and between studies, Level 3: $\chi^2(1) = 6, 1853, p = .013$, indicating that the outcome of PBDPs depends on intervention, sample or study characteristics. Results demonstrated that 40.4% of the overall variance in the outcome was attributable to differences between effect sizes within the same studies (Level 2), whereas 31% of the total variance was attributable to differences between studies (Level 3). To examine this variation, moderator analyses were conducted to provide insight into the extent to which study, sample, and intervention characteristics moderated the overall effect of PBDPs.

MODERATOR ANALYSES

Coding of studies led to the identification of 20 potential moderators, including 12 study, 6 sample, and 2 intervention characteristics. Table 3 displays the results of the moderator analyses. Findings demonstrated that there were three moderating effects (i.e., two study characteristics and one sample characteristic). First, outcome type significantly moderated the results, indicating that PBDPs were more effective in reducing the primary outcome criminal recidivism compared with secondary outcomes (i.e., social-emotional functioning). Second, a moderating effect of study design was found. The effects of PBDPs were larger for studies with a quasi-experimental (as opposed to randomized controlled) design. Third, age was a significant moderator, indicating that larger effects were found for older participants. Population type just failed to reach significance ($p = .052$), indicating larger treatment effects for individuals in Adult (vs. Juvenile) Justice Centers at trend level. None of the (other) study, sample, and intervention characteristics were found to significantly moderate the effects on PBDPs.

TRIM AND FILL ANALYSES

Results of the trim and fill analyses demonstrated that results were positively biased, as indicated by an asymmetrical distribution of effect sizes on the funnel plot and five missing negative effect sizes. Five new effect sizes were computed and included in subsequent analyses. Results indicated that publication bias may exist, because missing data (i.e., publication bias) did influence the results because the new overall effect size was smaller and did no longer reach the level of significance ($d = 0.082, p = .363$).

TABLE 3: Results of the Moderator Analyses

Moderator	No. of studies	No. of ES	Intercept/ Mean <i>d</i> [95% CI]	β [95% CI]	<i>F</i> (<i>df</i> ₁ , <i>df</i> ₂)	<i>p</i>
Study characteristics						
Outcome type	11	93				
Primary (RC)	3	6	0.414 [0.153, 0.676]**		<i>F</i> (1, 91) = 5.083	.027*
Secondary	9	87	0.108 [-0.020, -0.236]	-0.306 [-0.576, -0.036]*		
Publication year	11	93	0.166 [0.028, 0.304]*	-0.007 [-0.033, 0.019]	<i>F</i> (1, 91) = 0.300	.585
Publication status	11	93			<i>F</i> (1, 91) = 1.200	.276
Published (RC)	7	46	0.199 [0.039, 0.358]*			
Not published	6	47	0.089 [-0.097, 0.276]	-0.109 [-0.307, 0.089]		
Research country	11	93			<i>F</i> (1, 91) = 1.918	.170
United States (RC)	8	54	0.108 [-0.025, 0.240]			
Other	3	39	0.307 [0.053, 0.561]*	0.199 [-0.087, 0.486]	<i>F</i> (1, 91) = 2.094	.151
Study quality	11	93	0.174 [0.050, 0.298]**	-0.044 [-0.105, 0.017]	<i>F</i> (1, 91) = 1.828	.180
Assessment points	11	93				
Multiple (RC)	6	72	0.080 [-0.085, 0.246]	0.175 [-0.082, 0.431]	<i>F</i> (1, 91) = 1.439	.233
One	5	21	0.255 [0.059, 0.451]*			
Study type	11	93				
Prospective (RC)	8	82	0.108 [-0.048, 0.264]			
Retrospective	4	11	0.281 [0.033, 0.529]*	0.173 [-0.114, 0.460]	<i>F</i> (1, 91) = 5.539	.021*
Study design	11	93				
Quasi (RC)	8	78	0.218 [0.102, 0.335]**	-0.254 [-0.468, -0.040]*		
RCT	3	15	-0.036 [-0.215, 0.144]		<i>F</i> (2, 90) = 1.083	.343
Report type	11	93				
Self (RC)	8	73	0.122 [-0.044, 0.287]			
Staff	2	9	-0.014 [-0.340, 0.312]	-0.136 [-0.451, 0.179]		
Registered data	3	11	0.288 [0.031, 0.545]*	0.166 [-0.132, 0.465]		
Nature control group	5	55				
TAU (RC)	3	26	0.186 [-0.140, 0.512]		<i>F</i> (2, 52) = 0.478	.623
Waiting list	1	23	0.135 [-0.403, 0.673]	-0.051 [-0.679, 0.578]		
Intervention	1	6	-0.121 [-0.664, 0.422]	-0.307 [-0.940, 0.326]		

(continued)

TABLE 3. (continued)

Moderator	No. of studies	No. of ES	Intercept/ Mean <i>d</i> [95% CI]	β [95% CI]	<i>F</i> (<i>df</i> 1, <i>df</i> 2)	<i>p</i>
Control group intervention	9	84			<i>F</i> (1, 82) = 1.963	.165
Without animal (RC)	8	78	0.199 [0.032, 0.366]*			
With animal	1	6	-0.121 [-0.543, 0.302]	-0.320 [-0.775, 0.134]		.402
Control group dog contact	9	84				
No likely contact dog (RC)	4	27	0.243 [-0.015, 0.502]			
Likely contact dog	5	57	0.097 [-0.130, 0.325]	-0.146 [-0.490, 0.199]		
Sample characteristics						
Population	11	93				.052
Adult Justice Center (RC)	8	80	0.202 [0.087, 0.318]**			
Juvenile Justice Center	3	13	-0.021 [-0.215, 0.172]	-0.224 [-0.450, 0.002]		
Sample size	11	93				
<i>N</i> Total	11	93	0.172 [0.029, 0.314]*	-0.000 [-0.000, 0.000]	<i>F</i> (1, 91) = 0.770	.383
<i>N</i> Experimental	11	93	0.172 [0.025, 0.319]*	-0.000 [-0.001, 0.001]	<i>F</i> (1, 91) = 0.341	.561
<i>N</i> Control	11	93	0.178 [0.032, 0.324]*	-0.000 [-0.000, 0.000]	<i>F</i> (1, 91) = 0.770	.383
Age	8	78	0.189 [0.098, 0.279]**	0.016 [0.004, 0.028]*	<i>F</i> (1, 76) = 6.921	.010*
Addiction	11	93			<i>F</i> (1, 91) = 1.324	.253
Nonaddict (RC)	9	53	0.114 [-0.028, 0.256]			
Addict	2	40	0.288 [0.023, 0.554]*	0.174 [-0.127, 0.475]	<i>F</i> (2, 90) = 2.238	.113
Gender	12	93				
Male (RC)	7	78	0.144 [0.010, 0.279]*			
Female	3	9	0.366 [0.071, 0.662]*	0.222 [-0.103, 0.546]		
Mixed	2	6	-0.121 [-0.472, 0.230]	-0.265 [-0.641, 0.110]		
Cultural background, % other	8	66	0.111 [-0.131, 0.353]	-0.001 [-0.006, 0.004]	<i>F</i> (1, 64) = 0.206	.651
Intervention characteristics						
Program type	11	93				.115
Dog Training Program (RC)	9	73	0.109 [-0.013, 0.231]		<i>F</i> (1, 91) = 2.537	
AAI	2	20	0.355 [0.074, 0.635]*	0.245 [-0.061, 0.551]		.759
Contact dog ³	7	56				
Part-time (RC)	4	32	0.104 [-0.166, 0.374]			
Full-time	3	24	0.171 [-0.173, 0.516]	0.067 [-0.370, 0.505]		

Note. No. of ES = number of effect size; mean *d* = mean effect size; CI = confidence interval; β = regression coefficient; *F*(*df*1, *df*2) = results of the omnibus test; *p* = *p*-value of omnibus test; RC = reference category; Quasi = quasi-experimental design; RCT = randomized controlled trial; TAU = treatment as usual; AAI = Animal-Assisted Intervention, defined as a therapeutic intervention that incorporates one or more dog(s); (Cultural background) % other = proportion of other background; Contact dog³ = the dogs visit the correctional facility for the intervention (i.e., part-time); the dogs reside with the individual 24/7 (i.e., full-time). **p* < .05. ***p* < .01. ****p* < .001.

DISCUSSION

The present meta-analysis was conducted to provide an overview of the effectiveness of PBDPs because research evidence proved to be equivocal despite their popularity worldwide (Britton & Button, 2005; Mulcahy & McLaughlin, 2013). In total, 11 studies ($N = 3,013$) were included, yielding 93 effect sizes. The overall effect of PBDPs on primary and secondary outcomes proved to be significant and small (Cohen's $d = 0.153$), but may have been somewhat inflated due to possible publication bias. Results of the moderator analyses demonstrated that outcome type, study design, and age were significant moderators. We found a small-to-medium effect for PBDPs in terms of reducing the primary outcome of criminal recidivism, whereas no effect for PBDPs was found on the secondary outcomes (i.e., social-emotional functioning).

The lack of a significant effect on secondary outcomes suggests that factors other than social-emotional functioning are responsible for the positive effect of PBDPs on criminal recidivism. For example, PBDPs may be effective in reducing criminal recidivism by helping individuals who are incarcerated build an alternative "anticriminal" identity (Andrews et al., 2006; Hill, 2016), or by making them more susceptible to treatment targeting criminal recidivism (Jasperson, 2010; Mulder et al., 2011). The secondary outcomes in the current meta-analysis encompassed social-emotional outcomes that primarily concern people's well-being instead of established dynamic criminogenic needs (Bonta et al., 2014). Intermediate factors that promote improved social-emotional functioning and well-being (e.g., institutional group climate; Van der Helm & Stams, 2012) may not be sufficiently targeted in PBDPs, which would offer an explanation for the insignificant effect on secondary outcomes. Working mechanisms of PBDP's are an interesting area for future research to explore.

In addition, larger effects of PBDPs were found for studies with a quasi-experimental design, as compared with RCTs. RCTs are seen as the "golden" standard in intervention research because differences in outcomes cannot be attributed to initial (unmeasured) differences between participants in the experimental and control group, thus ruling out alternative explanations for the intervention's effectiveness (see Farrington, 2003). Quasi-experimental designs are—because of nonrandom assignment—more susceptible to the biasing effects of confounding (unmeasured) factors, which are responsible for the treatment effect. As larger treatment effects were found for quasi-experimental studies, unknown or unmeasured confounders may have inflated the effects of PBDPs. Furthermore, somewhat larger treatment effects were found for older PBDP participants. This finding suggests that effectiveness of PBDPs depends on the age of participants. These results support the general notion in mental health care that research is needed to assess what works for children or adolescents rather than implementing interventions that have been proven effective with adults (Patel et al., 2007).

The current meta-analysis showed only a small positive (overall) effect for PBDPs irrespective of program type (DTP vs. AAI), which appeared largely driven by the small-to-medium effect of PBDPs on criminal recidivism. These results are in line with effects found by Cooke and Farrington (2016) for DTPs on externalizing outcomes, although they found a larger effect, but do not confirm the small effect found by Cooke and Farrington (2016) on internalizing outcomes. A plausible explanation is found in the inclusion of six additional studies on PBDPs that were published after 2014, and in the more stringent inclusion criteria of the present meta-analysis, which resulted in the exclusion of three studies without a control group (included by Cooke and Farrington), which found strong, positive, program effects (i.e., Merriam, 2001; Moneymaker & Strimple, 1991; Walsh & Mertin, 1994). In addition, the present three-level meta-analysis included multiple effect sizes per study,

using all available information, whereas Cooke and Farrington (2016) included only one effect size per study. These changes could have resulted in different effects found for PBDPs in the current meta-analysis.

LIMITATIONS

There are some limitations worth mentioning. Despite our extensive search, only 11 studies with relatively low quality ratings were available (i.e., mean score of 3.9 out of 12). Often, studies received low scores on study design and methods, for example, few studies were prospective or RCTs, and the sample sizes were generally small (i.e., $N < 80$ for six of the 11 included studies). For example, moderator analyses demonstrated positive program effects on the primary outcome (i.e., criminal recidivism), but this finding was based on only three retrospective studies. Future prospective experimental studies on the effectiveness of PBDPs are necessary to examine to what extent this finding is robust (Weisburd, 2003). Nevertheless, this meta-analysis is based on the most robust, published and unpublished research on PBDPs that is currently available, and provides a comprehensive overview of the current state of empirical evidence from the perspective of the amount of studies and their quality.

Another limitation is that potentially important moderators, including program duration and intensity, could not be examined due to lack of information. Moreover, an interesting and important intervention aspect that could not be examined was implementation fidelity (Goense et al., 2016), as none of the studies reported to what extent the PBDPs were carried out (and received) as intended. As a consequence, the internal validity of the studies was threatened, preventing reliable conclusions to be drawn about whether the (lack of) effect is accounted for by PBDPs or by unknown (and unmeasured) factors that were accidentally added to or left out from the intervention (Bellg et al., 2004; Moncher & Prinz, 1991). Finally, there were some categories of moderators that yielded significant effects (e.g., AAI program type). However, the omnibus moderator tests did not show a significant difference between categories. Some of these moderator analyses were hampered by insufficient statistical power to detect small or medium effects sizes, because only few studies reported information on these moderators (e.g., nature of control group).

FUTURE DIRECTIONS

Fortunately, the research quality of studies on AAI/Ts has improved in recent years; this positive trend is also apparent in the literature on PBDPs included in the current meta-analysis (i.e., pre-2015 mean quality score is 3, vs. 5 post-2015; Hoagwood et al., 2017; May et al., 2016). It is important that this positive trend continues so that more definitive conclusions about the effectiveness of PBDPs can be drawn.

An additional avenue for future research to explore is PBDPs in combination with evidence-based treatments, because adding dog-assisted interventions onto established treatment programs might lead to better treatment responsivity, higher attendance, and more positive treatment outcomes overall (e.g., Calvo et al., 2016). This might be of particular interest for the forensic population, for whom lack of treatment adherence has been identified as a (dynamic) risk factor for criminal recidivism (Mulder et al., 2011). In this regard, it is also important for future research to include more process (e.g., treatment motivation) and instrumental (e.g., treatment attendance) oriented measures, in addition to outcome variables such as criminal recidivism and social-emotional functioning.

Given the status quo of current research into the effectiveness of PBDPs in combination with the rising popularity of PBDPs, we make an appeal for future robust research in this area, and hope that the findings of this meta-analysis will be utilized as a stepping-stone for high quality research. This study generates several new research questions for future studies on the effectiveness of PBDPs. For example, how do PBDPs affect criminal recidivism? Are PBDPs effective for most people in correctional facilities or are there specific subgroups that could benefit more (or less)? What is the “optimal” duration of PBDPs? In the current meta-analysis, we were only able to detect three significant moderators, leaving some of the heterogeneity between effect sizes unexplained. As suggested, other factors may have influenced PBDPs’ effectiveness that we could not account for (e.g., stronger effectiveness in certain subgroups, level of implementation fidelity). We therefore highly recommend the inclusion of detailed program, intervention, and sample descriptions in future studies, allowing future meta-analytic studies to test all relevant moderators, and to retain more of the included studies in moderator-analyses.

Conducting research in correctional facilities is challenging in a variety of ways due to the restricted and—at times—unpredictable nature of the environment (for an overview of challenges and strategies in prison research, for example, Apa et al., 2012; Wakai et al., 2009). It may not always be possible to conduct RCTs in such a tightly controlled setting with limited access to participants. Therefore, other research designs (e.g., quasi-experimental designs) could be considered in addition to RCTs (Hein & Weeland, 2019). One promising research design in the prison context may be the multiple case series design (MCSD), as this might be more easily implemented (i.e., inexpensive, requiring a smaller sample size). In a MCSD, a smaller group of individuals are assessed multiple times before, during, and after an intervention. Participants function as their own control through the multiple measurements across phases (Kratochwill et al., 2010). An advantage of the MCSD in the PBDP context is that it allows for a more close-up observation of the variation within and between individuals’ responses to PBDPs, which could not only help identify characteristics that moderate program effectiveness but may ultimately also help illuminate how PBDPs work (Kazdin, 2008). By examining within-participant changes over time with short intervals, potential working mechanisms of PBDPs could be identified. For example, MCSDs can be useful to study whether short term improvements in self-assessment and emotional regulation precede behavioral change at a later stage, and whether or not behavioral changes following PBDPs are contingent upon attachment to the dog. In general, this type of information could help fine-tune PBDPs and also aid the design of large-scale future research on PBDPs. Overall, more detailed recording and reporting of program and participants’ characteristics is needed to allow research that generates knowledge on the effectiveness of PBDPs. Such research enables policy makers, program administrators, and correctional facilities alike to make better informed decisions about when and under which conditions the PBDP is a suitable intervention.


A similar call for more robust research and reporting has been identified in recent systematic reviews and a meta-analysis on AAI/Ts with nonforensic populations (e.g., youth, patients with psychiatric disorders, elderly), illustrating the need for a stronger scientific foundation of AAI/Ts overall (Kendall et al., 2015; Lundqvist et al., 2017; Maujean et al., 2015; Virués-Ortega et al., 2012). Moreover, May and colleagues (2016) systematically examined the study designs of recent literature on AAT for youth, and concluded that methodological improvements are necessary to advance the field, because control groups, random assignment, follow-up assessments, and mixed methods were frequently lacking.

In addition to a call for more robust research, we would like to stress the importance of publishing studies that did not find significant or positive effects. Despite our extensive efforts to also include gray literature, which is seen as an important strategy to minimize publication bias (Paez, 2017), our findings (i.e., the asymmetrical funnel plot of effect sizes) demonstrated the possible presence of publication bias. Preregistration of future experimental studies on PBDPs is therefore strongly recommended, because this helps to reduce publication bias and overall prevents selective reporting (Aslam et al., 2013).

CONCLUSION

This meta-analysis found a small overall effect of PBDPs, which was largely driven by the program's effect on the primary outcome of criminal recidivism. The present meta-analysis complements the literature by presenting a comprehensive, up-to-date overview of the effectiveness of PBDPs. In line with the state of the science on AAI/Ts and dog training programs with nonforensic samples, more (robust) future research is needed to determine whether the rising dissemination of PBDPs in correctional facilities is justified. Even though these findings are promising, research was often hampered by weak study designs and methods. The benefits of human-animal interaction have long been recognized (see Friedmann et al., 1980). However, relatively little (research) attention has gone to the psychology of human-animal relationships in general, despite the important role animals play in human life (Amiot & Bastian, 2015). We hope that this meta-analysis contributes to the field and encourages future, robust research on the effectiveness of PBDPs, to systematically explore the potentials of the human-dog bond in forensic settings.

ORCID iD

Hanne M. Duindam  <https://orcid.org/0000-0002-5934-820X>

REFERENCES

- References marked with an asterisk indicate studies included in the meta-analysis. The in-text citations to studies selected for meta-analysis are not preceded with asterisks.
- Amiot, C. E., & Bastian, B. (2015). Toward a psychology of human-animal relations. *Psychological Bulletin*, *141*(1), 6–47. <https://doi.org/10.1037/a0038147>
- Andrews, D., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, *52*(1), 7–27. <https://doi.org/10.1177/0011128705281756>
- Apa, Z. L., Bai, R., Mukherejee, D. V., Herzig, C. T. A., Koenigsmann, C., Lowy, F. D., & Larson, E. L. (2012). Challenges and strategies for research in prisons. *Public Health Nursing*, *29*(5), 467–472. <https://doi.org/10.1111/j.1525-1446.2012.01027.x>
- Aslam, A., Imanullah, S., Asim, M., & El-Menyar, A. (2013). Registration of clinical trials: Is it really needed? *North American Journal of Medical Sciences*, *5*(12), 713–715.
- Assink, M., van der Put, C. E., Hoeve, M., de Vries, S. L. A., Stams, G. J. J. M., & Oort, F. J. (2015). Risk factors for persistent delinquent behavior among juveniles: A meta-analytic review. *Clinical Psychology Review*, *42*, 47–61. <https://doi.org/10.1016/j.cpr.2015.08.002>
- Assink, M., van der Put, C. E., Meeuwse, M. W. C. M., de Jong, N. M., Oort, F. J., Stams, G. J. J. M., & Hoeve, M. (2019). Risk factors for child sexual abuse victimization: A meta-analytic review. *Psychological Bulletin*, *145*(5), 459–489. <https://doi.org/10.1037/bul0000188>
- Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. <https://doi.org/10.20982/tqmp.12.3.p154>
- Barak, Y., Savorai, O., Mavashev, S., & Beni, A. (2001). Animal-assisted therapy for elderly schizophrenic patients. *The American Journal of Geriatric Psychiatry*, *9*(4), 439–442. <https://doi.org/10.1097/00019442-200111000-00013>
- Barker, S. B., & Dawson, K. S. (1998). The effects of animal-assisted therapy on anxiety ratings of hospitalized psychiatric patients. *Psychiatric Services*, *49*(6), 797–801. <https://doi.org/10.1176/ps.49.6.797>

- Bell, A. J., Resnick, B., Minicucci, D. S., Ogedegbe, G., Ernst, D., Borrelli, B., . . . Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443–451. <https://doi.org/10.1037/0278-6133.23.5.443>
- Berget, B., Ekenberg, O., & Braastad, B. O. (2008). Animal-assisted therapy with farm animals for persons with psychiatric disorders: Effects on self-efficacy, coping ability and quality of life, a randomized controlled trial. *Clinical Practice and Epidemiology in Mental Health, 4*(1), Article 9. <https://doi.org/10.1186/1745-0179-4-9>
- Bonta, J., Blais, J., & Wilson, H. A. (2014). A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior, 19*(3), 278–287. <https://doi.org/10.1016/j.avb.2014.04.014>
- Britton, D. M., & Button, A. (2005). Prison pups: Assessing the effects of dog training programs in correctional facilities. *Journal of Family Social Work, 9*(4), 79–95.
- Brooks, H. L., Rogers, A., Kapadia, D., Pilgrim, J., Reeves, D., & Vassilev, I. (2013). Creature comforts: Personal communities, pets and the work of managing a long-term condition. *Chronic Illness, 9*(2), 87–102. <https://doi.org/10.1177/1742395312452620>
- Calvo, P., Fortuny, J. R., Guzmán, S., Macías, C., Bowen, J., García, M. L., . . . Fatjó, J. (2016). Animal Assisted Therapy (AAT) program as a useful adjunct to conventional psychosocial rehabilitation for patients with schizophrenia: Results of a small-scale randomized controlled trial. *Frontiers in Psychology, 7*, Article 631. <https://doi.org/10.3389/fpsyg.2016.00631>
- *Chianese, N. (2009). *Girls, jails, and puppy dog tails: An evaluation of the new leash on life program* [Master's thesis, Department of Sociology, California State University, Fullerton].
- *Contalbrigo, L., de Santis, M., Toson, M., Montanaro, M., Farina, L., Costa, A., & Nava, F. A. (2017). The efficacy of dog assisted therapy in detained drug users: A pilot study in an Italian attenuated custody institute. *International Journal of Environmental Research and Public Health, 14*(7), Article 683. <https://doi.org/10.3390/ijerph14070683>
- *Cooke, B. J. (2014). *Bad to the bone? The effects of dog training programs on factors related to desistance* [Master's thesis, Institute of Criminology, University of Cambridge].
- Cooke, B. J., & Farrington, D. P. (2016). The effectiveness of dog-training programs in prison: A systematic review and meta-analysis of the literature. *The Prison Journal, 96*(6), 854–876. <https://doi.org/10.1177/0032885516671919>
- Duncan, S. (2011). *PAWS program effectiveness analysis—Q20*. Quan-Qual Outcomes.
- Duval, S., & Tweedie, R. (2000a). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. <https://doi.org/10.1080/01621459.2000.10473905>
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
- Farrington, D. P. (2003). Methodological quality standards for evaluation research. *The ANNALS of the American Academy of Political and Social Science, 587*, 49–68. <https://doi.org/10.1177/0002716202250789>
- *Fournier, A. (2005). *An investigation of the psychosocial impact of human-animal interaction on a forensic population* [Doctoral dissertation, Virginia Polytechnic Institute and State University].
- *Fournier, A., Geller, E., & Fortney, E. (2007). Human-animal interaction in a prison setting: Impact on criminal behavior, treatment progress, and social skills. *Behavior and Social Issues, 16*(1), 89–105.
- Friedmann, E., Katcher, A., Lynch, J., & Thomas, S. (1980). Animal companions and one-year survival of patients after discharge from a coronary care unit. *Public Health Reports, 95*(4), 307–312.
- Furst, G. (2006). Prison-based animal programs. *The Prison Journal, 86*(4), 407–430. <https://doi.org/10.1177/0032885506293242>
- Gilger, C. G. (2007). *Cell dogs: No effect of dog training programs on prisoners' self-efficacy* [Doctoral dissertation, Emporia State University].
- Goense, P. B., Assink, M., Stams, G. J., Boendermaker, L., & Hovee, M. (2016). Making “what works” work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior. *Aggression and Violent Behavior, 31*, 106–115. <https://doi.org/10.1016/j.avb.2016.08.003>
- *Grommon, E., Carson, D. C., & Kenney, L. (2018). An experimental trial of a dog-training program in a juvenile detention center. *Journal of Experimental Criminology, 41*(1), 62–80. <https://doi.org/10.1007/s11292-018-9346-2>
- Hein, S., & Weeland, J. (2019). Introduction to the special issue: Randomized Controlled Trials (RCTs) in clinical and community settings: Challenges, alternatives, and supplementary designs. *New Directions for Child and Adolescent Development, 2019*(167), 7–15. <https://doi.org/10.1002/cad.20312>
- *Hill, L. B. (2016). *Becoming the person your dog thinks you are: An assessment of florida prison-based dog training programs on prison misconduct, post-release employment and recidivism* [Doctoral dissertation, College of Criminology & Criminal Justice, Florida State University].
- *Hill, L. B. (2018). *Becoming the person your dog thinks you are: An assessment of Florida prison-based dog training programs on postrelease recidivism. Corrections: Policy, Practice, and Research*. Advance online publication. <https://doi.org/10.1080/23774657.2018.1433564>
- Hoagwood, K. E., Acri, M., Morrissey, M., & Peth-Pierce, R. (2017). Animal-assisted therapies for youth with or at risk for mental health problems: A systematic review. *Applied Developmental Science, 21*(1), 1–13. <https://doi.org/10.1080/1088691.2015.1134267>

- Holbrook, M., Stephens, D., Day, E., Holbrook, S., & Strazar, G. (2001). A collective stereographic photo essay on key aspects of animal companionship: the truth about dogs and cats. *Academy of Marketing Science Review*, 2001(1), 1–16.
- Humby, L., & Barclay, E. (2018). Pawisitive solutions: An overview of prison dog programs in Australia. *The Prison Journal*, 98(5), 580–603. <https://doi.org/10.1177/0032885518793951>
- Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). SAGE.
- Jaspersen, R. A. (2010). *Therapeutic interventions and animal assisted therapy with incarcerated females* [Doctoral dissertation, College of Social Work, The University of Utah].
- *Jaspersen, R. A. (2013). An animal-assisted therapy intervention with female inmates. *Anthrozoös*, 26(1), 135–145. <https://doi.org/10.2752/175303713X13534238631678>
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63(3), 146–159. <https://doi.org/10.1037/0003-066X.63.3.146>
- Kelley, G. A., & Kelley, K. A. (2012). Statistical models for meta-analysis: A brief tutorial. *World Journal of Methodology*, 2(4), 27–32. <https://doi.org/10.5662/wjm.v2.i4.27>
- Kendall, E., Maujean, A., Pepping, C. A., Downes, M., Lakhani, A., Byrne, J., & Macfarlane, K. (2015). A systematic review of the efficacy of equine-assisted interventions on psychological outcomes. *European Journal of Psychotherapy & Counselling*, 17(1), 57–79. <https://doi.org/10.1080/13642537.2014.996169>
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case design technical documentation*. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf
- Kruger, K. A., Trachtenberg, S. W., & Serpell, J. A. (2004). *Can animals help humans heal? Animal-assisted interventions in adolescent mental health*. Center for the Interaction of Animals & Society, School of Veterinary Medicine, University of Pennsylvania.
- Leonardi, R. J., Buchanan-Smith, H. M., McIvor, G., & Vick, S. J. (2017). “You think you’re helping them, but they’re helping you too”: Experiences of Scottish male young offenders participating in a dog training program. *International Journal of Environmental Research and Public Health*, 14(8), Article 945. <https://doi.org/10.3390/ijerph14080945>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE.
- Lundqvist, M., Carlsson, P., Sjö Dahl, R., Theodorsson, E., & Levin, L. Å. (2017). Patient benefit of dog-assisted interventions in health care: A systematic review. *BMC Complementary and Alternative Medicine*, 17(1), Article 358. <https://doi.org/10.1186/s12906-017-1844-7>
- Maujean, A., Pepping, C. A., & Kendall, E. (2015). A systematic review of randomized controlled trials of animal-assisted therapy on psychosocial outcomes. *Anthrozoös*, 28(1), 23–36. <https://doi.org/10.2752/089279315X14129350721812>
- May, D. K., Seivert, N. P., Cano, A., Casey, R. J., Johnson, A., & May, D. (2016). Animal-assisted therapy for youth: A systematic methodological critique. *Human-Animal Interaction Bulletin*, 4(1), 1–18.
- Mercer, J., Gibson, K., & Clayton, D. (2015). The therapeutic potential of a prison-based animal programme in the UK. *Journal of Forensic Practice*, 17(1), 43–54. <https://doi.org/10.1108/JFP-09-2014-0031>
- Merriam, S. (2001). *Discovering project pooch: A special program for violent incarcerated male juveniles*. Discover N’ Education and Research.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–270.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247–266.
- Moneymaker, J. M., & Strimple, E. O. (1991). Animals and inmates: A sharing companionship behind bars. *Journal of Offender Rehabilitation*, 16(3–4), 133–152.
- *Mulcahy, C. (2011). *Companions on the inside: Pet ownership and the introduction of a prison dog program at darling downs correctional centre* [Master’s thesis, School of Population Health, The University of Queensland].
- Mulcahy, C., & McLaughlin, D. (2013). Is the tail wagging the dog? A review of the evidence for prison animal programs. *Australian Psychologist*, 48(5), 369–378. <https://doi.org/10.1111/ap.12021>
- Mulder, E., Brand, E., Bullens, R., & Van Marle, H. (2011). Risk factors for overall recidivism and severity of recidivism in serious juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology*, 55(1), 118–135. <https://doi.org/10.1177/0306624X09356683>
- Nimer, J., & Lundahl, B. (2007). Animal-assisted therapy: A meta-analysis. *Anthrozoös*, 20(3), 225–238. <https://doi.org/10.2752/089279307X224773>
- Paez, A. (2017). Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, 10(3), 219–240. <https://doi.org/10.1111/jebm.12266>
- Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: A global public-health challenge. *The Lancet*, 369(9569), 1302–1313. [https://doi.org/10.1016/S0140-6736\(07\)60368-7](https://doi.org/10.1016/S0140-6736(07)60368-7)
- Prothmann, A., Bienert, M., & Ettricht, C. (2006). Dogs in child psychotherapy: Effects on state of mind. *Anthrozoös*, 19(3), 265–277.
- R Development Core Team. (2018). *R: A language and environment for statistical computing*. <http://www.rproject.org/>
- *Richardson-Taylor, K., & Blanchette, K. (2001). *Results of an evaluation of the pawisitive directions canine program at nova institution for women*. Correctional Service of Canada.

- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.
- *Seivert, N. P., Cano, A., Casey, R. J., Johnson, A., & May, D. K. (2016). Animal assisted therapy for incarcerated youth: A randomized controlled trial. *Applied Developmental Science*, *22*(2), 139–153. <https://doi.org/10.1080/10888691.2016.1234935>
- Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, *320*(7249), 1574–1577. <https://doi.org/10.1136/bmj.320.7249.1574>
- Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, *1*(3), 176–184. <https://doi.org/10.1111/j.1524-475X.2004.04006.x>
- Tripodi, S. J., Bledsoe, S. E., Kim, J. S., & Bender, K. (2011). Effects of correctional-based programs for female inmates: A systematic review. *Research on Social Work Practice*, *21*(1), 15–31. <https://doi.org/10.1177/1049731509352337>
- Van der Helm, G. H. P., & Stams, G. J. J. M. (2012). Conflict and coping by clients and group workers in secure residential facilities. In K. Oei, & M. Groenhuizen (Eds.), *Progression in forensic psychiatry: About boundaries* (pp. 553–564). Amsterdam, Netherlands: Kluwer.
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. <https://doi.org/10.3758/s13428-012-0261-6>
- Van den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63*(5), 765–790. <https://doi.org/10.1177/0013164402251027>
- *Van Wormer, J., Kigerl, A., & Hamilton, Z. (2017). Digging deeper: Exploring the value of prison-based dog handler programs. *The Prison Journal*, *97*(4), 520–538. <https://doi.org/10.1177/0032885517712481>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.
- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- Virués-Ortega, J., Pastor-Barriuso, R., Castellote, J. M., Población, A., & de Pedro-Cuesta, J. (2012). Effect of animal-assisted therapy on the psychological and functional status of elderly populations and patients with psychiatric disorders: A meta-analysis. *Health Psychology Review*, *6*(2), 197–221. <https://doi.org/10.1080/17437199.2010.534965>
- Wakai, S., Shelton, D., Trestman, R. L., & Kesten, K. (2009). Conducting research in corrections: Challenges and solutions. *Behavioral Sciences & the Law*, *27*, 743–752. <https://doi.org/10.1002/bsl.894>
- Walsh, P. G., & Mertin, P. G. (1994). The training of pets as therapy dogs in a women’s prison: A pilot study. *Anthrozoös*, *7*(2), 124–128. <https://doi.org/10.2752/089279394787002014>
- Weisburd, D. (2003). Ethical practice and evaluation of interventions in crime and justice. The moral imperative for randomized trials. *Evaluation Review*, *27*(3), 336–354.
- Wilson, D. B., Gallagher, C. A., & McKenzie, 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.

Hanne M. Duindam, MA, is a PhD student of Forensic Child and Youth Care Sciences at the Research Institute of Child Development and Education (RICDE) at the University of Amsterdam, the Netherlands. For her dissertation, she examines the effectiveness of a prison-based dog training program in the Netherlands.

Jessica J. Asscher, PhD, is a professor of Forensic Child and Youth Care Sciences at the University of Utrecht, Faculty of Social Sciences, and University of Amsterdam, Faculty of Social and Behavioral Sciences, Department of Educational Sciences. She has conducted longitudinal and experimental research in the field of forensic child and youth care, and has a special interest in intervention research under clinically representative conditions.

Machteld Hoeve, PhD, is associate professor of Forensic Child and Youth Care Sciences at the RICDE at the University of Amsterdam, the Netherlands. Her research interests include explaining juvenile delinquency and in particular parenting skills and psychopathology in association with delinquency among adolescents and young adults.

Geert Jan J. M. Stams, is a professor of Forensic Child and Youth Care Sciences at the University of Amsterdam, Faculty of Social and Behavioral Sciences, Department of Educational Sciences. He has conducted research on child–parent attachment relationships, moral development, child maltreatment, and juvenile delinquency.

Hanneke E. Creemers, PhD, is assistant professor of Forensic Child and Youth Care Sciences at the RICDE at the University of Amsterdam, the Netherlands. Her research focuses on the development of severe child and parenting problems, including juvenile delinquency and child maltreatment and neglect, and the effectiveness and working mechanisms of interventions to reduce such problems.