A META-ANALYSIS OF THE EFFECTS OF WILDERNESS THERAPY ON DELINQUENT BEHAVIORS AMONG YOUTH

NATALIE BECK
JENNIFER S. WONG
Simon Fraser University

The purpose of the present meta-analysis was to determine the effectiveness of wilderness therapy in addressing youth delinquency. A systematic review of the literature was conducted using 27 electronic databases and numerous gray literature sources, surveying literature published from 1990 to 2020. The search identified 189 potential studies for inclusion, resulting in a final study pool of 11 studies contributing 14 effect sizes from a total sample of 1,874 treatment youths. Both self-reported delinquency and caregiver-reported delinquency were examined using separate random-effects models. Pooled analyses yielded large, positive, and significant effects of 0.832 and 1.054 respectively, indicating that wilderness therapy is potentially an effective tool for addressing delinquent behaviors among youth. Limitations of the study include a lack of moderator analyses due to the small sample sizes. Wilderness therapy is a promising form of diversion programming and further investigation into this treatment modality is warranted.

Keywords: wilderness therapy; adventure therapy; delinquency; adolescents; meta-analysis; diversion programming

Wilderness therapy (WT) is a form of youth diversion and intervention programming (Johnson et al., 2020), which involves the use of rural outdoor settings in conjunction with therapeutic activities. Participants come from a variety of backgrounds: Many are enrolled in WT programs by their parents for externalizing problems such as defiance, substance use, school problems, and truancy; some are referred by community agencies, schools, or police; other programs host adjudicated youth who are mandated to attend (Bettmann et al., 2011). WT programs are diverse, ranging from primitive survivalist treks wherein participants build their own shelters, cook over fires, and hike continuously, to residential base-camps focusing on substance use and clinical treatment (Gillis, Kivlighan, & Russell, 2016; Harper et al., 2007). Furthermore, WT engages participants in various forms of individual, group, and/or family therapy components. The current study uses meta-analysis to synthesize the effectiveness of WT programs in addressing delinquency and antisocial behaviors. By using a recent date range, distinct program type (expedition and base-camp

AUTHORS’ NOTE: Correspondence concerning this article should be addressed to Jennifer S. Wong, Associate Professor, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6; e-mail: jenwong@sfu.ca.
program models), and focused outcome measures (delinquency and antisocial behaviors), this study contributes new information to the existing WT literature.

**WHAT IS WT?**

WT has been described as a treatment model centered in established practices (e.g., group counseling), which is set in the context of life in the wilderness (Russell, 2006). Wilderness programs often involve hiking, canoeing, camping, and various survival skills activities and can involve short trips or months-long stays in rural areas. Three popular models of WT programs include (a) the expedition model, lasting less than 2 months; (b) the base-camp or short-term residential model, typically lasting 6 to 12 weeks; and (c) long-term residential programs, wherein youth live in rural camps for up to 2 years (Bettmann et al., 2016; Johnson et al., 2020). Programs using an expedition model involve a continuous trekking experience in which youth frequently set up camps in new locations, such as the Catherine Freer Wilderness Therapy Expeditions program that involves a 21-day hiking trip with a focus on primitive living and survival skills (Harper et al., 2007). In a base-camp, or short-term residential format, participants live in one location for the duration of the program and embark on short wilderness excursions (Bettmann et al., 2016). For example, Base Camp in Canada offers a 3-month residential program at a rural facility, with participants engaging in 1- to 10-day wilderness excursions at least twice per month (Nikkel, 2014). Conversely, in long-term residential programs participants do not engage in wilderness expeditions; rather, a wilderness component is introduced in daily activities or in the facility setting (Bettmann et al., 2016).

**THEORETICAL BASIS OF WT**

The genesis of WT programming has most commonly been linked to Kurt Hahn’s “Outward Bound,” developed in the 1940s, although deeper roots of WT have been traced back to the late 19th century (White, 2012). In the context of WT, Walsh and Golins’s (1976) Outward Bound process of change model can be illustrated as follows: (a) The participant is placed into a unique and unfamiliar physical and social environment; (b) in this environment, problem-solving tasks are presented; (c) leaders facilitate an environment in which participants can master these tasks; and (d) this mastery results in increased self-esteem, self-awareness, and sense of belonging, as well as lasting problem-solving behavior (Gass et al., 2012). The Outward Bound framework has been the basis for many subsequent programs and has been adapted to accommodate various WT treatments (Turner, 2009; Wichmann, 1990).

Although Outward Bound is the most commonly referenced origin of WT, others have pointed to the clear influence and appropriation of Indigenous worldviews in Western outdoor recreation practices (Harper et al., 2018; Mullins et al., 2016; Skidmore, 2017). Indigenous cultures have long valued relationships with nature—values that Western identities have often borrowed. For example, WT programs include activities from Indigenous practices such as talking circles and talking sticks, canoe tripping, rituals and rites of passage, and vision quests (Harper et al., 2018; Mullins et al., 2016; Skidmore, 2017). In her book, *Braiding Sweetgrass*, Kimmerer (2013) reminds us that the healing power of nature is not a Western invention: “in some Native languages the term for plants translates to “those who take care of us”” (p. 229).
Other approaches have been identified in the development of wilderness programming; for instance, family systems theories have informed some of the therapeutic aspects of WT programs (Clark et al., 2004). This can be illustrated by the inclusion of family-centered components such as family counseling or, commonly, a family day at the beginning or end of treatment (Lowe, 2005). In their evaluation of two WT programs, Harper and Russell (2008) highlighted the importance of the family systems theory approach; that is, to address problems in childhood and adolescence, the role of the family should be accounted for in any treatment processes. This might include online parenting workshops, parents joining their youth for certain WT program days, or letter writing and phone calls between youth participants and their families during treatment (Harper & Russell, 2008). Other programs are based on theories that address specific behaviors. For example, many WT programs are designed for youth with substance use disorders. Theoretical foundations of substance use prevention such as the 12-step model (Vissell, 2004) and models that promote resilience or assets such as empathy and self-efficacy are commonly used in programs that target substance use (Gass et al., 2012).

WHAT MAKES IT “THERAPY”?  

To be considered a true WT program, an element of therapeutic process must be clear within the program structure, including a theoretical basis in established forms of treatment, and the intention to facilitate therapeutic change (Gass et al., 2012). Therapeutic processes are often facilitated by social workers, counselors, psychologists, physicians, or other therapists (Gass et al., 2012). Many current standards for WT stress the importance of staff being trained in both wilderness facilitation roles and therapeutic techniques (Outdoor Behavioral Healthcare Council, 2014). Without the presence of this therapeutic role, wilderness programs may be better categorized as youth camps or challenge programs. However, many programs cite the wilderness aspect alone as being a therapeutic element. For instance, without relying on formal or clinical therapy processes, the experience of being in nature or participating in challenging activities may promote change and growth (Christensen, 2008). Common features that distinguish WT programming from wilderness programs that are not based in therapeutic modalities include state licensure, program supervision by clinical staff, individual and group therapy, family participation in treatment, staff trained to serve specific client populations, tailored treatment plans for clients, and cooperation between program staff and aftercare agencies and families (Russell, 2001). In contrast to these principles, nontherapeutic programs such as wilderness boot camps often use punishment and psychological abuse to elicit compliance in youth, rather than the nonforceful, nurturing approach that is used in WT (Harper & Russell, 2008; Russell, 2006).

Other attempts to define WT have established themselves more clearly within the health care sector. For instance, “outdoor behavioral health care” (OBH) has been defined as “the prescriptive use of wilderness experiences by licensed mental health professionals to meet the therapeutic needs of clients” (Outdoor Behavioral Healthcare Council, 2014). The creation of the OBH Council placed WT strictly within the health care sector by setting professional and clinical standards for recognized programs, such as the presence of individual or group therapy with licensed clinicians, state licensure, safety standards, and allowing participants to obtain insurance reimbursements for program fees (Outdoor Behavioral Healthcare Council, 2014; Tucker et al., 2017). Endeavors that align themselves with the
OBH Council’s guidelines tend to be multiweek expedition or base-camp programs that involve physical challenges, individual and/or group therapy, and team-building activities.

**EMPIRICAL EVIDENCE FOR WT PROGRAMS**

WT evaluations have focused on a variety of outcomes, for example, self-esteem, locus of control, depression/anxiety, behavioral observations, mental health issues, weight control, academic performance, self-efficacy, and resilience (Bettmann et al., 2016; Bowen & Neill, 2013; Tucker et al., 2016; Walsh, 2009). The current study is focused on the impacts of WT on outcomes of delinquency. Recent evaluations measuring the effect of WT programs on measures of youth crime and criminogenic behaviors have provided mixed results. For instance, Raymond (2016) measured the short-term offending behaviors of 222 participants in an 8-day wilderness expedition program, wherein participants hiked a 100 km circuit through remote Australian wilderness. The evaluation found no significant differences in offending behavior between the treatment and control groups (Raymond, 2016). Similarly, in an evaluation of Wilderness Endeavors, a 21-day residential program in the Midwestern United States, Walsh (2009) found no significant differences between the offense rates of 43 treatment group participants and 43 control group participants. However, WT may be more successful with other criminogenic outcome variables. Antisociality and delinquency, for example, refer to delinquent behaviors such as aggression, problematic sexual behaviors, truancy, and other disruptive acts (Sawyer et al., 2015). Paquette and Vitaro (2014) found significant reductions in antisociality among 220 young adults participating in the Chance for Change program, a 10- to 20-day expedition model for at-risk youth in Scotland. Furthermore, WT has been shown as an effective treatment for problem behaviors, such as measurements of social problems on the Youth-Outcome Questionnaire (Y-OQ; Bettmann et al., 2013; Johnson et al., 2020).

Many studies have used the Y-OQ and the Youth-Outcome Questionnaire–Self-Report (Y-OQ-SR) to measure WT outcomes (e.g., Johnson et al., 2020; Lowe, 2005; Russell et al., 2018). The Y-OQ includes a measure of social problems, which encompasses behaviors such as “truancy, promiscuity, running away, substance use, violence, delinquent or aggressive behaviors, conflict difficulties, aggressiveness, breaking of social mores, and destruction of property” (Lowe, 2005, p. 135). Bettmann et al.’s (2013) evaluation of 40 youths in an 8-week WT program found significant reductions on the social problems subscale of the Y-OQ-SR from admission to discharge, with continued improvement evident in the 6-month follow-up. Similarly, Turner (2009) found significant improvements in social problems on both the Y-OQ caregiver report and the Y-OQ-SR after a 21-day wilderness expedition.

Despite existing empirical support for WT programming as a treatment for antisociality and delinquency, the link between WT’s theory of change and its applicability to the aforementioned criminogenic behaviors is not well established. While programs that incorporate established theoretical models such as the risk–need–responsivity model (RNR; Bonta & Andrews, 2007) or cognitive-behavioral therapy (Butler et al., 2006) have empirical support, programs that do not adhere to such principles may be considered ill-founded (Latessa et al., 2002). Theoretical models such as Walsh and Golins’s (1976) Outward Bound theory, which links self-esteem with a reduction in recidivism, have been contested (Bushman et al., 2009), while others have found small but significant negative relationships between self-esteem and delinquency (Mier & Ladny, 2018). Recent attempts to outline a clear
theory of change in WT have suggested a clinical psychotherapy model in which three dimensions—wilderness, the physical self, and the psychosocial self—are addressed concurrently to treat problem behaviors and mental health issues (Fernee et al., 2017). This theory posits the wilderness component of WT as a means of removing participants from negative stimuli and stress in their normal lives while allowing them to reflect and grow confidence.

Several meta-analyses on WT programming have been published in the past two decades; we begin by summarizing the ways in which the current study contributes to the existing literature. First, the date range (January 1990–February 2021) reflects an updated understanding of WT’s effect on delinquent behaviors and reflects the change in WT standards in the early 1990s that set WT apart from boot camps and other nontherapeutic programs and was followed by the establishment of the OBH Council in 1996. The current study also focuses on a narrower relationship than many of the previous meta-analyses by limiting eligible studies to those examining youth, and by limiting program type to a stricter definition of WT. Last, the current study improves on understanding WT’s effectiveness in treating delinquent behaviors by only combining outcome measurements that use similar instruments and respondents.

With respect to the existing literature, five prior meta-analyses examined the effects of WT programs on youth criminogenic variables. Specifically, Wilson and Lipsey (2000) and Bedard (2004) focused on delinquency outcomes, while Bettmann et al. (2016), Bowen and Neill (2013), and Gillis, Speelman, et al. (2016) included some form of antisocial problem behaviors within larger sets of outcome variables. An overview of these studies in comparison with the current research is presented in Table 1, including a summary of the findings in each meta-analysis, the number of studies in each, the number of studies overlapping with the present study, and the delinquency-related outcomes and conclusions of each meta-analysis.

As shown in Table 1, a significant amount of time has passed (17 years) since a meta-analysis has focused specifically on the effects of WT programming on delinquency; very few studies in the current analysis overlap with those used in either Wilson and Lipsey (2000) or Bedard (2004). Four of the five meta-analyses identified include publications that pre-date the scope of the current study, with some samples dating back to 1967. It is reasonable to assume that programs and participants from more than 30 years ago would be different from more recent WT programs.

Although most WT programs tend to focus on youth populations, adult-centered programs exist as well. Bettmann et al. (2016) and Bowen and Neill (2013) included both youth and adult samples in their meta-analyses; Bowen and Neill used a moderator analysis to examine participant age and found that adults may experience different outcomes than youth. Furthermore, Bedard (2004), Bettmann et al. (2016), and Bowen and Neill (2013) included samples from programs designed for specific populations, such as substance users, sex offenders, youth with serious illnesses such as cancer and diabetes, and youth with disabilities. Programs designed for specific populations also frequently include population-specific program models, such as the 12-step model for youth with substance use issues. For example, Gillis and Gass (2010) evaluated a wilderness program for juvenile sex offenders wherein participants completed workbooks on sexual behavior, engaged in group activities while respecting the physical boundaries of their peers, and were prompted to examine the power dynamics in their relationships to foster appropriate sexual behaviors. Alternatively,
| Study                              | Study date range | Program type                                                                 | No. of studies included in outcome of interest<sup>a</sup> | No. of studies overlapping with current study | Outcome of interest                                                                                   | Conclusions                                                                 |
|-----------------------------------|------------------|------------------------------------------------------------------------------|----------------------------------------------------------|---------------------------------------------|--------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Wilson and Lipsey (2000)          | 1967–1992        | Wilderness programs targeting criminogenic factors, involving challenge and   | 22                                                       | 1                                           | Antisocial behavior and delinquency, inc. antisocial behavior, official reports of recidivism, self-reported delinquency | Positive and significant effect of 0.18 on antisocial behavior and delinquency |
|                                   |                  | interpersonal elements, targeted to at-risk youth                             |                                                          |                                             |                                                                                                        |                                                                            |
| Bedard (2004)                     | 1968–1995        | Programs involving a wilderness component in a primitive environment          | 13                                                       | 1                                           | Recidivism, defined as police contact or contact with the criminal justice system                      | Positive and significant effect of 0.31 on recidivism                    |
| Bettmann et al. (2016)            | 1982–2014        | Programs in a primitive wilderness component, minimum duration of 5 days;     | 14                                                       | 4                                           | Behavioral outcomes, defined as a measure similar to the Y-OQ                                       | Positive and significant effect of 0.75 on behavioral outcomes           |
|                                   |                  | analysis samples <50% adjudicated youth                                       |                                                          |                                             |                                                                                                        |                                                                            |
| Bowen and Neill (2013)            | 1967–2012        | Programs focusing on adventure-based activities for therapeutic purposes     | 84                                                       | 5                                           | Behavioral outcomes, defined as ability to modify behavior based on the environment                   | Positive and significant effect of 0.41 on behavioral outcomes           |
| Gillis, Speelman, et al. (2016)   | 1998–2013        | Wilderness programs (no specification) and other mental health programs used  | 10                                                       | 3                                           | Y-OQ-SR and Y-OQ parent reports, total scores                                                       | Positive and significant effect of 0.72 on the Y-OQ-SR and 1.38 on Y-OQ parent reports |

<sup>a</sup> All five meta-analyses included a larger set of primary studies overall (e.g., Wilson & Lipsey, 2000 = 28 studies; Bedard, 2004 = 23 studies); however, many of the primary studies were not included in the authors’ analyses of delinquency/problem behavior outcomes. Here, only those studies that were included in relevant outcome analyses are listed. Y-OQ = Youth-Outcome Questionnaire; Y-OQ-SR = Youth-Outcome Questionnaire–Self-Report.
WT programs for adolescents with cancer may not include behavior management principles at all, but instead focus on quality of life or psychological and emotional outcomes (Stevens et al., 2004). Such programs clearly employ models that are distinct from WT programs not designed for specific client groups and may be less generalizable to a wider population; consequently, the current study excluded programs targeting specific populations.

In addition, most of the meta-analyses surveyed synthesized programs with very different formats. In an attempt to create a homogeneous sample, the current study focused on a strict operational definition of WT, which excluded ropes course programs, wilderness integrated family programs, inpatient or outpatient adventure counseling programs, school-based programs, and day camps. Last, many prior meta-analyses have focused on broad outcome measures, combining self-reports with parent or teacher reports and officially reported data from health and justice agencies, or combining data from multiple instruments to measure the same outcome. Although this is not uncommon, it does present some challenges. For example, adolescent reports of their own behaviors have different reliability and validity ratings than parent reports (Cannon et al., 2010; Hartung et al., 2005). Gillis, Speelman, et al.’s (2016) meta-analysis yielded different effect sizes between caregiver-reported and self-reported Y-OQ scores; however, as this meta-analysis only provided total scores, WT’s effectiveness on delinquent behaviors in particular could not be determined. It is possible that, for example, adolescents may be better at rating their own problem behaviors while parents are better at rating the adolescent’s social skills. For this reason, the current study only pooled self-reports of antisocial behaviors with other comparable self-reports. Parent reports were combined with reports completed by counselor and teachers. Little evidence has evaluated the convergence between these respondents on measures of problem behavior; however, Cannon et al. (2010) note that parent reports and reports from respondents other than self have a similar rate of change.

THE CURRENT STUDY

The present study aims to evaluate the effectiveness of WT programming in addressing juvenile delinquency and antisociality. To achieve this goal, we used systematic review to identify all eligible studies that evaluate the effects of WT programs on delinquent and antisocial behaviors, followed by meta-analysis to pool effect sizes from eligible studies to measure the overall effects of WT on such behaviors.

METHOD

Meta-analysis is a systematic, transparent, and reproducible process in which the population of relevant literature is identified and pooled quantitatively into an aggregate treatment estimate. The procedure involves converting individual study results into common measures of effect size, which are sensitive to both the direction and magnitude of effects, and prevent the reliance on individual study conclusions with respect to statistical significance. Criticisms of meta-analysis include that the methodological process involves a large number of decision-making steps that introduces subjectivity and potential validity concerns (see Ferguson & Kilburn, 2010; Ioannidis, 2016; Lakens et al., 2016).
SYSTEMATIC REVIEW

Inclusion Criteria

Nine inclusion criteria were developed for the current study; studies were assessed for eligibility based on these criteria by two reviewers. First, programs were required to be based entirely in remote settings using primitive shelters or facilities. Participants needed to spend at least three consecutive nights in the remote settings; this qualification was largely based on the importance of an unfamiliar environment in WT theories of change (Fernee et al., 2017; Walsh & Golins, 1976). Acceptable models were expedition or base-camp models; residential models were not eligible as these programs included elements that reduced the primitive element of treatment (such as community-based activities, academic credit and classes, or home visits) and were not comparable in program length (Johnson et al., 2020). Last, programs needed to be designed based on established theories with the purpose of facilitating therapeutic change, including elements such as individually tailored treatment plans and individual and/or group counseling sessions (Russell et al., 2018; Tucker et al., 2015).

Participants in eligible studies were predominantly youth aged 10 to 21 years; however, if the average (mean) age of participants was within the 10 to 21 years age range and a minority of participants were over 21, the study was still eligible for inclusion. Included studies must have presented a quantitative criminogenic outcome variable, such as antisocial behaviors, antisocial attitudes, problem behaviors, official reports of delinquency or recidivism, or unofficial reports (self-reports or caregiver reports) of delinquency or recidivism.

Eligible studies must have used a pretest to posttest research design, or a group contrast method. Specifically, studies could use randomized control trial designs, single-group pretest–posttest designs, or quasi-experimental designs using comparison groups. The Maryland Scale of Scientific Methods (SMS), a tool used to rate the internal validity and study quality of evaluations, was used to assess eligible studies for methodological rigor; only studies with a rating of 2 or higher on the SMS were included. The SMS rates studies from 1, for evaluations measuring correlation at one time point, to 5, for randomized control trials (Madaleno & Waights, 2015). In addition, studies included in the current analysis must have used samples with a minimum size of 15 participants in the treatment group.

To be included, studies must have been conducted in Canada, the United States, Western Europe, Australia, or New Zealand, and must have been published in English or German. These limitations were set to ensure researcher access and to limit the sample of programs and participants to those that are culturally similar and comparable. A date limiter was set to include studies published between January 1, 1990, and February 28, 2021. The purpose of this date limit was to provide an understanding of WT and delinquency that is more current than previous meta-analyses, and to sample a time frame that reflects modern standards and practice in WT. Both published and unpublished studies were eligible for inclusion.

Conversely, studies were not eligible for inclusion if the entire sample of participants belonged to a specific population due to program focus: for example, studies or programs exclusive to youth living on military bases, Aboriginal youth living on reserves, youth with specific mental illnesses (e.g., schizophrenia), youth with critical illnesses (e.g., cancer), or youth with substance use disorders. As previously noted, studies focused exclusively on target populations belonging to special populations were excluded to promote generalizability
of the meta-analytic results. However, youth belonging to specific populations could be included in eligible studies, as long as program formats were applicable to a general youth population. Last, studies were also excluded if they did not measure outcomes that were specifically and exclusively related to delinquency or antisocial attitudes.

**Search Constructs**

Four search constructs capturing population, program type, program features, and study designs eligible for inclusion were developed for the present study; terms were tested, added, deleted, expanded, narrowed, and modified until the search results were found to be sufficiently expansive to capture all potential studies. Constructs contained many synonyms and terms, as WT literature has been shown to use varied definitions and descriptors for treatment. Furthermore, the Boolean operators AND, OR, and NOT; parentheses; quotations; and a truncation symbol were used during the search process. The search terms were used in 27 electronic databases and date limiters were set from January 1, 1990, to February 28, 2021. See Table 2.

**Gray Literature Search**

*Hand searching.* Eleven journals were identified as relevant sources for WT evaluations. Other sources for hand searching included organization websites, the curriculum vitae (CV) of prominent authors, Google, and Google Scholar. Organization websites and prominent authors were identified during the initial review of the literature and further sources were identified during the electronic database search. Generally, if an author contributed to two or more studies that had been selected for retrieval, their CV was reviewed (authors were not contacted directly). A list of all the journals, author CVs, previous meta-analyses, and organization websites are included in Appendix A. Many gray literature sources do not permit the use of complex search constructs or date limiters. In this case, the following simplified search terms were utilized to conduct separate searches: “wilderness therapy,” “adventure therapy,” and “outdoor behavioral health care.” While searching Google and Google Scholar, these three terms were used as follows: “wilderness therapy” OR “adventure therapy” OR “outdoor behavioral health care.” The first 300 results were reviewed on Google and the first 800 results on Google Scholar were reviewed based on the recommendations of Haddaway et al. (2015). Where possible, date limiters were applied.

*Backward searching.* Backward searches were conducted extensively, although a record of the number of reference lists surveyed was not kept as this strategy was put into practice over the course of many months. In addition, backward searches were conducted on the reference lists of the five existing meta-analyses in the field.

**DATA COLLECTION AND ANALYSIS**

**Study Selection and Coding**

Two reviewers conducted identical searches on the electronic databases and any records flagged for further review were entered into a shared Excel spreadsheet. Studies flagged in the gray literature search, which was conducted by one reviewer, were also added to the Excel spreadsheet. This initial selection process was based on very general criteria, such as
publication date, program type, program location, and target population. Once a final list of flagged studies had been compiled, reviewers independently coded abstracts for full study retrieval by marking them with “retrieve” or “do not retrieve”; studies marked as “retrieve” by either reviewer were retrieved in full for further evaluation. If studies could not be retrieved directly, requests were made through our university’s Inter Library Loans (ILL) system.

After the initial set of studies had been retrieved and inclusion and exclusion criteria had been applied, study-level data were extracted. Throughout the coding process, any new records identified through backward searching were added to the Excel spreadsheet for agreement before retrieval. Coding was completed by two reviewers on 50 variables in the following categories: publication characteristics (e.g., year, location), program characteristics (e.g., program goals, components), intervention characteristics (e.g., program length, staffing), study characteristics (e.g., research design, SMS rating), sample characteristics (e.g., sample size, gender distribution), outcome measure (e.g., measurement tools, timing of data collection), general study conclusions, and outcome data (e.g., pre- and posttest scores, \( t \)-statistics). All coding done by the first reviewer was validated by the second reviewer. Disagreements on the coding of variables were discussed thoroughly to establish consensus on 100% of the data set. See Appendix B for the coding form.

**Effect Size Calculations**

To pool study effects, they must be comparable; this is achieved by ensuring that studies are measuring the same treatment effect and that effects are in the same standardized metric (Morris & DeShon, 2002). As effect sizes from different designs may not always measure the same population parameter, transformations must be made prior to study aggregation. Three types of effect size calculations were used in the current study: (a) modified Hedges’s \( g \) with pretest scores (one study), (b) Cox-adjusted odds ratio with pretest scores (one study), and (c) single-group pretest–posttest Cohen’s \( d \) with Morris and DeShon (2002) transformation applied (nine studies). Details on effect size calculation formulas are available in Beck (2021).

With respect to the single-group pretest–posttest studies, there have been mixed opinions on methodological criteria and the combination of single-group designs with other designs in meta-analyses; mainly, concerns surround the potential negative impact of single-group
pretest–posttest studies on the quality of findings in a meta-analysis (see Borenstein & Hedges, 2019; Cuijpers et al., 2017; Lipsey & Wilson, 2001). However, in many fields of research it is often impractical for randomized control trials or two-group designs to be implemented. In these situations, meta-analyses excluding studies with weaker designs may result in biased results which ignore numerous program evaluations (Lee & Wong, 2021). Single-group designs are very popular in the field of WT research, wherein youth are often deemed at-risk and in need of effective intervention. In these cases, withholding treatment from youth to implement randomized and two-group designs may be unethical and impractical. In the current study, we contend that combining single-group pretest–posttest studies with two-group designs was justified as representing WT research more thoroughly. Furthermore, numerous techniques were implemented to maintain conceptual continuity among effect sizes (see Morris & DeShon, 2002).

**Outcome Measures**

The current study included two outcome measures: self-reported delinquency and caregiver-reported delinquency. Measures of self-reported delinquency focused on behaviors such as truancy, substance use, inappropriate sexual activities, lying, stealing, fighting, or damaging property. The most common instrument used was the Social Problems subscale on the Y-OQ-SR. Questions on the Y-OQ-SR are rated on a 5-point Likert-type scale, asking youth about their delinquent behaviors ranging from 0 (never) to 4 (almost always). Measures of caregiver-reported delinquency were focused on the same behaviors; the sources of caregiver reports were parents, guardians, program instructors, and counselors. The Y-OQ was the most common instrument used to measure caregiver-reported delinquency, which includes very similar questions to the Y-OQ-SR and uses the same Likert-type scale. All available times of measurement (pretest, posttest, and follow-up) were recorded during the coding process.

While official measures of police contact, arrests, charges, and/or convictions were initially planned for inclusion, following article coding it was evident that there were insufficient studies with officially reported data to enable a meaningful separate meta-analysis. As official reports were not deemed commensurate with self-reported behavior and thus could not be combined to create a larger pool of studies, they were excluded from the current analysis. Given that self-reported data are known to underreport delinquent behaviors, to provide a more balanced picture of the relationship between WT and delinquency the current study included caregiver reports that tend to document higher levels of problem behaviors (Kroner et al., 2007; Salbach-Andrae et al., 2009).

**Decision Rules**

To ensure independence of effect sizes, we implemented three decision rules during data collection: (a) When multiple delinquency outcome measures were present in a single study, the outcome deemed to be the most conceptually and structurally similar to the rest of the study pool was included. (b) For studies that evaluated multiple programs within the same study, if enough details on each program site were presented then the study was eligible for inclusion, either as one effect or multiple effects. However, if the program characteristics of each included site were not reported adequately, the study (or site) was excluded. (c) If
multiple reports of the same study were identified, such as a dissertation and a journal article, the document with the most comprehensive information was selected for inclusion.

**Dealing With Missing Data**

Missing data were common across the identified studies—in many cases, this resulted in study exclusion from the analysis set. For example, studies were excluded for not reporting standard deviations of mean scores or for not reporting necessary details surrounding treatment or program components. If contact information for researchers was available, requests for missing data were sent and, at times, fulfilled (e.g., Russell et al., 2018). Where data were missing but contextual clues were present, coders replaced missing data if they were approximately 85% sure of a code. This rule was only used for the coding of program components or report details (e.g., parent involvement in the program)—not for outcome measures or effect size data.

**Data Synthesis**

The statistical model for meta-analysis is generally chosen based on an assumption of the origin of heterogeneity in a pool of studies and on the desired level of generalization (Card, 2011). The DerSimonian and Laird random-effects model was used in the current study, as heterogeneity in the study pool is assumed to come from differences in treatment, evaluation methods, study implementation, or other variables in addition to sampling error differences (Lipsey & Wilson, 2001). Forest plots were used to display outcomes.

**Publication Bias**

Publication bias arises when published research on a given topic reports very different outcomes than unpublished reports (such as statistically significant vs. null results), resulting in a skewed understanding of the effect of a phenomenon (Vevea et al., 2019). The current study sought to identify publication bias using funnel plots and Egger’s test of small-study effects.

**Influence Analysis**

We tested for sensitivity of the findings to potential outliers by conducting a remove-one-study influence analysis. To do so, each study in the meta-analysis is omitted, one at a time, and the pooled effect is recalculated without that study to determine whether its removal would have a notable impact on the pooled findings of the meta-analysis. In addition, as a robustness test, analyses were run with single-group and two-group designs separated.

**Assessment of Heterogeneity**

To examine heterogeneity among the included effects sizes, $Q$-statistics and $I^2$ statistics were examined. The $Q$-test assesses the presence of homogeneity, whereas the $I^2$ index is an indicator of the magnitude of heterogeneity in a set of studies.

**RESULTS**

**SEARCH RESULTS**

Searching the 27 electronic databases resulted in a hit count of 1,977 records. After both reviewers surveyed the titles and abstracts, 152 results were compiled in an Excel spreadsheet
for further review. An additional 37 records were identified during the gray literature search, for a total of 189 records. Forty-three of the 189 records were not retrieved due to not meeting basic inclusion criteria such as location, population type, or date, leaving 146 articles that were selected for retrieval. All studies were retrieved or requested from ILL and inclusion and exclusion criteria were applied, resulting in the exclusion of 135 studies—11 of which could not be retrieved (see Appendix D for a list of exclusion codes and counts). A final pool of 11 studies was included in the analysis; nine of the 11 studies contributed effect sizes to the self-reported delinquency outcome, whereas five studies contributed effect sizes to the caregiver-reported delinquency outcome. The PRISMA flow diagram in Figure 1 represents the process of the systematic review and study selection (Page et al., 2021).

CHARACTERISTICS OF INCLUDED STUDIES

All but two studies characterized their treatment group participants as “at-risk”; programs primarily enrolled participants through parent, clinical, community, and justice agency referrals, although several studies did not provide this information. All programs were conducted in the United States except one, which was located in Australia. Program duration ranged from 10 to 90 days and most included multiple forms of individual, group, or family therapy. Participants were predominantly male and Caucasian and participant age ranged from 11 to 26 years. Activities in the programs were diverse, although most involved hiking, camping, canoeing, rock-climbing, and shelter building and survival activities. Posttest measures taken at program discharge were used for all but one study (which provided only a follow-up). An overview of the studies and programs included in the current study is shown in Table 3 and details on study-level information are presented in Appendix C.

META-ANALYSIS OF SELF-REPORTED DELINQUENCY

Nine independent effect sizes were included in the meta-analysis of self-reported delinquency. The random-effects model produced a large, positive, and significant effect of 0.832 ($Z = 5.103$, $p < .001$). This finding indicates beneficial effects of WT programming on self-reported delinquency among youth. Nearly all of the individual studies reported positive, significant, and moderate to large effect sizes, with the exception of two studies that reported negative and nonsignificant effect sizes (Brand, 1998 and Deschenes et al., 1996). The $Q$-statistic of 109.41 ($df = 8$, $p < .001$) shows a significant level of heterogeneity within the pooled studies and the $I^2$ statistic suggests that a majority (92.7%) of this heterogeneity can be attributed to differences other than random variation.

A forest plot of the individual effects with their associated 95% confidence intervals (CIs) and relative weight in the model is presented in Figure 2; the diamond at the bottom represents the pooled effect size of 0.832 (95% CI = [0.513, 1.152]). Weights were derived using a random-effects model. The forest plot suggests that both peer review and date may influence effect sizes; for example, the two oldest and the only negative self-reported delinquency effects are non-peer-reviewed studies that were implemented during the key era of change in WT standards.

Publication Bias

The funnel plot (see Appendix E) shows the proximity of each study to the pooled effect size based on the precision of each effect. Five studies fall outside of the pseudo 95% CI, indicating the potential for publication bias and for outliers. However, Egger’s test
Figure 1: PRISMA Flow Diagram\textsuperscript{a}

\textsuperscript{a} See Appendix D for exclusion reason codes.
of small-study effects produced a statistic of $-0.817$ ($SE = 2.241$, $t = -0.36$, $p = .726$), indicating nonsignificant asymmetry and suggesting that publication bias or small-study effects are not likely a concern.

| Characteristics                  | Self-reports ($N = 9$) | Caregiver reports ($N = 5$) |
|----------------------------------|------------------------|----------------------------|
| **Publication type**             |                        |                            |
| Journal article                  | 6 (66.67)              | 2 (40.00)                  |
| Report                           | 1 (11.11)              | 3 (60.00)                  |
| Dissertation/thesis             | 2 (22.22)              |                            |
| **Publication year (range 1996–2020)** |                      |                            |
| 1990–2012                        | 3 (33.33)              | 3 (60.00)                  |
| 2013–2020                        | 6 (66.67)              | 2 (40.00)                  |
| **Type of therapy**             |                        |                            |
| Individual                       | 7 (77.78)              | 4 (80.00)                  |
| Group                            | 7 (77.78)              | 3 (60.00)                  |
| Family                           | 2 (22.22)              | 1 (20.00)                  |
| **Program model**               |                        |                            |
| Expedition                       | 6 (66.67)              | 4 (80.00)                  |
| Base-camp                        | 2 (22.22)              | 1 (20.00)                  |
| Missing                          | 1 (11.11)              |                            |
| **Program length**              |                        |                            |
| 10–21 days                       | 2 (22.22)              | 1 (20.00)                  |
| Longer than 3 weeks              | 5 (55.56)              | 4 (80.00)                  |
| Flexible duration                | 2 (22.22)              |                            |
| **Type of staff**                |                        |                            |
| Clinical staff                   | 7 (77.78)              | 4 (80.00)                  |
| Wilderness/recreation specialists | 1 (11.11)              | 1 (20.00)                  |
| Other program facilitators      | 1 (11.11)              |                            |
| **Parent participation**         |                        |                            |
| No                               | 7 (77.78)              | 3 (60.00)                  |
| Yes (for <3 days of the program) | 2 (22.22)              | 2 (40.00)                  |
| **Research design**             |                        |                            |
| Single-group pre–post            | 7 (77.78)              | 5 (100.00)                 |
| Independent groups pre–post      | 2 (22.22)              |                            |
| **M treatment n at pretest**     | 172.11 participants    | 65 participants            |
| Range at pretest                 | 32–816 participants    | 19–189 participants        |
| **Age range of sample**          | 11–26 years            | 13–26 years                |
| **Gender of analysis sample**    |                        |                            |
| Mixed (31%–69%)                  | 4 (44.44)              | 3 (60.00)                  |
| Mostly male (70%–89%)            | 3 (33.33)              | 2 (40.00)                  |
| All male (90%+)                  | 1 (11.11)              |                            |
| Missing                          | 1 (11.11)              |                            |
| **Ethnic composition of sample** |                        |                            |
| 70%+ Caucasian                   | 5 (55.56)              | 3 (60.00)                  |
| Mixed/minority (<70% Caucasian)  | 3 (33.33)              | 1 (20.00)                  |
| Missing                          | 1 (11.11)              | 1 (20.00)                  |

*Multiple types of therapy could be utilized; therefore, percentages for “type of therapy” do not sum to 100%.*
To further test for bias, an influence analysis was conducted; no individual study’s effect impacted the pooled effect drastically enough to affect the positive, significant pooled result. Influence analysis results are available in Appendix E. We also implemented the analysis with the 2 two-group design studies dropped from the study set, leaving seven single-group pretest–posttest designs in the study pool. The overall effect size for this analysis was 1.077 (Z = 7.221, p < .001); while this is larger than the pooled effect including the two-group designs, the substantive conclusions remain unchanged.

**META-ANALYSIS OF CAREGIVER-REPORTED DELINQUENCY**

The meta-analysis of caregiver-reported delinquency included five independent studies and used a random-effects model. A large, positive, and significant effect size of 1.054 (Z = 3.171, p < .003) was found. Heterogeneity was large and significant as evidenced by a Q-statistic of 81.60 (df = 4, p < .001), with most of the heterogeneity resulting from between-study differences (I^2 = 95.1%). As shown in Figure 3, each individual study produced a positive and significant effect and the overall pooled effect indicates that WT programming is effective in reducing caregiver-reported delinquency among youth.
Weights are calculated using a DerSimonian and Laird random-effects model. Similar to the meta-analysis of self-reported delinquency, non-peer-reviewed, older studies tended to report smaller or negative effects, whereas more recent, peer-reviewed studies reported larger effects.

**Publication Bias**

The funnel plot (see Appendix E) shows that four of the five studies fell outside of the pseudo 95% CI; however, the plot is relatively symmetric and does not indicate any major outliers. Egger’s test of small-study effects produced a statistic of 3.067 ($SE = 8.204$, $t = 0.37$, $p = .733$), indicating nonsignificant asymmetry and suggesting that bias in the current meta-analysis is not likely to be a result of publication bias or small-study effects.

**Influence Analysis**

The influence analysis of the caregiver-reported delinquency studies showed that no individual study’s effect influenced the pooled effect substantially enough to affect the positive, significant pooled result. Results are available in Appendix E.

---

**Figure 3: Forest Plot for Caregiver-Reported Delinquency**

Weights are calculated using a DerSimonian and Laird random-effects model. Similar to the meta-analysis of self-reported delinquency, non-peer-reviewed, older studies tended to report smaller or negative effects, whereas more recent, peer-reviewed studies reported larger effects.

**Publication Bias**

The funnel plot (see Appendix E) shows that four of the five studies fell outside of the pseudo 95% CI; however, the plot is relatively symmetric and does not indicate any major outliers. Egger’s test of small-study effects produced a statistic of 3.067 ($SE = 8.204$, $t = 0.37$, $p = .733$), indicating nonsignificant asymmetry and suggesting that bias in the current meta-analysis is not likely to be a result of publication bias or small-study effects.

**Influence Analysis**

The influence analysis of the caregiver-reported delinquency studies showed that no individual study’s effect influenced the pooled effect substantially enough to affect the positive, significant pooled result. Results are available in Appendix E.
DISCUSSION

The current study used systematic review and meta-analysis to evaluate the effectiveness of WT programming at reducing self- and caregiver-reported delinquency. Eleven studies reporting 14 effect sizes met inclusion criteria and results indicate that WT programs are effective in reducing both self- and caregiver-reported delinquent behaviors among youth.

WT programming has generally focused on skills-building tasks, such as creating shelter, building fires, and other survival-based skills, as well as recreational activities that require mental and physical endurance. Theoretical considerations of WT have frequently attributed its effectiveness in addressing problem behaviors to the sense of mastery and self-efficacy gained by engaging in these activities, which may be criticized for not adhering to accepted models such as RNR (Latessa et al., 2002; Russell, 2001). However, more recent explanations point toward a psychotherapeutic model or positive social interactions (Fernée et al., 2017; Gass et al., 2012). For example, “solo” time, in which participants complete short (~3-day) excursions without their group, has been attributed to beneficial program outcomes (Russell & Phillips-Miller, 2002). Routine outcome monitoring has been suggested as a way to improve understandings of the process of change in WT (Dobud et al., 2020).

Previous meta-analyses have mostly cited significant benefits of WT: for example, improvements in self-esteem, behavior, and recidivism outcomes (Bedard, 2004; Bettmann et al., 2016; Bowen & Neill, 2013; Gillis, Speelman, et al., 2016; Wilson & Lipsey, 2000). Findings from the present study support prior evidence of reductions in self-/caregiver-reported delinquent behaviors; however, the current study presents stronger effects than previously found in Wilson and Lipsey’s (2000) and Bedard’s (2004) meta-analyses. This concurs with Bowen and Neill’s (2013) findings that effect sizes in WT evaluations have increased from the 1960s to the 2000s, perhaps indicating overall improvements in WT programming. The current study also identified distinct findings for self- versus caregiver-reported delinquency; as such, the commensurability of outcome measures should be carefully considered. The difference found between these measures is in line with evidence that suggests self-reported outcomes among youth may underreport problem behaviors (Kroner et al., 2007; Salbach-Andrae et al., 2009). Gillis, Speelman, et al.’s (2016) meta-analysis of Y-OQ caregiver-report and Y-OQ-SR total scores found a similar distinction, wherein larger reductions in dysfunction were reported by caregivers.

LIMITATIONS

The present meta-analysis had several limitations. First, the pool of studies was small, in part due to the fairly strict requirements with respect to WT program type, and in part because limited quantitative evaluation research exists. Limited available data in empirical WT literature also restricted the current study to “soft” outcomes of delinquency, such as lying, fighting, and stealing, as opposed to “hard” outcomes of delinquency such as arrests or criminal convictions, which may be considered more relevant in criminal justice research. Second, we did not restrict our inclusion criteria to two-group designs. As discussed previously, an ongoing debate in the literature addresses the appropriateness of pooling effect sizes from single-group and two-group designs in meta-analysis. Given the preponderance of single-group designs identified through our systematic search, we believe that including
these studies permits a more comprehensive examination of the literature with respect to the effectiveness of WT programs.

Third, many studies did not report certain data that were a requirement for inclusion (e.g., standard deviations or sample sizes). Although authors were contacted for information, very few were able to provide missing data. Fourth, the current meta-analysis did not include primary studies that were focused on specific programs exclusive to military youth or Indigenous youth, youth with substance use disorders, or youth with critical illnesses and as such results are not generalizable to these programs. The present study sought to include only those WT programs that could be considered comparable; therefore, a trade-off was made in which the study pool remained small to maintain a sample that was commensurate.

Last, limited by a small-study pool, we were unable to include any moderator analyses (such as program components and participant characteristics) or times of measurement other than pretest to posttest outcomes. With respect to the latter, although additional times of measurement were coded where available, few studies presented this information, resulting in a notable gap in understanding of the durability and longevity of WT programming effects. The absence of follow-up data prompts questions about whether the effects of excursion-based programs fade once participants return home. With respect to the former, characteristics of the treatment participants in the study pool were sparse, as many evaluations fail to include data on factors like referral reasons. This leaves unanswered the question of for whom WT programs are most effective and prompts further investigation.

CONCLUSION AND RECOMMENDATIONS

Despite the limitations of a small-study pool and the lack of a moderator analysis, results indicate that WT programming is a potentially viable treatment for at-risk youth. Further research is needed to apply this knowledge to policy and practice in useful ways. Evaluators and program operators can further our knowledge of the impact that WT programming has on at-risk youth by including follow-up measures, coding specific program components and activities, and providing additional quantitative data such as subscale scores. Outcomes that indicate more serious delinquency such as negative police contacts, arrests, and convictions should be examined in future evaluations of WT programming to provide a clearer understanding of how WT can be used in the criminal justice system. In addition, gender, age, participant backgrounds, treatment duration, therapeutic alliance, and various program components (such as solo expeditions or journaling) are important variables that should be examined as predictors of treatment outcomes. Future research should also aim to uncover a meaningful understanding of which participants WT is most applicable to and in what scenarios. Program components that most strongly affect the process of change can be further developed and possibly applied to other interventions when addressing at-risk youth, and components that are costly but have little or no effect on outcomes can be removed. As delinquent behaviors can signal higher risks of criminal involvement and contribute to cumulative disadvantages among youth, effective treatments such as WT should continue to be developed and evaluated to help at-risk youth.
## APPENDIX A

### TABLE A1: Literature Search Sources

| Source type                  | Source name                                                                 |
|------------------------------|----------------------------------------------------------------------------|
| **Electronic databases**     | Academic Search Premier                                                    |
|                              | Canadian Research Index                                                    |
|                              | Cochrane Central Register for Controlled Trials                           |
|                              | Cochrane Database of Systematic Reviews                                   |
|                              | Criminal Justice Abstracts                                                 |
|                              | Database of Abstracts of Reviews and Effects                               |
|                              | EBM Reviews Full Text                                                      |
|                              | EBSCO Open Dissertations                                                   |
|                              | Education Source                                                           |
|                              | ERIC (EBSCO)                                                               |
|                              | Government of Canada Publications                                          |
|                              | Medline                                                                    |
|                              | National Criminal Justice Reference Service (NCJRS)                        |
|                              | Open Access Theses and Dissertations                                       |
|                              | ProQuest Dissertations and Theses Abstracts Index                          |
|                              | ProQuest Sociology Collection                                               |
|                              | ProQuest Sociology Database                                                |
|                              | PsycARTICLES                                                              |
|                              | PsycBooks                                                                  |
|                              | PsycINFO                                                                   |
|                              | PubMed Central                                                            |
|                              | Social Sciences Abstracts                                                  |
|                              | Social Sciences Full Text                                                  |
|                              | Sociological Abstracts                                                     |
|                              | Social Services Abstracts                                                  |
|                              | Theses Canada                                                              |
|                              | Web of Science                                                             |
| **Hand-searched journals**   | Australian Journal of Outdoor Education                                  |
|                              | Child & Youth Care Forum                                                   |
|                              | Journal of Adolescent and Family Health                                    |
|                              | Journal of Adventure Education and Outdoor Learning                        |
|                              | Journal of Child and Family Studies                                       |
|                              | Journal of Experiential Education                                          |
|                              | Journal of Outdoor Recreation, Education, and Leadership                   |
|                              | Journal of Outdoor and Environmental Education                            |
|                              | Journal of Research in Crime and Delinquency                              |
|                              | Journal of Therapeutic Schools and Programs                                |
|                              | Residential Treatment for Children and Youth                              |
| **Organization websites**     | American Psychological Association                                         |
|                              | American Society of Criminology                                            |
|                              | Association of Experiential Education (AEE)                                |
|                              | Australian Association for Bush Adventure Therapy                          |
|                              | Center for Court Innovation                                                |
|                              | Crimesolutions.gov list of evidence-based programs and practices           |
|                              | Department of Justice (DOJ)—Canada                                        |
|                              | Department of Justice (DOJ)—United States                                 |
|                              | Interagency Working Group on Youth Programs (IWGYP)                        |
|                              | National Association of Therapeutic Schools and Programs (NATSAP)          |
|                              | National Council on Crime & Delinquency (NCCD)—including the Children’s   |
|                              | Research Center                                                            |
|                              | National Institute of Justice (NIJ)                                        |
|                              | Office of Juvenile Justice and Delinquency Prevention (OJJDP)              |
|                              | Outdoor Behavioral Healthcare Council (Research Cooperative)               |
|                              | Outward Bound (Journal of Education volumes)                              |
|                              | Prime Minister’s Youth Council                                             |
|                              | Public Health Agency of Canada’s Best Practices Portal                     |
|                              | Public Health Institute at Liverpool                                       |
|                              | Public Safety Canada                                                       |
|                              | SAMHSA’s National Registry of Evidence-Based Programs and Practices       |
|                              | United States Department of Agriculture (Treeresearch)                     |

(continued)
TABLE A1. (continued)

| Source type                           | Source name                                                                 |
|---------------------------------------|-----------------------------------------------------------------------------|
| Researcher CVs (n = 29)               | Dene S. Berman, Joanna E. Bettmann, Daniel J. Bowen, Dell Brand, David D. Christian, Jeffrey P. Clark, Katie M. Combs, Jennifer Davis-Berman, Steven M. DeMille, Elizabeth P. Deschenes, Mathew D. Deurden, H. Preston Elrod, Michael A. Gass, Harold L. Gillis, Nevin J. Harper, Matthew J. Hoag, Bruce A. Larson, Sarah F. Lewis, Timothy A. Lowe, Kevin I. Minor, James T. Neill, Christine L. Norton, Pamela M. Orren, Ivan J. Raymond, Keith C. Russell, Robert L. Sveen, Anita R. Tucker, Michael A. Walsh, Anja Whittington |
| Previous meta-analyses (n = 5)        | Bedard (2004), Bettmann et al. (2016), Bowen and Neill (2013), Gillis, Speelman, et al. (2016), Wilson and Lipsey (2000) |

APPENDIX B

TABLE B1: Coding Form

| Publication characteristics          |                                                                 |
|--------------------------------------|------------------------------------------------------------------|
| Author (date)                        | String variable (e.g., Wichmann, 1990)                           |
| Outcome number                       | No specification (1, 2, 3, etc.) used when one study contributed more than one outcome effect |
| Targeted population                  | 0 = Referred/selected for program 1 = Open to all youth         |
| Publication year                     | No specification (1995, 2001, etc.)                              |
| Publication type                     | 0 = Journal article 1 = Book chapter 2 = Report 3 = Dissertation/thesis |
| Peer reviewed                        | 0 = No 1 = Yes                                                   |
### TABLE B1. (continued)

| Program characteristics |  |
|-------------------------|--|
| **Program name**        | String variable (e.g., Alaska Crossings) |
| **Program location**    |  |
| 0 = North America       |  |
| 1 = Europe              |  |
| 2 = Australia/New Zealand |  |
| **Program delivery year** | No specification (1998, 2010, etc.) |
| **Program type**        |  |
| 0 = Wilderness without formal therapy |  |
| 1 = Wilderness with formal therapy |  |
| 2 = School integrated program |  |
| 3 = Ropes course        |  |
| 4 = Adventure therapy (not wilderness) |  |
| 5 = Wilderness therapy integrated with family program |  |
| 6 = Mixed program type  |  |
| **Curriculum**          |  |
| 0 = Original            |  |
| 1 = Adapted from another program |  |
| **Program description** | String variable (e.g., 3-week backcountry wilderness expedition) |
| **Program activities**  | String variable (e.g., hiking, canoeing, camping) |
| **Topics covered**      | String variable (e.g., communication, goal setting) |
| **Program purpose/goals** | String variable (e.g., behavior management) |
| **Program model**       |  |
| 0 = Expedition          |  |
| 1 = Base-camp            |  |
| **Wilderness primary (vs. integrated into a larger program modality?)** |  |
| 0 = No                  |  |
| 1 = Yes                 |  |
| **OBH Council certification** |  |
| 0 = No                  |  |
| 1 = Yes                 |  |
| **Clinical staff**      |  |
| 0 = No                  |  |
| 1 = Yes                 |  |
| **Individual therapy**  |  |
| 0 = No                  |  |
| 1 = Yes                 |  |
| **Group therapy**       |  |
| 0 = No                  |  |
| 1 = Yes                 |  |
| **Family therapy**      |  |
| 0 = No                  |  |
| 1 = Yes                 |  |
| **Program duration**    | No specification (e.g., 21 days, 8 weeks) |
| **Number per group**    | No specification (e.g., 8 youth) |
| **Program delivered by?** |  |
| 1 = Clinical staff      |  |
| 2 = Master’s students   |  |
| 3 = Wilderness specialists |  |
| 4 = Other professional staff |  |
| 5 = General program facilitators |  |
| **Parent involvement**  |  |
| 0 = No                  |  |
| 1 = Yes, for <3 days of the program (e.g., debriefing days) |  |
| 2 = Yes, for 3 days or more of the program |  |

### Study design characteristics

| Research design |  |
|-----------------|--|
| 0 = Randomized control trial |  |
| 1 = Quasi-experiment with matched comparison group |  |
| 2 = Quasi-experiment with weakly matched comparison group |  |
| 3 = Single-group pretest–posttest |  |
| 4 = Minimally exposed cohort as comparison |  |

(continued)
### Study design characteristics

| Maryland SMS rating | 1 | Posttest only |
|---------------------|---|---------------|
|                     | 2 | Single-group pretest–posttest |
|                     | 3 | Quasi-experiment with weakly matched control group |
|                     | 4 | Quasi-experiment with strongly matched control group |
|                     | 5 | Randomized control trial |

| Unit of assignment | 0 | Individual |
|--------------------|---|------------|
|                    | 2 | Group      |

| Control group type | 0 | No control |
|--------------------|---|------------|
|                    | 1 | Nonmatched |
|                    | 2 | Matched    |
|                    | 3 | Waitlist control |
|                    | 4 | RCT control |

| Researcher involvement | 0 | Evaluation only |
|------------------------|---|-----------------|
|                        | 1 | Involved in delivering intervention |
|                        | 2 | Involved in developing intervention |
|                        | 3 | Developed and delivered intervention |

| Pretest | 0 | No |
|---------|---|----|
|         | 1 | Yes |

| Size of treatment group | No specification (e.g., n = 64) |
|-------------------------|---------------------------------|
| Size of control group   | No specification (e.g., n = 23) |
| Attrition from sample   | No specification (e.g., 26% from pretest to posttest) |
| Age range               | No specification (e.g., 14–21 years) |
| M age                   | No specification (e.g., 18.2 years) |
| Standard deviation of mean age | No specification (e.g., 0.9 years) |

| Gender mix | No specification (e.g., 74% male) |
|------------|----------------------------------|
| Ethnicity mix | No specification (e.g., 82% White) |

| At-risk sample | 0 | No |
|----------------|---|----|
|                | 1 | Yes |

| Direction and magnitude of differences between treatment and control groups | No specification (e.g., treatment group significantly higher in proportion male than control group) |

### Outcome measures

| Outcome measure name | String variable (e.g., self-reported delinquency) |
|----------------------|---------------------------------------------------|
| Direction of measure | 0 = Increase in score is good 1 = Increase in score is bad |
| Preexisting tools used to measure outcome | String variable (e.g., Y-OQ) |

| Source of measure | 0 = Self-report 1 = Parent report 2 = Counselor report 3 = Official report |
|-------------------|-------------------|
| Measurement       | 0 = Continuous 1 = Dichotomous |

| Time of posttest | No specification (e.g., on last day of program) |
|------------------|-----------------------------------------------|
| Time of follow-up(s) | No specification (e.g., 6 months following posttest) |

| Findings | String variable (e.g., significant reduction in social problems from pretest to posttest, no significant differences from posttest to 6-month follow-up) |

*Note. OBH = outdoor behavioral health care; SMS = Scale of Scientific Methods; Y-OQ = Youth-Outcome Questionnaire.*
## APPENDIX C

### TABLE C1: Study-Level Details

| Author (date) | Program model | Therapy type | Program length (days) | SMS scale | TX age M (range) | Post/follow-up used | Sample size N | TX gender | TX ethnicity | Outcome (instrument) |
|---------------|---------------|--------------|-----------------------|-----------|------------------|---------------------|---------------|-----------|--------------|----------------------|
| Bettmann et al. (2013) | Base-camp | Group | $M = 64.7$ | 2 | 15.8 (missing) | Discharge | Pre: 40 | 66% female | 82% Caucasian, 18% other | Social problems (Y-OQ) |
| Brand (1998) | Expedition | Unspecified | 10 | 3 | Missing (11–13) | Discharge | Pre: 70 | All male | Missing | Conduct disorder (Jessor & Jessor) |
| Deschenes et al. (1998) | Base-camp | Individual, group, family | 90 | 4 | 16.5 (14–18) | Follow-up (24 months) | Pre: 48 | Missing | 64% African American, 30% Caucasian, 6% other | Delinquency (Elliott et al., 1985) |
| Hagan (2002) | Expedition | Individual, group, family | 42 | 2 | Missing (13–17) | Discharge | Pre: 19 | 63% male | 95% Caucasian, 5% other | Social problems (Y-OQ) |
| Johnson et al. (2020) | Expedition | Individual | 70–84 | 2 | 15.4 (13–17) | Discharge | Pre: 816 | 59% male | 69% Caucasian, 21% other | Social problems (Y-OQ) |
| Lewis (2013) | Expedition | Individual, group | $M = 57.5$ | 2 | 15.7 (13–17) | Discharge | Pre: 190 | 66% male | 87% Caucasian, 13% other | Conduct subscale (Child TOP) |
| Russell et al. (2018) | Expedition | Individual, group | $M = 58.34$ | 2 | 15.6 (12–17) | Discharge | Pre: 78 | 71% male | 52% Indigenous, 38% Caucasian, 10% other | Conduct problems (Y-OQ-30.2) |
| Tucker et al. (2015) | Expedition | Individual, group | 63 | 2 | 15.8 (13–17) | Discharge | Pre: 63 | 50% male | 87% Caucasian, 13% other | Social problems (Y-OQ) |
| Tucker et al. (2016) | Expedition | Individual, group, family | $M = 79.8$ | 2 | 16.2 (13–18) | Discharge | Pre: 212 | 70% male | 79% Caucasian, 21% other | Social problems (Y-OQ) |
| Turner (2009) | Expedition | Individual, group | 21 | 2 | 18.2 (14–26) | Discharge | Pre: 32 | 89% male | 83% Caucasian, 17% other | Social problems (Y-OQ) |
| Wichmann (1990) | Base-camp | Group | 30 | 2 | 15.5 (13–18) | Discharge | Pre: 36 | Missing | Missing | Acting out (WABIS) |

**Note.** SMS = Maryland Scale of Scientific Methods; Y-OQ = Youth-Outcome Questionnaire.

*aSMS score: Level 2. Temporal sequence between the program and the crime or risk outcome clearly observed, or the presence of a comparison group without demonstrated comparability to the treatment group. Level 3. A comparison between two or more comparable units of analysis, one with and one without the program. Level 4. Comparison between multiple units with and without the program, controlling for other factors, or using comparison units that evidence only minor differences (Sherman et al., 1998, p. 5). TX = treatment group; TOP = Youth Version of the Treatment Outcome Package; WABIS = Wichmann-Andrew Behavior Intervention Scale.*
## APPENDIX D

### Table D1: Exclusion Reason Codes

| Code label                  | Code description                                                                                                          | Frequency n (%) |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------|-----------------|
| Could not retrieve          | Study identified as relevant and selected for further screening, but could not be retrieved through Interlibrary Loans or by contacting the study author | 11 (8.2)        |
| Date                        | Study published prior to 1990                                                                                           | 0 (0)           |
| Sample size                 | Sample size of less than 15 participants in the treatment group                                                          | 9 (6.7)         |
| Irrelevant outcomes         | Outcomes evaluated were not related to delinquency, or subscale scores for measures including delinquency were not reported | 53 (39.3)       |
| Cannot calculate effect size| Did not report data necessary to calculate an effect size, for example, standard deviation of the mean                   | 23 (17)         |
| Program does not qualify    | Program did not fit the operational definition of wilderness therapy used in the current study, such as ropes courses or sail training programs | 22 (16.3)       |
| Overlapping samples         | Used a sample which was used in another relevant evaluation                                                              | 7 (5.2)         |
| Sample too specific a subgroup | Program focused on subgroups which restricted generalizability to a greater youth population; for example, programs focused on youth with substance use disorders, terminal illnesses, or disabilities | 2 (1.5)         |
| Missing data                | Study was missing information necessary for study inclusion; for example, when no description of program components, implementation, name, or setting were reported and information could not be found through related sources | 8 (5.9)         |

Note. Exclusion codes were applied in hierarchical order; for example, if a study had a sample size of less than 15 participants and reported on a sample that was a specific subgroup, the exclusion code applied was “sample size.” The hierarchy of codes is presented here from top to bottom, with “could not retrieve” being the most dominant code and “inappropriate control” being the least dominant code.

## APPENDIX E

### Funnel Plots and Influence Analysis Results

**Meta-Analysis of Self-Reported Delinquency**

![Funnel Plot for Self-Reported Delinquency](image)

**Figure E1:** Funnel Plot for Self-Reported Delinquency:
**TABLE E1: Influence Analysis on Self-Reported Delinquency**

| Author (date)         | Effect size | 95% confidence interval |
|----------------------|-------------|-------------------------|
| Bettmann et al. (2013) | 0.756       | 0.423 - 1.090           |
| Brand (1998)         | 0.948       | 0.643 - 1.253           |
| Deschenes et al. (1996) | 0.941     | 0.624 - 1.257           |
| Johnson et al. (2020) | 0.812       | 0.355 - 1.269           |
| Lewis (2013)         | 0.862       | 0.484 - 1.240           |
| Russell et al. (2018) | 0.778       | 0.430 - 1.126           |
| Tucker et al. (2015)  | 0.804       | 0.451 - 1.157           |
| Tucker et al. (2016)  | 0.718       | 0.423 - 1.013           |
| Turner (2009)        | 0.862       | 0.515 - 1.208           |
| Combined             | 0.832       | 0.512 - 1.152           |

---

**Figure E2: Influence Analysis on Self-Reported Delinquency**

**Meta-Analysis of Caregiver-Reported Delinquency**

**Figure E3: Funnel Plot for Caregiver-Reported Delinquency**
Table E2: Influence Analysis on Caregiver-Reported Delinquency

| Author (date)   | Effect size | 95% confidence interval |
|-----------------|-------------|-------------------------|
| Hagan (2002)    | 1.215       | [0.434, 1.997]          |
| Johnson et al. (2020) | 0.854   | [0.315, 1.392]          |
| Tucker et al. (2015) | 0.924   | [0.175, 1.672]          |
| Turner (2009)   | 1.033       | [0.271, 1.794]          |
| Wichmann (1990) | 1.234       | [0.510, 1.958]          |
| Combined        | 1.054       | [0.403, 1.706]          |

Figure E4: Influence Analysis on Caregiver-Reported Delinquency

ORCID iD
Jennifer S. Wong https://orcid.org/0000-0001-7523-7430

References

Studies included in the self-report meta-analysis are marked with one asterisk*, studies in the caregiver report analysis are marked with two asterisks**, and studies in both meta-analyses are marked with three asterisks***.

Beck, N. (2021). The effects of wilderness therapy on delinquent behaviors among youth: A systematic review and meta-analysis [Unpublished Honors thesis]. Simon Fraser University.

Bedard, R. M. (2004). Wilderness therapy programs for juvenile delinquents: A meta-analysis [Doctoral dissertation, Colorado State University] (Publication No. 3143810). ProQuest Dissertations Publishing.

Bettmann, J. E., Gillis, H. L., Speelman, E. A., Parry, K. J., & Case, J. M. (2016). A meta-analysis of wilderness therapy outcomes for private pay clients. *Journal of Child and Family Studies*, 25(9), 2659–2673.

Bettmann, J. E., Lundahl, B. W., Wright, R., Jasperson, R. A., & McRoberts, C. H. (2011). Who are they? A descriptive study of adolescents in wilderness and residential programs. *Residential Treatment for Children & Youth*, 28(3), 192–210.

*Bettmann, J. E., Russell, K. C., & Parry, K. J. (2013). How substance abuse recovery skills, readiness to change and symptom reduction impact change processes in wilderness therapy participants. *Journal of Child and Family Studies*, 22, 1039–1050.

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. Public Safety Canada. https://www.publicsafety.gc.ca/cnt/rsrscs/plctns/rsk-nd-rspsvty/rsk-nd-rspsvty-eng.pdf

Borenstein, M., & Hedges, L. V. (2019). Statistically describing and combining study outcomes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 207–243). Russell Sage Foundation.

Bowen, D. J., & Neill, J. T. (2013). A meta-analysis of adventure therapy outcomes and moderators. *The Open Psychology Journal*, 6(1), 28–53.
Lee, C., & Wong, J. S. (2021). Examining the effects of teen dating violence prevention programs: A systematic review and meta-analysis. *Journal of Experimental Criminology*. https://doi.org/10.1007/s11292-020-09442-x

*Lewis, S. F. (2013). Examining changes in substance use and conduct problems among treatment-seeking adolescents. *Child and Adolescent Mental Health, 18*(1), 33–38.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE.

Lowe, T. A. (2005). *The effectiveness of Anasazi: A wilderness treatment program* (Publication No. 3192842) [Doctoral dissertation, Brigham Young University]. ProQuest Dissertations Publishing.

Madaleno, M., & Wights, S. (2015). *Guide to scoring methods using the Maryland Scientific Methods Scale*. What Works Centre for Local Economic Growth.

Mier, C., & Ladny, R. T. (2018). Does self-esteem negatively impact crime and delinquency? A meta-analytic review of 25 years of evidence. *Deviant Behavior, 39*(8), 1006–1022.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *British Medical Journal, 7*(1), 105–125.

Mullins, P., Lowan-Trudeau, G., & Fox, K. (2016). Healing the split head of outdoor recreation and outdoor education. In B. Humberstone, H. Prince, & K. A. Henderson (Eds.), *Routledge international handbook of outdoor studies* (pp. 49–58). Routledge.

Nikkel, L. J. (2014). *Adventure therapy for youth with addictions in residential treatment: An analysis of program processes and proximate outcomes* [Doctoral dissertation, University of Manitoba] (Publication No. MS26251). ProQuest Dissertations Publishing.

Outdoor Behavioral Healthcare Council. (2014). *About us*. https://obhcouncil.com/about/

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal, 372*(71), 1–9.

Paquette, J., & Vitaro, F. (2014). Wilderness therapy, interpersonal skills and accomplishment motivation: Impact analysis on antisocial behavior and socio-professional status. *Residential Treatment for Children & Youth, 31*(3), 230–252.

Raymond, I. J. (2016). *Can intensive wilderness programs be a catalyst for positive change for young people at risk of future offending, educational disengagement or poor wellbeing?* [Unpublished doctoral dissertation]. Flinders University.

Russell, K. C. (2006). *Brat camp, boot camp, or . . . ? Exploring wilderness therapy program theory*. *Journal of Experiential Education*, 28(4), 70–79.

Russell, K. C. (2006). *Brat camp, boot camp, or . . . ? Exploring wilderness therapy program theory*. *Journal of Adventure Education and Outdoor Learning*, 6(1), 51–67.

Russell, K. C., Gillis, H. L., & Harvey, J. D. (2018). An evaluation of Alaska Crossings: Comparison of the Client Status Review and the Youth Outcome Questionnaire. *Journal of Therapeutic Schools and Programs, 10*(1), 127–154.

Russell, K. C., & Phillips-Miller, D. (2002). Perspectives on the wilderness therapy process and its relation to outcome. *Child & Youth Care Forum, 44*(5), 671–686.

Skidmore, K. (2017, June 25). Stealing wisdom: Cultural appropriation and misrepresentation within adventure therapy and outdoor education. *OutdoorEd.com*. https://www.outdoored.com/articles/stealing-wisdom-cultural-appropriation-and-misrepresentation-within-adventure-therapy-and

Stevens, B., Kagan, S., Yamada, J., Epstein, I., Beamer, M., Bilodeau, M., & Baruchel, S. (2004). Adventure therapy for adolescents with cancer. *Pediatric Blood & Cancer, 43*(3), 278–284.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. D. (1998). Preventing crime: What works, what doesn’t, what’s promising. *National Institute of Justice Research in Brief. U.S. Department of Justice*. https://www.ncjrs.gov/pdffiles/171676.PDF

*Tucker, A., Norton, C. L., DeMille, S., & Hobson, J. (2016). The impact of wilderness therapy: Using an integrated care approach. *Journal of Experiential Education, 39*(1), 15–30.

Tucker, A., Widmer, M., Faddis, T. J., Randolph, B., & Gass, M. (2017). Engaging families in outdoor behavioral healthcare. In J. D. Christenson & A. N. Merritts (Eds.), *Family therapy with adolescents in residential treatment* (pp. 263–283). Springer.

***Tucker, A. R., Bettmann, J. E., Norton, C. L., & Comart, C. (2015). The role of transport use in adolescent wilderness treatment: Its relationship to readiness to change and outcomes. *Child & Youth Care Forum, 44*(5), 671–686.

***Turner, J. S. (2009). *Social support interactions in therapeutic adventure education programs* [Unpublished doctoral dissertation]. University of Georgia.

Vevea, J. L., Coburn, K., & Sutton, A. (2019). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 153–172). Russell Sage Foundation.
Vissell, R. (2004). *Effects of wilderness therapy on youth at risk’s concept of self and other: A deeper understanding of the journey* (Publication No. 3146377) [Doctoral dissertation, Institute of Transpersonal Psychology]. ProQuest Dissertations Publishing.

Walsh, M. A. (2009). *Wilderness adventure programming as an intervention for youthful offenders: Self-efficacy, resilience, and hope for the future* [Doctoral dissertation, University of Minnesota] (Publication No. AAI3373431). ProQuest Dissertations Publishing.

Walsh, V., & Golins, G. (1976). *The exploration of the Outward Bound process*. Colorado Outward Bound.

White, W. (2012). *A history of adventure therapy*. In M. A. Gass, H. L. Gillis, & K. C. Russell (Eds.), *Adventure therapy: Theory, practice, & research* (pp. 19–46). Routledge.

**Wichmann, T. F. (1990). *Interpersonal problem solving and asocial behavior in a therapeutic wilderness program* [Doctoral dissertation, Southern Illinois University] (Publication No. 9129891). ProQuest Dissertations Publishing.

Wilson, S. J., & Lipsey, M. W. (2000). *Wilderness challenge programs for delinquent youth: A meta-analysis of outcome evaluations*. *Evaluation and Program Planning, 23*(1), 1–12.

**Natalie Beck**, BA (Hons), is an MA student in the School of Criminology at Simon Fraser University. Her research interests include prevention programming, delinquency, and rehabilitation. She has been a research assistant to Dr. Wong on projects focusing on topics such as recidivism prevention, sexual violence prevention, and gang prevention programming. She uses primarily quantitative methods of analyses such as meta-analysis and systematic review.

**Jennifer S. Wong** is an associate professor in the School of Criminology at Simon Fraser University. Her work employs applied, quantitative, and qualitative methods to study issues of delinquency/crime prevention and intervention, focusing primarily on program evaluation and meta-analysis in the areas of crime prevention and intervention. Recent work includes evaluations of intimate partner violence intervention programs, examining the impact of prior deportation of illegal immigrants on recidivism, and meta-analyses on gang prevention programs, restorative justice programs for at-risk youth, halfway houses, and intensive supervision programs.