

Running Head: PSYCHOLOGICAL DEBRIEFING

Examining the Effectiveness of Psychological Debriefing Following a Critical Incident:

A Meta-Analysis

Alyssa Taylor

A thesis submitted to the faculty of Graduate Studies and Research in partial
fulfillment of the requirements of the Masters of Arts degree

Department of Psychology

Carleton University

Ottawa, ON

August 2007

Copyright ©, 2007 Alyssa Taylor



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-33774-5
Our file *Notre référence*
ISBN: 978-0-494-33774-5

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

Psychological debriefing (PD) is a widespread and commonly relied upon crisis intervention technique used to help individuals post-critical incident. However, the acceptance of PD has outpaced scientific research, as attempts to summarize PD literature have yet to determine its effectiveness. The current thesis represents the most comprehensive empirical summary of the literature. A total of 24 studies on PD were examined using a meta-analytic approach. Results indicate that there was a small effect supporting the use of PD, especially for reducing anger and improving general health and functioning. Furthermore, the moderator analysis showed that PD is most effective when using Mitchell's CISM model, when targeting occupational personnel in group settings who were exposed to occupationally-related traumatic events, and when sessions are mandatory. Given the limitations of this thesis, the most appropriate suggestion is to use PD with caution in the settings that have been identified as promoting PD effectiveness.

Acknowledgements

It is rare to have one's expectations constantly met, let alone to always have those expectations surpassed. Craig Bennell is one individual who possesses this ability. As my thesis supervisor, he not only fought to salvage a previous thesis project that was halted due to unforeseen circumstances, but he motivated and helped me to complete this current thesis in order to ensure my timely graduation. This is above and beyond anything expected by a faculty member and it is why I consider him a true mentor. Thank you for all that you do. Your time and hard work is appreciated more than you know.

Thank you to my committee for providing thoughtful feedback and challenging questions. I am grateful to the supportive and informative administration staff within the department. I would also like to thank my graduate student colleagues, especially my friend Shevaun. Your support and guidance has been the source of much of my success. You make me strive to be better.

Most importantly, I would like to acknowledge my parents, who have been the driving force behind my education. Thank you for listening to me through the worst of it and praising me through the best of it. Dad, thank you for reading numerous drafts and always providing me with thoughtful advice. Mom, you are my academic inspiration. Your passion for knowledge, countless research ideas and moral support made this degree possible. I love you both.

Finally, I thank my sister, Davina (you may wonder why you are being thanked), but if it was not for your down-to-earth attitude and ability to make me laugh about 'academics', I may have quit long ago. I will need your humour to get me through the next degree.

Table of Contents

Psychological Debriefing.....1

Review of Psychological Debriefing.....3

Potential Moderators of Psychological Debriefing Effectiveness.....7

 The Model of Psychological Debriefing.....8

 The Type of Incident and Context of the Debriefing.....11

 The Facilitator of the Debriefing Session.....13

 The Individuals Being Debriefed.....14

 The Optimal Time to Debrief.....16

 Debriefing as a Requirement.....17

The Current Study.....17

Methods.....19

 Design.....19

 Advantages and Disadvantages of Meta-Analyses.....20

 Sample of Studies.....23

 Coding Manual.....24

 Data Analysis.....28

Results.....32

 Analysis of Main Effects.....35

 Moderator Analysis.....36

Discussion.....54

 Is Debriefing Effective?.....54

 Which Moderators Impact the Effectiveness of Debriefing?.....55

Limitations of this Research.....62

Implications and Future Directions.....64

Conclusion.....66

References.....68

Appendices.....83

 Appendix A.....83

 Appendix B.....84

 Appendix C.....92

 Appendix D.....94

 Appendix E.....96

List of Tables

Table 1.	Phases of PD Models.....	9
Table 2.	A Partial List of the Variables Coded and the Outcome Measures.....	25
Table 3.	Studies Included in the Current Meta-analysis.....	33
Table 4.	Unweighted Mean d , Hedges' \hat{g} , Significance Levels, and U_3 of Debriefing Versus No Debriefing for Aggregated Outcome Measures....	35
Table 5.	Analysis of Heterogeneity for the Outcome Measures.....	37
Table 6.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Model of Debriefing Moderator for all Outcome Measures.....	40
Table 7.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Context of the Debriefing Moderator for all Outcome Measures.....	42
Table 8.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Type of Incident Moderator for all Outcome Measures.....	44
Table 9.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Identity of the Debriefing Facilitator Moderator for all Outcome Measures.....	46
Table 10.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Role of the Individual Moderator for all Outcome Measures.....	48
Table 11.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Timing of the Debriefing Moderator for all Outcome Measures.....	50
Table 12.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Assessment Time Point Post-debriefing Moderator for all Outcome Measures.....	51
Table 13.	Hedges' \hat{g} , Q Statistics, and Significance Levels for the Debriefing Option Moderator for all Outcome Measures.....	53

List of Appendices

Appendix A. List of Researchers Contacted for Studies for Inclusion.....83

Appendix B. Coding Manual.....84

Appendix C. Comprehensive List of Outcome Measures and Corresponding
Aggregated Categories.....92

Appendix D. Unweighted Mean d , Hedges' \hat{g} , Significance Levels, and U_3
for Unaggregated Outcome Measures.....94

Appendix E. Hedges' \hat{g} , Q statistics, and Significance Levels for Standard
Moderators for all Outcome Measures.....96

Examining the Effectiveness of Psychological Debriefing Following a Critical Incident:

A Meta-Analysis

This thesis seeks to answer two questions: (1) to what extent is psychological debriefing (PD) an effective approach for combating the negative effects associated with critical incident exposure and (2) what factors influence the success of this approach. Over recent years, PD has become a widespread, commonly accepted intervention for both victims and emergency workers following exposure to trauma (Devilley, Gist, & Cotton, 2006). However, despite its acceptance and frequency of use, research concerning the efficacy of PD remains unclear (Gibson & Campling, 2006; Mitchell & Everly, 2000). Thus, practitioners are currently unable to determine whether they should rely on PD when intervening with potentially traumatized individuals and, if they do rely on PD, they are unable to determine exactly how it should be implemented in order to maximize success.

Psychological Debriefing

Psychological debriefing is a term that was first used by Raphael (1984) and it is most suitably equated with “emotional first aid” (Devilley & Cotton, 2003, p. 144). While there are countless definitions for the general term, PD, for the purpose of this thesis it will be operationalized as “any brief, short-term intervention aimed at mitigating long-term distress or preventing the emergence of posttraumatic stress” (Devilley et al., 2006, p. 318). Typically, PD is one part of a larger crisis intervention program aimed at helping individuals who have been involved in a critical incident. A critical incident is defined as “any event which has the potential to overwhelm one’s usual coping mechanisms resulting in psychological distress and an impairment of normal adaptive functioning”

(Everly & Mitchell, 1999, p. 11). Individuals exposed to such events typically feel some degree of traumatization. This can cause feelings of stress, which is a state of cognitive, physical, emotional, and behavioural arousal that, if it goes unchecked, can develop into an array of physiological and/or psychological disorders (Gibson & Campling, 2006; Mitchell, 2004).

While there is no agreed upon, all-encompassing theory that explains why PD might be effective, there is a logical rationale for its use. Individuals often possess coping mechanisms that allow them to deal with many forms of stress (Lazarus, 1977). However, since most individuals have no experience with the trauma associated with critical incidents their coping mechanisms for these types of events will be quite limited or even non-existent. This means that individuals will either have insufficient coping strategies to manage critical incident-related stress or will rely on maladaptive coping mechanisms (e.g., denial) (Latack, 1986). This explains why individuals exposed to such traumas often experience high levels of stress and why, subsequently, this stress can develop into a range of stress disorders (e.g., post-traumatic stress disorder; PTSD).¹

Psychological debriefing is an intervention dedicated to improving individuals' reactions to a critical incident, which is accomplished in part by providing the individuals with the necessary coping "tools" that they lack (Schopler & Galinsky, 1993). More specifically, there are three general goals of PD: (1) mitigation of the impact of the event, (2) facilitation of normal recovery processes, and (3) restoration of adaptive functioning (Mitchell, 2004). A typical PD session attempts to mitigate the severity of an event through a process of normalizing the reactions that individuals are experiencing through

¹ For the purpose of this thesis, the PTSD literature will not be extensively discussed, so as to primarily focus on PD research.

group discussion, emotional ventilation, and the sharing of feelings by those that encountered the trauma. A debriefing educates individuals about normal stress responses that they may encounter (e.g., sleep or memory disturbances) following exposure to a critical incident and about various types of adaptive coping mechanisms (Mitchell & Everly, 1997). Finally, PD also establishes a positive contact with mental health professionals (MHPs) and informs the participants about the types of support that may be required in the future in order to resolve any long term stress reactions (e.g., peer-support, spouse or family counselling). These PD strategies are thought to accelerate the rate of normal recovery for individuals who are having normal reactions to traumatic events (Mitchell & Everly, 1997).

Reviews of Psychological Debriefing

Despite the widespread use of PD, its efficacy has been the subject of extensive debate in recent years (Arendt & Elklit, 2001). While there are numerous studies examining the effectiveness of PD, there remains a large amount of controversy in the field and a lot of confusion as to whether PD actually works, and under what conditions it works best. This confusion primarily stems from the mixed conclusions reached by researchers. For instance, several reports conclude that there is no empirical evidence supporting the usefulness of PD (Bisson & Deahl, 1994; Deahl, Gillham, Thomas, Searle, & Srinivasan, 1994; Raphael, Meldrum, & McFarlane, 1995). In fact, several studies even report that PD can produce negative effects (e.g., an increase in PTSD symptoms for individuals participating in PD) (Bisson, Jenkins, Alexander, & Bannister, 1997; Kenardy, Webster, Lewin, Carr, Hazell, & Carter, 1996; McFarlane, 1988; Rose, Bisson, & Wessely, 2000). On the other hand, several reviews show the opposite, indicating that

PD is an effective intervention for helping individuals after trauma exposure (Bisson, McFarlane, & Rose, 2000; Everly & Boyle, 1999; Everly & Mitchell, 1999; Everly, Flannery, & Mitchell, 2000).

Several studies have attempted to clear up this confusion by conducting non-statistical reviews of the PD literature (e.g., Arendt & Elklit, 2001; Everly et al., 2000; Lewis, 2003; Malcolm, Seaton, Perera, Sheehan, & Van Hasselt, 2005; Regehr, 2001). Typically when using this approach, researchers group studies into those with either positive or negative results and hypothesize about which factors led to particular outcomes. While this may be useful to some degree, this non-statistical approach is problematic. Most notably, the lack of objective statistics for summarizing the body of research can lead to biased, or completely erroneous, conclusions (Cooper & Rosenthal, 1980). Indeed, many researchers have commented on the fact that, while these types of qualitative reviews can prove valuable to the field, they often leave the reader even more confused about the conflicting results emerging from studies in the area (Rosenthal & DiMatteo, 2001). In an attempt to remedy this situation, researchers have begun to conduct quantitative reviews of the PD literature, using meta-analytic techniques. In fact, several meta-analyses have already been completed in regards to PD effectiveness.

First, Everly and Boyle (1999) conducted a meta-analysis on the most popular model of debriefing, Mitchell's Critical Incident Stress Debriefing (CISD; Mitchell, 1983). As discussed in more detail below, this model is a group crisis intervention tool designed to assist people that have been exposed to the same significant traumatic event. It forms one part of a much more comprehensive, systematic, multi-component approach to critical incident stress management (CISM; Mitchell, 2004). Only studies specifically

assessing this model of debriefing were included in the meta-analysis, which resulted in an examination of five studies. These studies resulted in a subject pool of 341 adults that were all exposed to critical incidents as part of their job (e.g., police officers) and the type of critical incidents to which the participants were exposed was diverse (e.g., mass shooting, hurricanes, etc.). The results from this study supported the use of PD given the large mean effect size ($d = .86$) that was found. More specifically, the findings supported the utility of CISD for mitigating symptoms of psychological distress including PTSD symptoms, anxiety, depression, and anger. Everly and Boyle (1999) recognized that a larger sample would have been desirable in order to obtain more reliable results and they also highlighted the fact that their results could not be generalized to other models or variations of PD.

A second meta-analysis was conducted by the same authors where they examined the same original five studies plus an additional five studies, in order to examine a wider array of PD models (Everly, Boyle, & Lating, 1999). However, this study still restricted the inclusion criteria to individuals who were exposed to critical incidents as part of their occupation. In this case, a somewhat smaller average effect size was found, but it was still within the moderate range ($d = .54$) attesting to the usefulness of PD for these types of individuals. In their conclusion, the authors stated that, while CISD is the most popular intervention, it does not have to be the only intervention, as other forms of PD have been shown to be effective and are empirically justified.

The most recent meta-analysis to be conducted was by van Emmerik, Kamphuis, Hulsbosch, and Emmelkamp (2002). They examined single-session debriefing after exposure to psychological trauma. The inclusion criteria stated that the debriefing had to

be done within one month of the trauma, psychological symptoms had to be assessed with widely accepted clinical outcome measures, and data from these assessments had to be collected before and after the intervention. The studies included a range of critical incidents (e.g., burns, road traffic accidents, shootings, etc.) and a range of participants (e.g., emergency workers, victims, soldiers, etc.). Seven studies, none of which were included in previous meta-analyses, were analyzed. The authors grouped the interventions into one of three categories: CISM-type interventions, non-CISM interventions (e.g., counselling), and no intervention (i.e., control group). Results showed a very small effect size ($d = .13$) indicating that CISM-type interventions did not lead to a reduction in PTSD symptoms. In fact, CISM was found to be no more effective than non-CISM intervention or no intervention. Based on these results, the authors cautioned that claims that single session PD can prevent trauma-related symptoms are not justified.

While the completion of these analyses should have cleared up the confusion surrounding the efficacy of PD, they have merely created more confusion. It is currently unclear why discrepancies remain between studies that have examined the effectiveness of PD (Gibson & Campling, 2006). However, a number of possible explanations exist. The discrepancies may be explained in part by the fact that each of the above meta-analytic studies only examined a very small number of studies, thus decreasing the reliability of the results. In addition, and related, it is likely that some of the discrepancies can be attributed to the fact that researchers have only ever examined PD under very specific conditions (e.g., often by focusing on one particular model of PD using one specific population) and that the effectiveness of PD varies across the conditions under which it has been scrutinized. Furthermore, different researchers have focused on

different outcome measures when evaluating PD (e.g., including measures related to anger, anxiety, general well-being, etc.) and PD may be more effective at mitigating the negative impact of certain outcomes over others.

In general, based on problems with current research, it is impossible to determine whether PD actually works or to identify the conditions under which PD can be used with maximum success. In other words, the lack of clarity resulting from current meta-analyses makes it difficult to determine whether the findings emerging from these studies should be treated as evidence that PD works (or does not work), or whether the findings should simply be interpreted as indicating whether PD worked (or did not) under the very specific circumstances focused on in the studies. The former is the more common approach taken in the research literature (e.g., Everly & Boyle, 1999; Everly et al., 1999; van Emmerik et al., 2002), though the latter approach is arguably more appropriate. It is this latter approach that will be adopted in the current study. It will be assumed that there may be conditions under which PD works, and conditions under which it does not, and this can be determined by examining PD effectiveness across a range of potential moderator variables.

Potential Moderators of Psychological Debriefing Effectiveness

The PD literature has identified several important factors that may influence the effectiveness of crisis interventions. The following section will address some of the factors that may moderate the effectiveness of PD. While this is not a comprehensive list, these factors are the ones that are typically focused on when attempts are made to explain the discrepancies within the research literature and they will be the moderators that are focused on in the current study.

The Model of Psychological Debriefing

As previously stated, there are several models of PD and currently no official universal standard of how PD should be delivered. For example, PD is frequently conceptualized as a debriefing consisting of a short group meeting, led by some type of trained professional, in order to facilitate individuals to talk about their critical incident experience (Deville et al., 2006). However, alternative versions of PD exist. For instance, individualized debriefing sessions can be used instead of group sessions and the debriefing can be delivered by peers as opposed to MHPs. It is not currently clear whether some or all of the various PD models actually work and the results of studies in this area may differ drastically depending on the specific model being examined. As there is no exhaustive list of models or adaptations, two of the most popular models will be addressed here (see Table 1 for a condensed version of the models).

Table 1

Phases of PD Models

Mitchell's CISD	Dyregrov's PD model
Introduction/Rules: outline of process is explained, participants are urged to talk	Introduction/Rules: outline of process is explained, participants are urged to talk
Facts: discussion of what happened during the incident	Expectations/Facts: discussion of what happened prior to the incident and during
Thoughts: cognitive emphasis when talking of thoughts about the trauma	Thoughts/Decisions: emphasis on discussing decision-making process during the incident
Reactions: emotional emphasis when talking of feelings and fears about the trauma	Sensory Impressions: sensory information about the scene is discussed
Symptoms: any stress symptoms during or after the incident are discussed	Emotional Reactions: emotional emphasis when talking about feelings and fears about the trauma
Teaching: information is given about 'normal' stress reactions	Normalization: emphasis on normalizing behaviours and feelings
Re-entry: summarizing all that has occurred and address future issues	Future Planning/Coping: future plans for coping strategies to manage stress
	Disengagement: answering of questions and a summary of the session is provided

The first formal PD model was proposed by Mitchell (1983). It has since developed into the most popular and widely used model, known as CISD. This model, originally consisting of six-phases, has evolved to contain seven distinct phases: Introduction/Rules, Facts, Thoughts, Reactions, Symptoms, Teaching, and Re-entry. Its goal is to use the ventilation of feelings, peer-group support, and health education to help individuals better manage critical incident stress. This type of debriefing comes after

more immediate types of intervention (e.g., defusing), but before more reactive strategies (e.g., PTSD counselling), and is always meant to be part of a larger CISM program. Since its development, CISD has evolved into other PD models, including stand-alone debriefing intervention. This has created much controversy from supporters of CISD because it is believed that these interventions are less effective than CISD and potentially detrimental to its reputation (Mitchell, 2004).

A second model of PD, developed by Dyregrov (1989), is based on Mitchell's (1983) CISD and is often thought of as being interchangeable. However, there are several noteworthy differences between the two models. For example, compared to CISD, Dyregrov's model begins with a discussion about what occurred just prior to the critical incident, whereas the discussion in Mitchell's model targets events that begin with the onset of the traumatic incident. Dyregrov's model also has a stronger focus on the individual's decision-making process during the trauma and pays closer attention to the types of sensory information experienced by the individual during the incident (Rose & Tehrani, 2002). This model also has a greater focus on the reactions and responses of the participants than Mitchell's original CISD model and it contains eight phases: Introduction/Rules, Expectation/Facts, Thoughts/Decisions, Sensory Impressions, Emotional Reactions, Normalization, Future Planning/Coping, and Disengagement.

In addition to these models of PD that focus on group debriefing strategies there are a number of other models that provide PD on an individual basis (i.e., SAFER model, Tehrani model) (Gibson & Campling, 2006; Tehrani & Westlake, 1994). Furthermore, several other models of PD have been generated for specific populations (e.g., police

officers) and/or specific events (e.g., line-of-duty deaths) (Bohl, 1995; Mitchell & Levenson, 2006).

Similar to the majority of PD research, findings in regards to these various models of PD remain mixed. The vast majority of the PD literature has examined CISD. Mitchell (2004) summarized many of the empirical studies examining CISD and found that 10 controlled studies and over 20 case studies and literature reviews supported its use. In contrast, several other researchers have found that CISD is less than effective (Bisson et al., 1997; Carlier, Voerman, & Gersons, 2000; Lee, Slade, & Lygo, 1996). Similar things can be said of Dyregrov's model, with research studies either supporting its use or refuting its effectiveness (Deahl et al., 1994; Dyregrov & Gjestad, 2003), and of the many other PD models that have been proposed (e.g., Armstrong, O'Callahan, & Marmar, 1991; Sijbrandij, Olf, Reitsma, Carlier, & Gersons, 2006; Slaikou, 1984).

Thus, after decades of research examining a wide range of PD models it remains unclear as to which ones are most effective and the precise conditions under which the various models work best is certainly unknown (although many of the moderators that will be discussed in the remainder of this section might give some insight into these issues).

The Type of Incident and Context of the Debriefing

Currently, it is also unclear whether the effectiveness of PD varies depending on the type of incident that is targeted in debriefing. For example, some researchers have made a distinction between critical incidents experienced within a group (e.g., emergency personnel exposed to a natural disaster) and non-group critical incidents (e.g., an individual experiencing a hostage taking) (Mitchell, 2004). These researchers maintain

that PD is likely to be most effective in cases of group trauma. This may be in part because the traumatized individuals can share their experiences and verbalize their own distress about the incident. Some suggest that this helps individuals to know that they are not experiencing the situation alone and that their feelings of stress are shared by others, which can ultimately help to validate their own reactions (Mitchell & Everly, 1997).

Despite a wealth of research examining the effectiveness of PD with group versus non-group trauma, we are still no further in our understanding of this issue. Research has shown that PD has sometimes been effective for individuals involved in group crises (e.g., Bohl, 1991; Wee, Mills, & Koehler, 1999), however other research reveals the opposite (e.g., Deahl et al., 1994; Leonard & Alison, 1999; Marchand, Guay, Boyer, Iucci, Martin, & St-Hilare, 2006). Likewise, research has shown that PD can be effective when used on those involved in non-group trauma (Humphries & Carr, 2001) and at other times it has been shown to be ineffective (e.g., Bisson et al., 1997; Rose, Brewin, Andrews, & Kirk, 1999).

As indicated above in relation to the meta-analytic studies, it is likely that some of the discrepancies in research that has examined the impact of PD for varying types of incidents can be explained by the narrow focus taken by the researchers in these studies (i.e., the very restricted conditions under which hypotheses are tested). It is possible that interaction effects need to be examined in order to fully understand whether PD can be effective across a range of incidents (e.g., the type of incident may interact with the model of PD or any of the other moderators to be discussed below). Given sufficient data, the current study will allow this issue to be examined.

The Facilitator of the Debriefing Session

A range of people have filled the role of facilitator of a PD session (e.g., trained peers, social workers, psychologists, nurses, welfare professionals, researchers, etc.) and this has led to much debate over who is most suitable for this position. This debate is extremely important given that, currently, peers are typically relied on for debriefings in many occupational contexts (e.g., Miller, 2006), whereas MHPs are relied on in other settings (e.g., Sijbrandij et al., 2006). Advocates for peer debriefing suggest that MHPs add to the already abnormal and frightening situation surrounding trauma, and they point out that peers understand the language, culture, and behaviours of emergency workers and so they are viewed as better able to understand what the sufferers are going through (Breznitz, 1980; Hoff, 1989). Supporters of MHPs conducting debriefings argue that they have the proper training, expertise, and experience to deal with all of the issues that typically arise in PD (Mitchell & Everly, 1997).

Thus far, there is little to no empirical support that MHPs get better results from PD than trained and supervised paraprofessional helpers (Berman & Norton, 1985; Hattie, Sharpley, & Rogers, 1984; Stein & Lambert, 1984). In fact, there is evidence supporting and refuting the use of both facilitator groups (e.g., Bohl, 1991; Carlier et al., 2000; Deahl et al., 1994; Stallard, Velleman, Salter, Howse, Yule, & Taylor, 2006; Warren, 1995). One of the only studies to directly compare the effectiveness of MHPs versus peer supporters suggests that peer supporters are preferable, but this was only tested under a narrow range of conditions. More specifically, Simms-Ellis and Madill (2001) conducted a qualitative study comparing this moderator with armed robbery victims. Results indicated that individuals preferred to be debriefed by peers because they

were seen as able to normalize the experience and were considered to be “one of us”. The MHPs were perceived as less caring because they were perceived as “just doing a job”.

Mitchell (2004) has argued that studies failing to find positive effects of clinician-led debriefing reach such conclusions because the clinicians in these studies had no formal debriefing training from the International Critical Incident Stress Foundation, the association that is primarily responsible for conducting such training (notably, the president of this organization is Mitchell, the developer of CISD). McNally (2003) states that Mitchell’s argument is premature, since everyone receiving training from this foundation are being taught how to implement an intervention that has yet to be empirically validated. Regardless of these arguments, the current study will represent the largest study to date that will allow the impact of this moderator variable to be tested.

The Individuals Being Debriefed

Given that PD was originally created for dealing with the stress experienced by emergency workers, it may not be surprising that a number of studies have shown that PD is effective when this population is targeted (Bohl, 1991, 1995; Wee et al., 1999). Having said this, there are times when PD has not worked with these individuals (Carlier et al., 2000; Leonard & Alison, 1999). More recently, PD has been extended to primary trauma victims (e.g., victims of burns, miscarriages, road traffic accidents, etc.). While studies do exist showing that PD can be used to effectively mitigate stress reactions within this population as well (e.g., Chemtob, Tomas, Law, & Cremniter, 1997; Nurmi, 1999), there are more studies suggesting the opposite (e.g., Jacobs, Horne-Moyer, & Jones, 2004; McFarlane, 1988; Mitchell, 2004; Sijbrandij et al., 2006). Thus, questions remain about the suitability of PD for various target groups.

Why should these groups respond differently to PD? Those that argue that there will be differences typically highlight the fact that these groups differ from each other in meaningful ways and these differences directly or indirectly influence the effectiveness of PD (resulting in a situation where emergency workers can benefit from PD, but not other populations). For example, emergency workers may be more likely than other populations to experience group trauma (e.g., disaster relief) and the negative effects resulting from exposure to these sorts of incidents may be more conducive to PD-type interventions (Mitchell & Everly, 1997). In addition, emergency service workers may differ from the general population in terms of their personality characteristics and this may lead to more favourable responses to PD (e.g., compared to other groups, emergency workers are likely to be more action-oriented, have a high level of dedication, possess risk-taking tendencies, etc.) (Mitchell & Everly, 1997). Furthermore, an inherent part of the emergency personnel's job is exposure to critical incidents and this may lead to a set of conditions that allows PD to work for this group (Boudreaux & McCabe, 2000; Moran, 1998). For instance, there is likely to be an expectation that such incidents will be encountered on the part of the worker, which may result in less serious stress reactions compared to unsuspecting victims, and there is an attempt on the part of the employer to prepare their emergency workers for critical incident exposure (e.g., through appropriate training where workers are taught adaptive coping skills).

Despite these potential differences between emergency workers and other groups, which are used by some to explain the differential response rates to PD, others still believe that PD can be effective for both groups. For example, Walker (1990) conducted a review of the debriefing literature and found that people who experienced a critical

incident responded in a consistent and systematic way, regardless of whether they were emergency workers or victims. Thus, the debate over who is a suitable target for PD continues.

The Optimal Time to Debrief

A common recommendation in the PD literature is that PD should take place immediately following the critical incident (except in cases where physical injuries need to be stabilized) and take place in conjunction with medical service (e.g., Campfield & Hills, 2001; Raphael, 1986; Talbot, 1990). In contrast, other researchers recommend that PD should be slightly delayed and take place 24-48 hours post-trauma (Kenardy et al., 1996), 24-72 hours post-trauma (Everly, 1995; Mitchell, 1983), or 48-72 hours after the incident (Busuttil & Busuttil, 1995). The evidence supporting one or the other of these scenarios is largely anecdotal and, therefore, there is an urgent need for empirical research of the sort that will be conducted here. Examining how the timing of PD interacts with some of the other moderators discussed above is particularly important given that some have argued that such interactions exist (e.g., a delay of 24-72 hours is thought by some to be appropriate when intervening with emergency workers, but average citizens may require earlier interventions; Mitchell, 1983). Given sufficient data, such an analysis will be conducted.

A related analysis will also be conducted in order to determine if the timing of the post-debriefing assessment (i.e., to measure the impact of the debriefing) has any effect on debriefing effectiveness. Follow-up periods ranging from days (e.g., Carlier et al., 2000; Matthews, 1998) to several months (e.g., Bisson et al., 1997; Hobbs, Mayou, Harrison, & Worlock, 1996) have been used in the research literature. Examining this

issue will indicate how quickly potential PD effects emerge (and fade) after the intervention.

Debriefing as a Requirement

Another common question regarding PD is whether or not it should be mandatory or voluntary. Some organizations make attendance voluntary (e.g., American Red Cross responders), whereas others make it mandatory (e.g., many police and fire departments) (Pender, 2006). Debriefings for emergency workers in particular are typically automatic, whereby anyone who is involved in a critical incident must participate (Hokanson & Wirth, 2000). There are several reasons for this. One reason is that emergency service personnel often do not feel as though they need help after a critical incident, even when they might actually benefit from it (Leonard & Alison, 1999). Another reason is that personnel who attend voluntary PD sessions can be viewed as “weak” by their coworkers, thus decreasing the chance that these individuals will seek out the help they require (Hokanson & Wirth, 2000). The major problem in determining whether PD should be mandatory or voluntary is that no empirical evidence exists on the topic (Hokanson & Wirth, 2000). However, this debate, like many of the previous ones regarding potential moderators, may be premature given that the effectiveness of debriefing has yet to be empirically determined.

The Current Study

The goal of the current meta-analysis is to determine whether PD is effective for individuals following a critical incident. The general effectiveness of PD will be determined and an exploration of moderators (including interactions between moderators) will also be carried out whenever possible in order to better understand the conditions

under which PD is most effective. The following is a list of the major research questions that were examined in the current study. Mixed results from existing research made it difficult to develop directional hypotheses.

1. In general, is PD an effective intervention for individuals following a critical incident?
2. Does PD effectiveness vary as a function of the outcome measure used to evaluate its effectiveness?
3. Is one model of PD most effective?
4. Is PD more effective with certain types of critical incidents (e.g., group versus individual trauma)?
5. Does the identity of the debriefing facilitator (e.g., clinician versus peer) have an impact on the effectiveness of PD?
6. Do certain types of individuals (e.g., emergency personnel versus primary victims) benefit more from PD?
7. Is there an optimal time (after exposure to trauma) to perform a PD?
8. Does the effectiveness of PD depend on the length of the time interval between the PD and the assessment of the individual?
9. Does being given the option to participate in PD (e.g., mandatory versus voluntary) influence its effectiveness?

While the main moderators discussed in the literature have been addressed within these research questions, there are countless other variables (e.g., severity of the trauma, previous trauma exposure, etc.) that may influence the effectiveness of PD.

Unfortunately, details regarding these variables are less commonly included in the literature, which prevented their examination.

Method

Design

Meta-analysis was used to determine the effectiveness of PD for individuals following a critical incident and to identify potential moderators of PD effectiveness. A meta-analysis is defined as the statistical analysis of “quantitative summaries of [empirical studies] that describe the typical strength of the effect or phenomenon, its variability, its statistical significance, and the nature of the moderator variables from which one can predict the relative strength of the effect or phenomenon” (Rosenthal, 1995, p. 183). The goal of a meta-analysis is to generalize research findings and to identify the practicality of those results within a specific area of research.

A meta-analysis is much more than a statistical technique; it is a methodology for systematically examining a particular phenomenon through the following phases: formulating hypotheses, establishing inclusion criteria, statistically synthesizing and combining data from studies, searching for potential moderators and mediators to explain effects, and reporting results (Rosenthal & DiMatteo, 2001). This method of analysis has become popular in the behavioural sciences, as this area produces a large amount of literature that is often difficult to process and interpret (e.g., due to conflicting results across studies) (Rosenthal & DiMatteo, 2001). In addition, Hall and Rosenthal (1995) have noted a shift towards conducting meta-analyses because of a growing recognition that a broader view of the distribution of results within an area is of greater importance than the results of individual studies.

The specific meta-analytic approach taken in the current study follows the fixed (versus random) effects model. The fixed effects model stipulates that the sample of effect sizes under study will be relatively homogeneous (Rosenthal, 1994). However, if effect size heterogeneity is found across the sampled studies, beyond that which can be attributed to sampling error, a fixed effects model can still be used (Lipsey & Wilson, 2001). In this case, it is simply assumed that the heterogeneity can be explained by a systematic source, or sources, such as the moderator variables discussed previously. The random effects model, in contrast, assumes that variability across studies is random and unexplainable (Lipsey & Wilson, 2001).

Advantages and Disadvantages of Meta-Analyses

A meta-analysis is a form of literature review, however it differs from the traditional literature review in that it is “more systematic, more explicit, more exhaustive, and more quantitative” (Rosenthal, 1984, p. 17). A meta-analysis provides many other advantages including:

- The statistical significance of results emerging from a meta-analysis is more meaningful in comparison to individual studies that may have had a small sample size and only lacked statistical significance due to low power (Lipsey, 1990).
- A meta-analysis can ensure that data is collected from a wide-range of sources (e.g., books, theses, dissertations, technical reports, unpublished manuscripts, etc.) and these sources can be coded for and statistically tested to determine whether the source of the study affects the results. This alleviates any possibility of publication bias (i.e., a tendency to publish studies that have found significance)

and with that the possibility of inflated effect size estimates is reduced (Rosenthal, 1984).

- A meta-analysis allows for an almost unlimited capacity of detail to be included in the analysis (Lipsey, 1992). This includes a statistical examination of the possible effects that various moderator variables (e.g., model of PD) may have on the results (Glass, McGaw, & Smith, 1981). This adds to the richness of findings provided by a meta-analysis.
- More reliable and valid conclusions are produced with a meta-analysis, as opposed to qualitative studies or narrative reviews, as the results are statistically quantified. This results in the reduction of interpretative biases and subjectivity (Rosenthal, 1984).

However, meta-analyses are not without disadvantages. These include:

- A persistent argument with meta-analytic literature is that of comparing “apples and oranges” (Hunt, 1997, p. 61), whereby some view meta-analyses as an attempt to summarize results from studies that are notably different in their operationalization and measurement of variables. One argument to this critique is that the comparison of the exact same type of studies offers very little in terms of generalizability (Rosenthal & DiMatteo, 2001). So long as the meta-analyst provides clear inclusion criteria, there can be no argument about what concept is being examined. Similarly, the methodology of studies can be coded for and statistically analyzed to determine whether it has affected the results.
- Some argue that there is an inherent bias in all meta-analyses by virtue of the inclusion criteria, since no matter how thorough the attempt is to compile all

relevant research, some studies will be left out. While this is likely to be true, meta-analyses clearly provide a less distorted view of a research area than would be the case if we focused solely on the findings of one study (as is more typical). In addition, this thesis improves on previous meta-analyses that focused on a narrow range of PD studies. Indeed, an attempt will be made to include all studies that have been included in each previous, individual meta-analysis, as well as other relevant studies that have since been completed.

- Hunt (1997) makes reference to another criticism, which he refers to as the “garbage in and garbage out” issue. This issue refers to the fact that meta-analyses often mix “good” and “bad” studies and, therefore, the results can be more or less meaningless. To remedy this difficulty, Rosenthal (1991) proposed “quality weighting” with the use of a point-scale on a variety of variables (e.g., random assignment, etc.) in order to moderate for the large spectrum of study quality and de-emphasize studies of poorer quality. Something similar to this was done in the current study to determine if study quality relates to PD effectiveness.
- There is also concern that multiple results from the same study are incorporated into a meta-analysis (Glass et al., 1981). This would result in biased results, as the data would not be independent. To combat this issue, only a single outcome measure relating to the same underlying construct (e.g., PTSD) from each study should be included in the meta-analysis (if multiple outcome measures for the same construct are available, researchers should randomly choose one measure or take an average of the measures) (Glass et al., 1981; Lipsey & Wilson, 2001; Rosenthal, 1984).

Sample of Studies

The sample of studies collected for this thesis focused on any and all studies dealing with PD in relation to critical incidents. Computer searches of PsycINFO, Dissertation Abstracts International, the International Critical Incident Stress Foundation, and Google, among other databases encountered, were conducted using the following search words: debrief; debriefing; psychological debriefing; critical incident; critical incident stress; CISD; and CISM. The reference sections of all articles, especially the previous meta-analyses that have been conducted, were also examined for other potential studies. In addition, several key researchers in this area were contacted in order to obtain any other relevant unpublished theses, articles, or data for inclusion (for a list of these researchers see Appendix A).

Articles selected for inclusion in the present study possessed the following characteristics:

- Only studies that described a specific model of PD (e.g., Mitchell's CISD, debriefing adapted from Mitchell, individual PD sessions, etc.) were included.
- The debriefing must have been delivered to individuals following a critical incident according to the above-mentioned operational definition (Everly & Mitchell, 1999) (i.e., an event that has the potential to overwhelm one's usual coping mechanisms resulting in psychological distress and an impairment of normal adaptive functioning).
- Studies had to include an experimental group (who received PD) and a control group (who received no PD or a different type of intervention that did not have

any aspect of debriefing). This meant that pre-post design studies that exposed all individuals to PD were excluded.

- An outcome measure of PD effectiveness had to be included. This measure had to provide enough information to allow for the calculation of an effect size (i.e., sample size, means, standard deviations, etc.).

Coding Manual

A coding manual was designed to be used for the meta-analytic review (see Appendix B). This manual focused on several key areas: (1) study and sample characteristics (e.g., publication source, affiliation of authors, etc.), (2) factors related to PD (e.g., model of PD, type of incident, etc.), (3) outcome variables (e.g., PTSD symptoms, anxiety, etc.), and (4) reported statistics (to calculate effect sizes). See Table 2 for a partial list of the coded variables.

Table 2

A Partial List of the Variables Coded and the Outcome Measures

Study and Sample Characteristics	Moderators	Outcome Measures
1. Publication source	1. Model of debriefing a) Mitchell CISD b) Adaptation of Mitchell c) Dyregrov's model	1. PTSD symptoms a) Impact of Events Scale b) PTSD Diagnostic Scale c) Structured Interview
2. Affiliation of authors	2. Type of incident a) Accident b) Natural disaster c) Shooting	2. Anxiety a) Hospital Anxiety Scale b) SCL-90 Anxiety Scale c) State-trait Anxiety Inventory
3. Participant assignment	3. Debriefing facilitator a) Clinician b) Peer c) Researcher	3. Depression a) Beck Depression Inventory b) Birleson Depression Inventory c) Hospital Depression Scale
4. Quality of study	4. Role of individual a) Accident victims b) Emergency workers c) Victims of crime	4. Anger a) Novaco Provocation Scale b) State-trait Anger Inventory
	5. Timing of debriefing a) Within 3 days b) Within 7 days c) Within 30 days	5. General health & functioning a) General Health Questionnaire b) Maladaptive Coping Scale c) Social Readjustment Scale

Study and sample characteristics. Several study and sample characteristics were coded. The majority of these variables (e.g., publication source, affiliation of authors, etc.) were coded because they are frequently examined in meta-analytic research and provide typical sources of variability in effect size distributions (Glass et al., 1981; Rosenthal, 1984). Coding these variables is also useful for replication purposes.

Psychological debriefing factors (moderators). The studies were coded so as to compare groups that received PD to groups that received no PD. This was done to determine whether there was a benefit to debriefing individuals following critical incidents. In addition, several other factors engrained in the debriefing process (e.g., model of PD, type of incident, etc.) were coded as these are likely to act as moderators of PD effectiveness (Gibson & Campling, 2006; Mitchell & Everly, 1996; Robinson, 2000).

Outcome variables. The dependent variable in the current study is PD effectiveness. Effectiveness in the context of PD is a rather complicated concept, as existing research defines effectiveness in several different ways. For example, some argue that PD effectiveness can be measured in terms of an individuals' perception that a debriefing was helpful (Robinson, Mitchell, & Murdoch, 1995). Others focus specifically on a reduction of specific symptoms (e.g., depression, anxiety, etc.) (Leonard & Alison, 1999; Rose et al., 1999). Still others believe that PD effectiveness should be based on an examination of whether targeted individuals do not go on to develop PTSD (Marchand et al., 2006). The current study did not restrict its examination of effectiveness to any one particular definition. Instead, PD effectiveness was operationalized as the difference in scores on any reported psychological test or measure between the experimental group (receiving debriefing) and the control group (not receiving debriefing).

Due to the wide range of psychological measures used in the sampled studies, several aggregate categories of outcome measures were constructed (although analyses were done on each outcome measure separately as well to ensure that the aggregate analysis did not distort any results). Given the outcome measures that are typically focused on in the existing literature, the following categories were used: PTSD

symptoms, anxiety, depression, anger, and general health and functioning (see Appendix C for a comprehensive list of the outcome measures and corresponding aggregated categories).

It was necessary for these measures to be organized into these five aggregate categories for two reasons. First, because some studies used unique psychological tests, aggregating results that measure similar concepts allowed for a sufficient number of effect sizes for purposes of analysis. Second, the aggregation process made results more manageable for communication purposes (e.g., instead of providing results for over 25 individual outcome measures, results from the five aggregate categories are provided).

Inter-rater reliability. All of the appropriate studies (according to the inclusion criteria) were coded using the manual described in Appendix B. A research assistant, blind as to the purpose of the study, was trained in the scoring of the manual and was given three preliminary studies to code for practice. This pilot testing session is essential to ensure there is consistency between the coder and the author (Orwin, 1994).

Subsequently, the research assistant and author discussed any coding discrepancies following each individual study that was coded and identified any problems that arose because of inconsistencies, ambiguity, or other problems in the coding manual. The author then coded all of the studies, and the research assistant coded six randomly selected articles (25% of the sample) without help or feedback from the author.

Percent agreement between the two coders on 25% of the studies was determined by dividing the number of coding classifications where there was agreement between the coders by the total number of coding classifications. Variables that were used for analysis in this thesis were included in calculating inter-rater agreement, however all open-ended

questions (which were included in the coding manual for future analysis) were excluded. Of the 288 total variables coded, there were only 12 discrepancies. Thus, the overall percentage agreement was 95.8%. Any discrepancies were examined by the author and discussed with the coder in order to properly adjust data entry into the meta-analysis program.

Data Analysis

Calculating effect sizes. In order to measure the effectiveness of PD, and identify potential moderator variables, reported and estimated effect sizes were used. An effect size in this case provides a quantified measure of the degree to which the experimental group (e.g., those receiving PD) and the control group (e.g., those not receiving PD) differ on a particular outcome measure (e.g., reduction of PTSD symptoms). Thus, an effect size provides a means by which to compare all of the study findings in the sample (Rosenthal, 1984).

The effect size statistic used in this thesis was the “standardized mean difference” (Cohen, 1988). This is the most commonly used statistic in meta-analytic studies where the goal is to compare outcome measures across groups (e.g., debriefed versus not debriefed) (Lipsey & Wilson, 2001). More specifically, Cohen’s d was used, which reflects the difference between two population means (debriefed versus not debriefed groups) divided by the average population standard deviation ($d = (M_1 - M_2)/sd\ pooled$) (Rosenthal, 1994). A statistical program called *Comprehensive Meta-Analysis* was used to calculate the effect sizes from the array of statistics provided in the studies (e.g., F , t , etc.) (Biostat Inc., 2006). *Comprehensive Meta-Analysis* is one of the most frequently used commercial software packages for carrying out meta-analyses.

It should be noted that Cohen's d suffers from a slight upward bias when based on small samples of less than 20 participants (Lipsey & Wilson, 2001). Given that some of the studies included in the current meta-analysis were based on samples of this size, a correction formula for d , which is provided by Hedges (1981), was used to calculate unbiased effect sizes. These transformations were done using the following formula: corrected $d = (1 - 3/[4N - 9])(d)$, where N is equal to the sample size of the particular study that generated the effect size. Applying this correction provided a more appropriate (i.e., less inflated) set of effect sizes for analysis.

To estimate the overall effect size across all studies (in order to determine if in fact PD is an effective intervention), average effect sizes were calculated. Average effect sizes are calculated as $d_{avg} = \Sigma d/n$, where d is the corrected effect size for each study, and n is the total number of studies (Wolf, 1986). These average corrected effect sizes are unweighted, meaning that they do not take into consideration the sample size of the study. Since studies containing more participants have less sampling error than those with fewer participants, those with larger samples should be weighted more than studies with smaller sample sizes (Lipsey & Wilson, 2001). A weighted effect size is calculated by summing each effect size for a particular contrast (e.g., debriefed versus not debriefed), multiplying that sum by its inverse variance ($w = 1/se^2$), and then dividing by the sum of the weights (i.e., inverse variances). This statistic is commonly referred to as Hedges' \hat{g} . Results are therefore reported using the initial, unweighted mean difference (d) and also Hedges' \hat{g} . The statistical significance of the mean weighted effect sizes was determined by calculating 95 percent confidence intervals (95% CIs). If the 95% CI does not include zero then the mean effect size is significant (Lipsey & Wilson, 2001).

Interpreting effect sizes. Cohen (1988) provides rough guidelines for interpreting weighted effect sizes. Effect sizes ranging from 0.20 to 0.50 represent small effects, 0.50 to 0.80 represent moderate effects, and 0.80 or greater represent large effects. In addition to interpreting the effect sizes in this way, Cohen's (1988) U_3 will be calculated for each weighted effect size (\hat{g}) to indicate the degree of overlap between the control and experimental groups. As an example, if $\hat{g} = .32$ for the contrast, debriefed versus not debriefed, then Cohen's $U_3 = .63$ (values of U_3 are obtained from a table provided by Cohen, 1988). This means that the average individual, who was debriefed, would have an outcome (e.g., a reduction in PTSD symptoms), which is greater than 63% of the individuals who were not debriefed.

Effect size independence. Two effect sizes from the same study, which relate to the same underlying construct, are considered to be non-independent (Glass et al., 1981; Lipsey & Wilson, 2001; Rosenthal, 1984). In the current meta-analysis, if multiple effect sizes for the same outcome measure (e.g., reduction of PTSD symptoms) were encountered in the same study, the separate effects were averaged to provide a single effect size. As indicated above, this is common practice in meta-analytic research.

Examining the impact of potential moderator variables. The fixed effects model assumes homogeneity of effect sizes, however if the distribution of effect sizes is found to be heterogeneous various moderators may explain the variation (Lipsey & Wilson, 2001). In general, a moderator variable "is a qualitative... [e.g., gender] or quantitative... [e.g., time of debriefing] variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable" (Baron & Kenny, 1986, p. 1174). The moderator variables included in the current thesis

are a mixture of sample characteristics and factors that have been identified by previous research as likely to influence PD effectiveness.

Initially, overall tests of heterogeneity were conducted on the moderators of interest based on the inclusion of individual outcome measures and all aggregate categories, as outlined above (Hedges, 1994). This was done using the Q statistic, which follows a chi-square distribution (Wolf, 1986). A significant Q indicates a heterogeneous subset of effect sizes (beyond subject-level sampling error). For example, if a significant Q statistic is found for a subset of effect sizes obtained from studies examining the effectiveness of PD for emergency workers versus primary victims, there is significant variability in the effect sizes within these studies.

Using a method of analysis that is analogous to ANOVA, the total variability of the effect sizes was then partitioned into the portion of the variance explained by the moderator variable of interest (e.g., emergency workers versus primary victims) (Q_b) and the residual or remaining portion (Q_w) (Lipsey & Wilson, 2001). More specifically, in this sort of analysis, Q_b is equal to the variation between the means (of the effect sizes) for each category of a moderator variable (e.g., emergency workers versus primary victims), and Q_w is equal to the variation within the categories of a moderator variable. According to Lipsey and Wilson (2001),

If significant variability is explained by the [moderator] variable (a significant Q_b) then the mean effect sizes across categories differ by more than sampling error, that is, show a statistically significant difference. If Q_w is not statistically significant, the [moderator] variable represented in Q_b is sufficient to account for the excess variability in the effect size distribution. (p. 136)

A significant Q_b , combined with a significant Q_w , indicates that the variance explained by the significant moderator is not sufficient to explain all the excess variance across studies (Wilson, 2001).

Results

Twenty-one research studies were included in this meta-analysis. Two of these studies had multiple independent groups (e.g., different groups of individuals that received different forms of debriefing) and so they were separated to represent different studies.² This meant that 24 independent studies were analyzed (see Table 3 for a list of all studies). Analysis of these studies produced effect sizes based on the unaggregated and aggregated outcome measures selected for investigation.

² These studies are indicated with a double asterisk in the reference list.

Table 3

Studies Included in the Current Meta-analysis

Studies Included	Type of Event	Model of Debriefing	Outcome Measure(s)
Bisson et al. (1997)	Accidents	Adapted from Mitchell's CISD	PTSD symptoms, Anxiety, Depression
Bohl (1991)	Critical incident (police)	Mitchell's CISD	Anxiety, Depression, Anger, General health & functioning
Bohl (1995)	Critical incident (fire)	Adapted from Mitchell's CISD	Anxiety, Depression, Anger, General health & functioning
Carlier et al. (2000)	Critical incident (police)	Adapted from Mitchell's CISD	PTSD symptoms
Conlon et al. (1999)	Accidents	Counselling session	PTSD symptoms
Deahl et al. (1994)	War exposure/body recovery	Dyregrov's model	PTSD symptoms
Deahl et al. (2000)	War exposure/body recovery	Combination of Dyregrov & Mitchell	PTSD symptoms, Anxiety, Depression, General health & functioning
Engelmann (1997)	Unspecified (occupational)	Other	General health & functioning
Humphries & Carr (2001)	Multiple types of trauma	Combination of Dyregrov & Mitchell	PTSD symptoms

Leonard & Alison (1999)	Critical incident (police)	Mitchell's CISD	Anger, General health & functioning
Marchand et al. (2006)	Violent acts	Adapted from Mitchell's CISD	PTSD symptoms
Matthews (1998)	Unspecified (occupational)	Mitchell's CISD	PTSD symptoms
Mayou et al. (2000)	Accidents	Other	PTSD symptoms, Anxiety, Depression, General health & functioning
Nurmi (1999)	War exposure/body recovery	Mitchell's CISD	PTSD symptoms
Regehr & Hill (2000)	Critical incident (fire)	Other	PTSD symptoms, Depression
Rose et al. (1999)	Violent acts	Adapted from Mitchell's CISD	PTSD symptoms, Depression
Sijbrandij et al. (2006)	Multiple types of trauma	Adapted from Mitchell's CISD	PTSD symptoms, Anxiety, Depression
Stallard et al. (2006)	Accidents	Dyregrov's model	PTSD symptoms, Anxiety, Depression, General health & functioning
Warren (1995)	Unspecified (occupational)	Mitchell's CISD	PTSD symptoms
Wee et al. (1999)	Unspecified (occupational)	Mitchell's CISD	Anger
Young (2003)	Critical incident (police)	Mitchell's CISD	PTSD symptoms, Depression

Analysis of Main Effects

Unweighted mean effect sizes (d), Hedges' \hat{g} , and their respective distributions, were calculated for unaggregated and aggregated outcome measures. For clarity, results of the aggregated outcome measures will be the focus of this analysis (see Table 4).

Examination of the effect sizes for the unaggregated outcome measures can be found in Appendix D.

Table 4

Unweighted Mean d , Hedges' \hat{g} , Significance Levels, and U_3 of Debriefing Versus No Debriefing for Aggregated Outcome Measures

Contrast (k) ³	N	$d(SD)$	CI_d	$\hat{g}(SD)$	$CI_{\hat{g}}$	U_3
Overall outcome measure (24)	2235	.11(.05)	.02 to .19	.11(.04)	.02 to .19	.54
PTSD symptoms (20)	1916	.09(.04)	.01 to .16	.09(.04)	.02 to .16	.54
Anxiety (8)	888	.09(.07)	-.05 to .23	.09(.07)	-.05 to .22	.54
Depression (11)	1207	.16(.06)	.04 to .27	.16(.06)	.04 to .27	.56
Anger (4)	257	.41(.08)	.26 to .56	.41(.08)	.25 to .55	.66
General health & functioning (7)	677	.27(.07)	.13 to .41	.27(.07)	.13 to .40	.60

Note. k = number of effect sizes; N = total sample size, d = mean effect size; SD = standard deviation; CI = 95% confidence interval around the effect size; \hat{g} = Hedges' effect size; U_3 = Cohen's U_3 (Cohen, 1988) for effect size interpretation.

In terms of the overall effect of debriefing, $\hat{g} = .11$ ($SD = .05$), with an associated 95% CI of .02 to .19, and a U_3 value of .54. This indicates that there is a slight positive effect with regard to debriefing following a critical incident, whereby the average individual who participated in a debriefing had a more positive outcome than 54% of

those who did not participate in a debriefing. However, given that the width of the 95% *CI* exceeds .10, this finding should be viewed with the appropriate caution until replicated (Snook, Eastwood, Gendreau, Goggin, & Cullin, 2007).⁴

An examination of the aggregated outcome measures revealed more specific findings. The aggregate category of PTSD symptoms ($\hat{g} = .09$, $SD = .04$, $CI = .02$ to $.16$, $U_3 = .54$), anxiety ($\hat{g} = .09$, $SD = .07$, $CI = -.05$ to $.22$, $U_3 = .54$), and depression ($\hat{g} = .16$, $SD = .06$, $CI = .04$ to $.27$, $U_3 = .56$) were associated with very small effect sizes, indicating a very slight effect of debriefing. The average effect sizes associated with general health and functioning ($\hat{g} = .27$, $SD = .07$, $CI = .13$ to $.40$, $U_3 = .60$) and anger ($\hat{g} = .41$, $SD = .08$, $CI = .25$ to $.55$, $U_3 = .66$) were slightly higher, indicating that debriefing had a small effect on the general functioning and anger levels of individuals taking part in the debriefing session(s).

Moderator Analysis

As previously stated, effect sizes are only meaningful if they are consistent across studies. If the effect sizes are highly inconsistent this indicates that a single effect size does not properly explain what transpired in the collection of studies (Wilson, 2001). Not surprisingly, given the inconsistencies found within PD research, the present sample of studies revealed a heterogeneous effect size distribution. This indicates that the variability in effect sizes is significantly more than would be expected by mere sampling error (Lipsey & Wilson, 2001). More specifically, results revealed a significant overall Q for the overall outcome measure that compared the debriefed and non-debriefed group. Also,

³If a study had multiple time points when assessments were taken (e.g., 3 months and 6 months), the first time point was used in order to make results as consistent as possible with those studies that only had one time point of assessment.

⁴This should be the case when viewing any finding reported in this thesis where the 95% *CI* exceeds .10.

a significant overall Q was found for each of the aggregated outcome measures that were examined (see Table 5).

Table 5

Analysis of Heterogeneity for the Outcome Measures

Moderating Variables	$Q(df)$	p
Overall outcome measure	121.37 (23)	.000
PTSD symptoms	80.83 (20)	.000
Anxiety	58.17 (7)	.000
Depression	63.09 (10)	.000
Anger	23.27 (3)	.000
General health & functioning	52.74 (6)	.000

A heterogeneous distribution of effect sizes is not cause for alarm and is often the desired outcome of a meta-analysis (Wilson, 2001). This result was to be expected for the current meta-analysis due to the moderators that have been highlighted by the previous literature. This heterogeneity justifies the further exploration of moderators, as it suggests that other variables (beyond simply being debriefed) will influence the effectiveness of PD.

While most moderators were analyzed as precisely as possible using the categories outlined in the original coding manual, slight changes to the categorization of some variables was required to make the moderator analysis possible. For example, for a single moderator variable (e.g., model of PD), one level of the moderator (e.g., Mitchell's 6-stage CISD) may have consisted of too few effect sizes. In these cases, levels of the

moderator variables had to be combined with others (e.g., combining Mitchell's 6-stage CISD model with the 7-stage model).⁵

Standard moderators used in meta-analyses. Meta-analytic literature highlights several moderators found to generally influence effect sizes. Four of these moderators were examined in this thesis: publication source, affiliation of authors, assignment of conditions, and the quality of the study. However, given that these moderator variables do not relate substantially to the research being conducted here, only the major results emerging from this analysis are highlighted below. A more detailed description of the results is provided in Appendix E. It should be mentioned that some outcome measures (e.g., anger) were unable to be examined in relation to the above moderators due to an insufficient number of studies.⁶

With respect to publication source, larger effects were generally found for non-peer reviewed versus peer-reviewed publications. The affiliation of authors revealed no consistent pattern of effect sizes with regards to academic, private, or multi-disciplinary affiliations; however the catch-all category of "Other" had large effect sizes showing that these authors reported results that were highly supportive of PD. The moderator variable related to assignment of conditions showed that non-randomized studies resulted in small to large positive effects, with randomized studies having near-zero effect sizes.

With respect to the quality of study moderator, this was coded on a 6-point scale with 0 indicating poor quality and 5 indicating high quality.⁷ Inter-rater reliability on the quality of study moderator was 83%. Scores were based on certain recognized quality

⁵ This combining procedure was also used with the following moderator variables: role of individual, identity of the debriefing facilitator, and type of incident.

⁶ Unless otherwise stated, any outcome measure or moderator variable category that is absent from a table is due to the fact that an insufficient number of studies made it impossible to analyze.

criteria (e.g., randomized experimental/control groups, baseline measures, sample size exceeding 100, etc.) (Public Safety Canada, 2007). The results of the moderator analysis indicated no clear pattern of effect sizes across the various quality ratings. Surprisingly, however, studies rated as low quality often did produce reasonably large effect sizes.

Moderators that are substantively related to PD. Other types of moderators that were specifically identified from a review of the debriefing literature were also analyzed in an attempt to determine their influence on the effectiveness of PD. The results of these moderators will be presented in the order they were listed in the previously stated research questions.

Model of debriefing. The model of debriefing was coded according to the coding manual. However, as previously stated, some types of models had to be collapsed to form broader categories to make analysis possible (i.e., to increase the sample size). This resulted in the following models being included in the analysis: Mitchell's CISD model, adapted version of Mitchell's CISD, Dyregrov's model, a combination of Mitchell's and Dyregrov's model, and an "Other" category. Effect sizes for each model and the significant Q_b statistics for all outcome measures can be found in Table 6. Results indicate that Mitchell's CISD model consistently produces significantly larger effect sizes, typically in the moderate to large range, compared to all other models of PD.

⁷ No studies were coded as having a quality of 5, 4 were rated as 4, 7 as 3, 7 as 2, 3 as 1, and 3 as 0.

Table 6.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Model of Debriefing Moderator for all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
Mitchell's CISD model	9	.67(.10)	.48 to .85	51.41(4)	.000	69.44(19)	.000
Adapted from Mitchell's CISD model	7	-.05(.07)	-.18 to .08				
Dyregrov's model	2	.07(.15)	-.22 to .36				
Combination of Mitchell & Dyregrov	2	.28(.18)	-.06 to .62				
Other	4	-.26(.12)	-.49 to -.04				
PTSD symptoms							
Mitchell's CISD model	7	.61(.11)	.39 to .82	43.62(4)	.000	37.77(16)	.002
Adapted from Mitchell's CISD model	6	-.15(.07)	-.28 to -.02				
Dyregrov's model	2	-.02(.15)	-.30 to .27				
Combination of Mitchell & Dyregrov	2	.23(.17)	-.11 to .56				
Other	4	-.29(.11)	-.52 to -.07				
Depression							
Mitchell's CISD model	2	.83(.20)	.44 to 1.22	14.86(2)	.001	47.08(6)	.000
Adapted from Mitchell's CISD model	5	.09(.08)	-.06 to .24				
Other	2	-.09(.14)	-.36 to .19				
General health & functioning							
Mitchell's CISD model	2	.91(.19)	.55 to 1.28	9.32(1)	.002	13.46(2)	.001
Dyregrov's model	2	.19(.15)	-.10 to .47				

Context of debriefing. The context in which debriefings were performed (in a group versus individually) was also analyzed. Findings showed significant Q_b statistics for all outcome measures except anxiety ($p = .181$) (see Table 7). An examination of the effect sizes revealed that providing PD to groups was more effective than providing PD to individuals (although, not always at a significant level). Group-based PD was always associated with small, but positive effect sizes. In contrast, individual-based PD resulted in no effect or slightly negative effects.

Table 7.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Context of the Debriefing Moderator for all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
Group	13	.27(.07)	.14 to .40	14.87(1)	.000	63.27(19)	.000
Individual	8	-.09(.06)	-.21 to .04				
PTSD symptoms							
Group	11	.16(.07)	.02 to .30	7.65(1)	.006	61.25(17)	.000
Individual	8	-.11(.06)	-.24 to .01				
Anxiety							
Group	2	.14(.15)	-.15 to .44	1.79(1)	.181	2.28(5)	.809
Individual	5	-.08(.08)	-.25 to .08				
Depression							
Group	4	.32(.11)	.09 to .54	5.48(1)	.019	20.73(8)	.008
Individual	6	.01(.07)	-.13 to .14				
General health & functioning							
Group	4	.37(.12)	.13 to .61	3.98(1)	.046	32.79(4)	.000
Individual	2	.01(.13)	-.25 to .27				

Type of incident. Studies included a range of diverse traumatic events and so the type of incident subcategories had to be aggregated for analysis. Table 8 shows significant Q_b statistics for all categories except anxiety ($p = .643$), indicating that the type of incident leading up to the PD has a significant impact on the effectiveness of the debriefing. Based on the analysis, PD seems to be most effective for individuals exposed to war exposure/body recovery and with emergency service personnel dealing with critical incidents. Studies examining these types of incidents resulted mainly in small to medium effect sizes. In contrast, violent acts, accidents, and multiple types of exposure did not show any strong effects across any of the five outcome measures.

Table 8.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Type of Incident Moderator for all Outcome Measures

Moderating Variables	<i>k</i>	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	<i>p</i>	$Q_w (df_w)$	<i>p</i>
Overall outcome measure							
Accidents	4	-.08(.11)	-.29 to .14	22.17(4)	.000	99.20(19)	.000
Violent acts	2	-.08(.10)	-.28 to .12				
Emergency services critical incident	10	.15(.08)	.00 to .30				
War exposure/body recovery	5	.53(.11)	.31 to .74				
Multiple types of exposure	3	-.02(.12)	-.25 to .21				
PTSD symptoms							
Accidents	4	-.11(.11)	-.33 to .10	23.51(4)	.000	60.16(17)	.000
Violent acts	2	-.10(.10)	-.30 to .10				
Emergency services critical incidents	8	-.10(.08)	-.26 to .07				
War exposure/body recovery	5	.49(.11)	.28 to .71				
Multiple types of exposure	3	.00(.12)	-.24 to .23				
Anxiety							
Accidents	3	-.05(.11)	-.27 to .17	.22(1)	.643	1.85(3)	.605
Multiple types of exposure	2	-.13(.13)	-.37 to .12				
Depression							
Accidents	3	.01(.11)	-.21 to .23	18.99(2)	.000	42.98(6)	.000
Emergency services critical incidents	4	.59(.13)	.34 to .84				
Multiple types of exposure	2	-.13(.12)	-.37 to .11				
General health & functioning							
Accidents	2	.02(.14)	-.26 to .30	31.37(2)	.000	21.37(4)	.000
Emergency services critical incidents	3	1.05(.15)	.75 to 1.36				
War exposure/body recovery	2	-.01(.16)	-.32 to .29				

Identity of the debriefing facilitator. As a variety of different individuals can perform a debriefing, the identity of the debriefer was examined. Significant Q_b statistics were found for overall outcome measures, PTSD, and general health and functioning, but not for anxiety ($p = .454$) or depression ($p = .647$) (see Table 9). No consistent pattern was found regarding this moderator.

Table 9.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Identity of the Debriefing Facilitator Moderator for all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
Health professional(s)	5	-.12(.09)	-.29 to .05	12.93(4)	.012	107.80(19)	.000
Researcher(s)	6	.29(.08)	.14 to .44				
Trained peers & other(s)	5	.08(.11)	-.13 to .29				
Unspecified	5	.08(.11)	-.14 to .30				
Other	3	.18(.15)	-.11 to .47				
PTSD symptoms							
Health professional(s)	4	-.26(.09)	-.44 to -.08	24.07(4)	.000	57.31(16)	.000
Researcher(s)	6	.26(.08)	.10 to .41				
Trained peers & other(s)	4	-.05(.12)	-.28 to .18				
Unspecified	5	.04(.11)	-.18 to .26				
Other	2	-.38(.17)	-.72 to -.04				
Anxiety							
Health professional(s)	4	-.09(.10)	-.28 to .10	.56(1)	.454	3.07(4)	.547
Researcher(s)	2	.03(.13)	-.23 to .29				
Depression							
Health professional(s)	4	.01(.10)	-.18 to .20	.87(2)	.647	23.29(6)	.001
Researcher(s)	3	.13(.01)	-.05 to .31				
Unspecified	2	.07(.14)	-.20 to .34				
General health & functioning							
Researcher(s)	2	.01(.13)	-.25 to .27	7.50(1)	.006	19.23(2)	.000
Other	2	.65(.19)	.28 to 1.03				

Role of individual. The individuals involved in a critical incident were categorized according to their role (either occupational or victim) in the incident. Similar to all other tested moderators, significant Q_b statistics were found for all outcome measures (see Table 10). The effect sizes showed a consistent pattern whereby those in an occupational role had higher positive effect sizes, and therefore experienced greater benefits from PD than those who were in a victim role.

Table 10.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Role of the Individual Moderator for all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure	15	.27(.06)	.50 to .40	16.08(1)	.000	99.55(21)	.000
Occupational Victim	8	-.09(.06)	-.21 to .04				
PTSD symptoms	12	.10(.07)	-.04 to .23	4.96(1)	.026	69.98(18)	.000
Occupational Victim	8	-.11(.06)	-.24 to .01				
Anxiety	3	.54(.14)	.28 to .81	15.54(1)	.000	42.63(6)	.000
Occupational Victim	5	-.08(.08)	-.25 to .08				
Depression	5	.52(.11)	.31 to .73	16.61(1)	.000	46.48(9)	.000
Occupational Victim	6	.01(.07)	-.13 to .14				
General health & functioning	5	.53(.11)	.31 to .75	8.97(1)	.003	43.80(5)	.000
Occupational Victim	2	.01(.13)	-.25 to .27				

Timing of debriefing. The timing in which a debriefing is carried out post-critical incident was analyzed. All Q_b statistics were significant and while there was no clear chronological pattern, Table 11 shows some consistencies. For instance, conducting a debriefing within 3 days of a critical incident resulted in positive effect sizes (with the exception of the PTSD outcome measure). Conducting a debriefing within 7 days of an incident resulted in large positive effects for the overall outcome measure and for PTSD. Conducting a debriefing at any time beyond seven days post-incident resulted in weaker effects, including negative effect sizes.

A related analysis examined the timing of the assessments (either short-term or long-term) post-debriefing. Studies that had assessments within 6 months of the debriefing had more positive effects, especially for depression and general health and functioning, compared to assessments taken after 6 months (see Table 12).

Table 11.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Timing of the Debriefing Moderator for all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
Within 3 days	6	.30(.10)	.12 to .49	40.10(3)	.000	67.20(15)	.000
Within 7 days	5	.81(.13)	.55 to 1.06				
Within 14 days	6	-.15(.08)	-.13 to .02				
Within 30 days	2	.10(.10)	-.09 to .30				
PTSD symptoms							
Within 3 days	3	-.07(.13)	-.32 to .17	39.66(3)	.000	28.73(12)	.004
Within 7 days	5	.81(.13)	.55 to 1.06				
Within 14 days	6	-.16(.08)	-.32 to .01				
Within 30 days	2	.05(.10)	-.14 to .25				
Anxiety							
Within 3 days	3	.52(.14)	-.08 to .25	14.80(1)	.000	43.37(4)	.000
Within 14 days	3	-.16(.11)	-.36 to .05				
Depression							
Within 3 days	3	.78(.14)	.50 to 1.05	28.64(2)	.000	28.38(5)	.000
Within 14 days	3	-.16(.10)	-.37 to .04				
Within 30 days	2	.15(.10)	-.05 to .34				

Table 12.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Assessment Time Point Post-debriefing Moderator for all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
Within 6 months (short-term)	13	.10(.06)	-.02 to .23	3.13(1)	.077	72.28(15)	.000
After 6 months (long-term)	4	-.10(.09)	-.28 to .08				
PTSD							
Within 6 months (short-term)	11	-.04(.07)	-.18 to .09	0.32(1)	.573	25.04	.023
After 6 months (long-term)	4	-.11(.09)	-.29 to .07				
Anxiety							
Within 6 months (short-term)	5	.18(.09)	.00 to .36	7.05(1)	.008	59.52(6)	.000
After 6 months (long-term)	3	-.22(.12)	-.44 to .01				
Depression							
Within 6 months (short-term)	5	.32(.09)	.13 to .50	8.13(1)	.004	61.18(7)	.000
After 6 months (long-term)	4	-.05(.09)	-.23 to .13				
General health & functioning							
Within 6 months (short-term)	4	.58(.12)	.35 to .82	8.86(1)	.003	43.84(4)	.000
After 6 months (long-term)	2	.02(.14)	-.26 to .30				

Debriefing option. Studies were also coded according to whether or not it was mandatory for the group that received a debriefing to participate in the debriefing (e.g., occupational requirements), as opposed to voluntary debriefings. Once again, across all outcome measures, Q_b was significant. More specifically, results revealed that mandatory debriefing consistently resulted in significantly higher positive effect sizes compared to voluntary (or not stated) debriefings, which resulted in near-zero effect sizes (see Table 13).

Interactions between moderators. As indicated previously, given previous findings on PD, it would be very useful to examine interactions between certain moderator variables. For example, existing research suggests that the role of the individual taking part in the debriefing may interact with the optimal time for the debriefing (with emergency personnel benefiting more from a delay in debriefing compared to civilians) (Mitchell, 1983). Analytically, it is possible to examine these sorts of interactions in a meta-analysis (Uziel, 2007). However, a relatively large number of studies are required to provide the necessary statistical power to detect interactions. It was regrettable that the small number of studies in the current meta-analysis (especially after studies were subcategorized by each moderator variable) did not allow any of these interactions to be examined in an appropriate fashion.

Table 13.

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Debriefing Option Moderator for all Outcome Measures

Moderating Variables	<i>k</i>	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	<i>p</i>	$Q_w (df_w)$	<i>p</i>
Overall outcome measure							
Mandatory	10	.39(.08)	.24 to .54	21.54(1)	.000	99.19(22)	.000
Voluntary/not stated	14	-.04(.05)	-.15 to .06				
PTSD symptoms							
Mandatory	7	.21(.09)	.04 to .38	8.25(1)	.004	73.14(19)	.000
Voluntary/not stated	14	-.09(.06)	-.19 to .02				
Anxiety							
Mandatory	2	.99(.19)	.62 to 1.37	25.93(1)	.000	32.24(6)	.000
Voluntary/not stated	6	-.06(.08)	-.21 to .09				
Depression							
Mandatory	2	1.38(.19)	1.01 to 1.76	44.84(1)	.000	18.25(9)	.032
Voluntary/not stated	9	.04(.06)	-.08 to .15				
General health & functioning							
Mandatory	4	.79(.13)	.52 to 1.05	32.62(5)	.000	20.14(1)	.000
Voluntary/not stated	3	.01(.11)	-.21 to .22				

Discussion

The current thesis represents the most inclusive meta-analytic review regarding the effectiveness of PD to date. This project used a quantitative approach in order to determine whether or not debriefing was effective in helping individuals following a critical incident. The general effectiveness of PD was examined, as were key moderating variables that were believed to be influencing its effectiveness.

Is Debriefing Effective?

The main effects analysis demonstrated that those individuals who participated in a debriefing had fewer adverse psychological symptoms (as recorded by the outcome measures) than did those who did not participate in a debriefing. However, while this was the general finding, the positive effect of a debriefing was minimal (with effect sizes ranging from .09 to .41). When one further examines the subcategories of the outcome measures, it is clear that debriefings help reduce certain types of symptoms better than others. For instance, PD appears to alleviate problems with anger reasonably well, and has a small but potentially important impact on an individual's general health and functioning. Surprisingly, debriefings were found to be far less effective in reducing PTSD symptoms and associated problems, specifically anxiety and depression. Interestingly, these are the types of symptoms that are often the focus of PD (Carlier et al., 2000; Mitchell, 1983).

Upon careful consideration, these results do make sense. Arguably, anger and problems with general health and functioning may be regarded as symptoms that are more likely to improve from debriefings, since treatment for such problems is often some form of individual or group therapy (Taft & Niles, 2004). It is perhaps less likely for a

debriefing to reduce other PTSD-related symptoms (e.g., anxiety, depression), because these may require additional forms of therapy (e.g., cognitive-behavioural therapy, pharmacotherapy) (Veterans Affairs Canada, 2004). In fact, this line of thinking accords well with arguments made by Mitchell (2004), who has always argued that CISD needs to be part of a more comprehensive stress management intervention (e.g., CISM) in order to be effective in reducing PTSD. Thus, it may be unfair to evaluate models such as Mitchell's CISD based on their ability to reduce PTSD symptoms, despite the fact many rely on CISD, and CISD-type interventions, to mitigate problems associated with PTSD.

In addition, it is important to re-emphasize methodological limitations that were encountered in this study, as they may play a role in explaining certain results emerging from the analysis of main effects. Perhaps most obvious are the potential problems caused by basing a meta-analysis on a small sample size. This was a particular issue for certain outcome measures, most notably anger. Given the small sample size upon which this analysis was based, it is possible that the effect size associated with anger is over-inflated, giving an unrealistic representation of the effectiveness of PD for reducing symptoms of anger. In fact, the width of the 95% *CI*s for all outcome measures used in the main effect analysis was unacceptably large (exceeding .10). Therefore, all of these results should be interpreted with an appropriate degree of caution until the results are replicated. The same can be said when considering the results of the moderator analyses.

Which Moderators Impact the Effectiveness of Debriefing?

Numerous moderator variables were analyzed in order to determine their impact, if any, on the effectiveness of PD. Most moderators chosen for analysis ended up having

a moderating effect on the effectiveness of debriefing, and these variables will be discussed here in more detail.

The analysis of moderators that are typically used in meta-analyses produced interesting results. First, non-peer reviewed studies (e.g., internal reports) reported larger effects than peer-reviewed publications, which is the opposite of what would be expected given publication biases (Lipsey & Wilson, 1993). Second, the affiliation of authors moderator did not reveal any meaningful results because the only category showing high effect sizes was the “Other” category. Third, non-randomized studies resulted in larger effect sizes than randomized studies, which can likely be explained by several important group differences that may emerge when using such an experimental design. For instance, in a non-random study, individuals who choose to take part in debriefing sessions may have more insight into the problems they are facing, be more engaged in the healing process, or have fewer/less severe psychological problems compared to individuals who do not choose to participate in PD. Finally, in terms of the quality of study moderator, lower quality studies were generally found to produce larger effect sizes compared to higher quality studies. It is not clear why this result was found, although it is possible that the criteria used to define study quality in this study were simply not appropriate. Defining study quality is a difficult task, as there are many potentially relevant variables that can be considered. In the future, more consideration should be given to the variables that other researchers have focused on for this purpose, such as those in the area of sex offender research (e.g., Public Safety Canada, 2007).

Arguably, the moderators of most interest to this thesis were those directly related to PD. While previous research, much of which is anecdotal, has identified certain

moderators as having the potential to impact debriefing effectiveness, this thesis was able to quantify those moderators to determine which of them had the greatest effect on PD. Based on the significant Q_b statistics found for all the moderators examined in this thesis, it is clear that these variables do play an important role in determining the effectiveness of PD. However, it is important to highlight that most of these moderators were also associated with significant Q_w statistics. This indicates that, while the moderator may account for a significant portion of the variability in effect sizes, that moderator alone does not account for all of the excess variability (Wilson, 2001). It is therefore necessary to be cautious before making conclusions about the impact of specific moderators in the current study.

Regarding the model of debriefing, Mitchell's CISD model was found to be the most effective. This finding was not overly surprising, given that previous research supports that CISD is the benchmark against which most debriefing models are created and evaluated (Everly & Boyle, 1999; Mitchell & Everly, 2000). However, despite this result, it is unclear as to what element(s) within the CISD procedure makes it more effective than other models. For example, it may be the actual CISD phases themselves or the way in which each phase is carried out that separates this model from other models. Alternatively, CISD may be more effective than other PD models because it has received the most support for its development (e.g., from organizations such as The International Critical Incident Stress Foundation). It is also possible that the studies examining CISD were conducted in a manner that differed from the studies examining other PD models.

It is interesting to note that studies which used an adapted version of Mitchell's CISD produced significantly lower effect sizes, perhaps indicating that the strict

protocols developed by Mitchell need to be followed precisely in order for CISD to be effective. More specifically, Dyregrov's model of PD, combination models, and less well defined debriefing models (e.g., counselling sessions) showed no clear indication that debriefing was effective in reducing any of the problematic symptoms caused by critical incidents. It was regrettable that, due to low sample sizes, more of the individual PD models could not be examined.

The context in which debriefings are performed was also found to be very important in that group debriefing sessions appear to be significantly more effective than individual sessions. This is consistent with some PD literature which states that a group atmosphere allows individuals to communicate with others who have experienced similar traumatic events, which can help to validate stress reactions (e.g., Mitchell & Everly, 1997). It is also consistent with literature from other fields, which is generally supportive of group therapy (e.g., correctional psychology) (Byrnes, Hansen, Malloy, Carter, & Curry, 1999).

Related to the context of debriefing, the specific type of traumatic event to which individuals were exposed was analyzed. For the first time, a wide range of events were examined as potential moderators and it appears as though PD is more effective for dealing with problems that emerge from certain types of events, particularly secondary traumatic events. More specifically, the largest effect sizes were found when PD was used to help those with war exposure/body recovery and emergency personnel-related critical incident exposure. Substantially lower effect sizes were found for primary traumatic events, such as exposure to violent acts and accidents. This makes sense given that tackling stress for those who have experienced secondary trauma was PD's original

intent. Indeed, PD was originally designed to help military personnel (Mitchell & Everly, 2000), and there is significant evidence that crisis intervention strategies were used to combat stress dating back to World War I (Wohlmuth, 2003). Since that time, PD has been primarily developed for fields ranging from disaster response to law enforcement (Mitchell & Everly, 2001).

Why is PD more effective when applied to people suffering from secondary traumatization? The results from the current analysis provide few answers to this question other than the fact that the two groups of individuals must clearly differ in important ways (e.g., in terms of the actual degree of stress that they experience). Such differences between the groups are certainly well established. For example, primary and secondary sufferers express different needs post-trauma, with victims being primarily worried about safety (Giller, 1998), and secondary sufferers (e.g., occupational personnel) being primarily worried about how soon they can return to their jobs (Renck, Weisaeth, & Skarbo, 2002). Additional research will be required to examine how these types of differences relate to PD effectiveness.

Directly related to the analysis of PD context and type of incident is the analysis of the role of the individual within the critical incident (victim versus occupational). This analysis revealed that those in an occupational role benefited far more from debriefings than did victims. Given what was found with respect to PD context and type of incident, this is not at all surprising. Thus, a clear picture is beginning to emerge, whereby PD appears to be most effective for individuals who experience secondary traumatic events as occupational workers and participate in debriefing sessions with others who have experienced similar trauma.

Findings related to the previously discussed moderators make other moderators even more important. First, it is essential to examine the impact of the facilitator of the debriefing since the debate over who should deliver PD is most important for the occupational population, as they are the population who has the option of a peer facilitated debriefing. Second, whether or not a debriefing is voluntary or mandatory is also important for occupational personnel, given that, once again, it is this population that is often required to participate in debriefings as part of occupational health and safety protocols.

Unfortunately, no real clear relationship was found regarding the facilitator of the debriefing. Arguably, it was when researchers facilitated the debriefing that debriefed individuals experienced the best results, while trained peers and “Other” debriefers showed no real influence in either direction. Perhaps the biggest surprise when analyzing this moderator variable was that, when MHPs played the role of debriefer, they produced more problems in the individuals they were debriefing. Given previous research on this issue, this finding can probably best be explained by the fact that those in occupational roles typically do not trust MHPs. For example, these individual often believe that MHPs do not genuinely care about them and sometimes look down upon them (Woody, 2005). In addition, occupational personnel may believe that MHPs cannot possibly understand the types of trauma to which these workers are regularly exposed (Simms-Ellis & Madill, 2001).

Regarding the option of attending a debriefing session, it is clear that mandatory debriefings had a positive influence on the effectiveness of PD for all outcome measures, whereas studies examining voluntary debriefings did far worse. This suggests that those

who would choose not to attend a voluntary debriefing, possibly because they do not want to appear weak or admit they need help, may actually benefit from participation if forced to do so. The only concern with these results is that it is possible that this variable was correlated with other variables since mandatory debriefings are only standard for occupational personnel and not for victims.

The final moderator of interest was the optimal time to perform a debriefing. While results were rather ambiguous, it appears as though debriefings led to the best results when performed sooner rather than later. Debriefings performed within 3 or 7 days had larger positive effects than those done within 14 or 30 days. This finding is consistent with what certain researchers have argued (e.g., Campfield & Hills, 2001; Mitchell, 2004), and suggests that relatively early intervention is important so as to shorten the suffering post-trauma and prevent the on-set of secondary maladaptive coping behaviours (e.g., substance abuse).

While this finding is important, it would have been more useful if the analysis could have been more precise in order to allow for an examination of shorter time periods (e.g., debriefings within 24 hours, 48 hours, 72 hours, etc.), given that these time frames are central to ongoing debates (Busuttil & Busuttil, 1995; Everly, 1995; Kenardy et al., 1996). The problem, of course, is that few studies examined debriefings that were performed less than 3 days post-trauma. That is not to say that debriefings are not typically offered that quickly, but rather that debriefings which are performed quickly (e.g., after school shootings) are simply not the focus of empirical research. In addition, many times with severe trauma there are many other primary concerns (e.g., proper medical care, the safety of loved ones) making an immediate debriefing impossible.

A related analysis examined the timing of the assessments (either short-term or long-term) post-debriefing. This was done in an attempt to see if the impact of a debriefing was only short-term in nature. This concern was supported, whereby assessments taken within 6 months post-debriefing had larger positive effects than assessments taken after 6 months. However, one would want to be cautious when interpreting these results as the sample size of studies obtaining assessments after 6 months was small.

Limitations of this Research

This study improved on several previous anecdotal and narrative reviews by using a quantitative approach to examine the PD literature. It was also able to improve on the previous meta-analyses on this topic, because this thesis included more studies than ever before and examined a wider-range of moderators. However, despite these advances there were still several noteworthy limitations.

First, as previously mentioned, small sample sizes were a problem. While the overall sample size was sufficient to conduct a meta-analysis (Cooper & Rosenthal, 1980), there were still too few studies related to certain aggregated outcome measures to conduct a sound moderator analysis (e.g., anger). While a wide-range of stress reactions are expected for those who have been involved in a critical incident, including anger-related symptoms (Artwohl & Christensen, 1997), it appears as though previous PD research has a tendency to focus primarily on PTSD symptoms.

Second, the current analysis did not take into account baseline measures of individuals' stress levels before they were exposed to PD. The results reported here could be further validated if baseline measures were compared to corresponding measures post-

debriefing for both experimental and control groups. This type of analysis was excluded from the current study for several reasons. First, it made the results more straightforward to simply compare effect sizes associated with outcome measures across experimental and control groups. Second, many PD studies do not include baseline measures post-trauma making this analysis impossible. Of the studies that do include baseline measures, many use psychometric tests that do not correspond to post-debriefing measures, which makes pre-post comparisons meaningless. Finally, baseline measures can be poor indicators of actual stress for trauma sufferers because of the impact of shock and the delayed nature of many stress reactions (Bryant & Harvey, 2002).

Third, as was made clear at the beginning of this thesis, the term PD includes a wide range of procedures and goes by several names (e.g., crisis intervention, trauma counselling, stress management, etc.). This means it is possible that studies using other types of interventions that were similar to debriefing, although not specifically defined as such, were not included in the current study. Also, several studies that examined the effectiveness of PD could not be included because they either did not have a clear experimental (that received PD) and control group (that did not receive PD), or because they lacked specific details regarding the debriefing (e.g., the sample size of each group) (e.g., Chemtob et al., 1997; Jenkins, 1996; Kenardy et al., 1996). So, while this meta-analysis attempted to be as inclusive as possible, it does recognize that some PD literature may have been excluded.

Lastly, and arguably the most important limitation, was the lack of details provided by the included studies which made some moderator analyses impossible. This is a problem that has been identified in other meta-analyses (e.g., Corey, 2006; Sitzmann,

Kraiger, Stewart, & Wisner, 2006). This absence of detail limited the type of moderators that could be examined, and potentially important moderators had to be ignored. For instance, it is likely that variables such as the severity of the trauma, the environment of the debriefing, and the amount of participation within a debriefing session could impact the effectiveness of PD (Gist & Devilly, 2002; Rick, 2000). However, these variables could not be examined. The lack of specificity when describing the actual debriefings was another concern. For example, there were often very few details that were provided with respect to the exact debriefing procedure that was used, the qualifications of the facilitator, and the topics that were discussed within the debriefing session. Without these details, it is difficult to generalize results supporting the effectiveness of PD.

Implications and Future Directions

As would be expected, given the previously reported mixed results on the effectiveness of PD, it was no surprise that results from the current study were not straightforward. Overall, results revealed a generally weak effect supporting the use of PD, especially for combating the effects of PTSD. In contrast, PD was found to be slightly more effective at reducing anger and improving one's general health and functioning. The current study also demonstrated that the effectiveness of PD increases under certain conditions, pointing to the fact that it is too simplistic to ask the question – does PD work? More specifically, the success of PD is maximized when: (1) using Mitchell's CISD model of debriefing, (2) targeting occupational personnel in group settings who have been exposed to occupationally-related traumatic events, and (3) the PD sessions are made mandatory. Despite what others have previously said in the

literature, other potential moderators, such as the identity of the PD facilitator and the exact timing of the PD sessions, appear to be less important.

The results of the main effects and moderator analyses do give some direction to practitioners and policy makers who make decisions about how PD should be implemented. However, given the relatively small sample size upon which the conclusions in this study are based, it is very important for these decision-makers to proceed with extreme caution. Before important decisions about PD can be made with confidence, further research needs to be conducted that examines the issues addressed in this thesis, especially in regards to the impact of various moderator variables on PD effectiveness.

Having said this, there is no point in continuing to study PD unless researchers begin to improve the methods they use to study the topic. Fortunately, researchers are beginning to propose gold standards for studying PD, and this research will greatly assist in improving the quality of future PD research (e.g., Foa, Keane, & Friedman, 2000). So far, quality criteria include, but are not limited to, clearly defined symptoms of interest, reliable and valid outcome measures, using multi-method assessments, the use of blind evaluators, debriefings performed by a trained facilitator, ensuring unbiased assignment to conditions, and providing numerous details about the debriefing process (e.g., environment in which the debriefing takes place). In addition, future studies could make an effort to include a wider-range of outcome measures that encompass more than traditional PTSD symptoms, as PTSD symptoms are often highly comorbid with other problems (e.g., substance abuse) (Green, 1996).

The fact that small to moderate effect sizes were the norm in this study, even under the conditions where PD was most effective, raises another important issue - the cost-effectiveness of PD. Little to no research has been conducted regarding this issue, despite its obvious importance (Arendt & Elklit, 2001). For example, stress currently costs Canadian industries over 16 billion dollars each year in terms of reduced productivity, absenteeism, and disability (Tangri, 2003). Thus, research is urgently needed that examines the specific costs associated with PD and compares these costs to the costs of other, more extensive stress management programs that have been shown to be effective (Mitchell, 2004). In turn, these costs need to be compared to the ultimate cost of helping individuals that have developed long-term trauma-related symptoms and disorders. However, it is clear is that the use of PD has outpaced the empirical research backing up its effectiveness. To remedy this situation, the quality of PD research has to improve, PD has to be more vigorously evaluated, and its cost-effectiveness has to be clearly established.

Conclusion

The current thesis represents the most inclusive quantitative review of the PD literature to date. The approach adopted here empirically summarized previous studies in order to determine the effectiveness of PD. The results supported the use of debriefing to some extent, but only under certain conditions. This brings into question the practicality and cost-effectiveness of this procedure, especially when it is used with little consideration given to contextual issues. It is regretful that, due to a range of study limitations encountered in this thesis, more solid results could not be found. In the future,

it is hoped that PD research will continue to improve so that the best possible psychological first-aid can be provided to those who are exposed to critical incidents.

References

References marked with an asterisk represent studies included in this meta-analysis.

Arendt, M., & Elklit, A. (2001). Effectiveness of psychological debriefing. *Acta*

Psychiatrica Scandinavica, *104*(6), 423-437.

Armstrong, K., O'Callahan, W., & Marmar, C. (1991). Debriefing Red Cross disaster

personnel: The multiple stressor debriefing model. *Journal of Traumatic Stress*,

4(4), 581-593.

Artwohl, A., & Christensen, L. W. (1997). *Deadly force encounters: What cops need to*

know to mentally and physically prepare for and survive a gunfight. Boulder, CO:

Paladin Press.

Baron, M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social

psychological research: Conceptual, strategic, and statistical considerations.

Journal of Personality and Social Psychology, *51*(6), 1173-1182.

Berman, J. S., & Norton, N. C. (1985). Does professional training make a therapist more

effective? *Psychological Bulletin*, *98*(2), 401-407.

Biostat, Inc. (2006). Comprehensive meta-analysis. Retrieved February 12, 2007, from

http://www.meta-analysis.com/pages/why_use.html

Bisson, J. I., & Deahl, M. (1994). Psychological debriefing and the prevention of

posttraumatic stress: More research is needed. *British Journal of Psychiatry*,

165(6), 717-720.

*Bisson, J. I., Jenkins, P. L., Alexander, J., & Bannister, C. (1997). Randomised

controlled trial of psychological debriefing for victims of acute burn trauma.

British Journal of Psychiatry, *171*(1), 78-81.

- Bisson, J. I., McFarlane, A. C., & Rose, S. (2000). Psychological debriefing. In E. Foa, T. M. Keane, & M. J. Friedman (Eds.), *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies* (pp. 39-59). New York, NY: Guilford Press.
- *Bohl, N. (1991). The effectiveness of brief psychological interventions in police officers after critical incidents. In J. Reese, J., Horn, & C. Dunning (Eds.), *Critical incidents in policing: Revised* (pp. 31-38). Washington, DC: US Government Printing Office.
- *Bohl, N. (1995). Measuring the effectiveness of CISD: A study. *Fire Engineering*, *148*(8), 125-126.
- Boudreaux, E. D., & McCabe, B. (2000). Critical incident stress management: Interventions and effectiveness. *Emergency Psychiatry*, *51*(9), 1095-1097.
- Brewin, C. R. (2001). Cognitive and emotional reactions to traumatic events: Implications for short-term intervention. *Advances in Mind-Body Medicine*, *17*, 163-168.
- Breznitz, S. (1980). Stress in Israel. In H. Selye (Ed.), *Selye's guide to stress research* (pp. 71-89). New York, NY: Van Nostrand Reinhold Company.
- Bryant, R. A., & Harvey, A. G. (2002). Delayed-onset posttraumatic stress disorder: A prospective evaluation. *Australian and New Zealand Journal of Psychiatry*, *36*(2), 205-209.
- Busuttil, A., & Busuttil, W. (1995). Psychological debriefing. *British Journal of Psychiatry*, *166*(5), 676-681.

- Byrnes, E. I., Hansen, G. K., Malloy, T. E., Carter, C., & Curry, D. (1999). Reductions in criminality subsequent to group, individual, and family therapy in adolescent residential and day treatment settings. *International Journal of Group Psychotherapy, 48*(3), 307-322.
- Campfield, K. M., & Hills, A. M. (2001). Effect of timing of critical incident stress debriefing (CISD) on posttraumatic symptoms. *Journal of Traumatic Stress, 14*(2), 327-340.
- *Carlier, I. V. E., Voerman, A. E., & Gersons, B. P. R. (2000). The influence of occupational debriefing on post-traumatic stress symptomatology in traumatized police officers. *British Journal of Medical Psychology, 73*(1), 87-98.
- Chemtob, C., Tomas, S., Law, W., & Cremniter, D. (1997). Postdisaster psychosocial intervention: A field study of debriefing on psychological distress. *American Journal of Psychiatry, 154*(3), 415-417.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). New York, NY: Academic Press.
- *Conlon, L., Fahy, T. J., Conroy, R. (1999). PTSD in ambulant RTA victims: A randomized controlled trial of debriefing. *Journal of Psychosomatic Research, 46*(1), 37-44.
- Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87*(3), 442-449.
- Corey, S. (2006). *Examining the influence of practice, feedback and fidelity on the effectiveness of use-of-force simulation training in the military domain: A meta-analytic review*. Unpublished master's thesis, Carleton University, Ottawa.

- *Deahl, M. P., Gillham, A. B., Thomas, J., Searle, M. M., & Srinivasan, M. (1994). Psychological sequelae following the Gulf War: Factors associated with subsequent morbidity and the effectiveness of psychological debriefing. *British Journal of Psychiatry*, *165*(1), 60-65.
- *Deahl, M., Srinivasan, M., Jones, N., Thomas, J., Neblett, C., & Jolly, A. (2000). Preventing psychological trauma in soldiers: The role of operational stress training and psychology debriefing. *British Journal of Medical Psychology*, *73*(1), 77-85.
- Devilly, G. J., & Cotton, P. (2003). Psychological debriefing and the workplace: Defining a concept, controversies and guidelines for intervention. *Australian Psychologist*, *38*(2), 144-150.
- Devilly, G. J., Gist, R., & Cotton, P. (2006). Ready! Fire! Aim! The status of psychological debriefing and therapeutic interventions: In the work place and after disasters. *Review of General Psychology*, *10*(4), 318-345.
- Dyregrov, A. (1989). Caring for helpers in disaster situations: Psychological debriefing. *Disaster Management*, *2*, 25-30.
- Dyregrov, A. & Gjestad, R. (2003). A maritime disaster: Reactions and follow-up. *International Journal of Emergency Mental Health*, *5*, 3-14.
- *Engelmann, N. (1997). *The impact of critical incident stress debriefing on correctional officers*. Unpublished master's thesis, Southern Connecticut State University, Connecticut.
- Everly, G. S. (1995). The role of critical incident stress debriefing (CISD) process in disaster counselling. *Journal of Mental Health Counselling*, *17*, 278-290.

- Everly, G. S., & Boyle, S. H. (1999). Critical incident stress debriefing (CISD): A meta-analysis. *International Journal of Emergency Mental Health, 3*(1), 165-168.
- Everly, G. S., Boyle, S. H., & Lating, J. M. (1999). The effectiveness of psychological debriefing with vicarious trauma: A meta-analysis. *Stress Medicine, 15*(4), 229-233.
- Everly, G. S., Flannery, R. B., & Mitchell, J. T. (2000). Critical incident stress management: The sinking of the "Jupiter". *Aggression and Violent Behavior, 5*(1), 23-40.
- Everly, G. S., & Mitchell, J. T. (1999). *Critical incident stress management (CISM): A new era and standard of care in crisis intervention* (2nd Ed.). Ellicott City, MD: Chevron Publishing Corporation.
- Flannery, R. B. Jr., & Penk, W. E. (1996). Program evaluation of an intervention approach for staff assaulted by patients: Preliminary inquiry. *Journal of Traumatic Stress, 9*(2), 317-324.
- Foa, E. B., Keane, T. M., & Friedman, M. J. (2000). *Effective treatments for PTSD*. New York: Guilford Press.
- Gibson, M., & Campling, J. (2006). *Order from chaos: Responding to traumatic events*. University of Bristol, UK: The Policy Press.
- Giller, J. (1998). Caring for "victims of torture" in Uganda: Some personal reflections. In P.J. Bracken & C. Petty (Eds.), *Rethinking the trauma of war* (pp. 128-145). London: Free Association Books.
- Gist, R., & Devilly, G. J. (2002). Post-trauma debriefing: The road too frequently travelled. *The Lancet, 360*(9335), 741-742.

- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.
- Green, B. L. (1996). Traumatic stress and disaster: Mental health effects and factors influencing adaptation. *International Review of Psychiatry, 2*, 177-210.
- Hall, J. A., & Rosenthal, R. (1995). Interpreting and evaluating meta-analysis. *Evaluation & the Health Professions, 18*(4), 393-407.
- Hattie, J. A., Sharpley, C. F., & Rogers, H. F. (1984). Comparative effectiveness of professional and paraprofessional helpers. *Psychological Bulletin, 95*(3), 534-541.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.
- Hedges, L. V. (1994). Fixed effects model. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 285-298). New York, NY: Russell Sage Foundation.
- Hobbs, M., Mayou, R., Harrison, B., & Worlock, P. (1996). A randomised controlled trial of psychological debriefing for victims of road traffic accidents. *British Medical Journal, 313*(7070), 1438-1439.
- Hoff, L. (1989). *People in crisis: Understanding and helping* (3rd Ed.). Redwood: Addison-Wesley Publishers.
- Hokanson, M., & Wirth, B. (2000). The critical incident debriefing process for the Los Angeles county fire department: Automatic and effective. *International Journal of Emergency Mental Health, 2*, 249-257.
- *Humphries, C. L., & Carr, A. (2001). The short term effectiveness of critical incident stress debriefing. *The Irish Journal of Psychology, 22*(3-4), 188-197.

- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York, NY: Sage Publications.
- Jacobs, J., Horne-Moyer, H. L., & Jones, R. (2004). The effectiveness of critical incident stress debriefing with primary and secondary trauma victims. *International Journal of Emergency Mental Health, 6*(1), 5-14.
- Jenkins, S. R. (1996). Social support and debriefing efficacy among emergency medical workers after a mass shooting incident. *Journal of Social Behavior and Personality, 11*(3), 477-492.
- Kenardy, J. A., Webster, R. A., Lewin, T. J., Carr, V. J., Hazell, P. L., & Carter, G. L. (1996). Stress debriefing and patterns of recovery following a natural disaster. *Journal of Traumatic Stress, 9*(1), 37-49.
- Latack, J. C. (1986). Coping with job stress. *Journal of Applied Psychology, 71*(3), 377-385.
- Lavender, T., & Walkinshaw, S. A. (1998). Can midwives reduce postpartum psychological morbidity? A randomized trial. *Birth, 25*(4), 215-219.
- Lazarus, R. (1977). Stress and coping: Some current issues and controversies. In A. Monat, & R. Lazarus (Eds.), *Stress and coping: An anthology* (pp. 1-12). New York, NY: Columbia University Press.
- Lee, C., Slade, P., & Lygo, V. (1996). The influence of psychological debriefing on emotional adaptation in women following early miscarriage: A preliminary study. *British Journal of Medical Psychology, 69*, 47-58.

- *Leonard, R. & Alison, L. (1999). Critical incident stress debriefing and its effects on coping strategies and anger in a sample of Australian police officers involved in shooting incidents. *Work & Stress, 13*(2), 144-161.
- Lewis, S. J. (2003). Do one-shot preventative interventions for PTSD work? A systematic research synthesis of psychological debriefings. *Aggression and Violent Behavior, 8*(3), 329-343.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage Publications.
- Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis & F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 83-127). New York, NY: Russell Sage Foundation.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioural treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-1209.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Malcolm, A. S., Seaton, J., Perera, A., Sheehan, D. C., & Van Hasselt, V. B. (2005). Critical incident stress debriefing and law enforcement: An evaluative review. *Brief Treatment and Crisis Intervention, 5*(3), 261-287.
- *Marchand, A., Guay, S., Boyer, R., Iucci, S., Martin, A., & St. Hilaire, M. H. (2006). A randomized controlled trial of an adapted form of individual crisis incident stress

debriefing for victims of an armed robbery. *Brief Treatment and Crisis Intervention, 6(2)*, 122-129.

- *Matthews, L. R. (1998). Effect of staff debriefing on posttraumatic stress symptoms after assaults by community housing residents. *Psychiatric Services, 49(2)*, 207-212.
- *Mayou, R. A., Ehlers, A., & Hobbs, M. (2000). Psychological debriefing for road traffic accident victims: Three-year follow-up of a randomised controlled trial. *British Journal of Psychiatry, 176(6)*, 589-593.
- McFarlane, A. C. (1988). The longitudinal course of posttraumatic morbidity: The range of outcomes and their predictors. *Journal of Nervous and Mental Disease, 176(1)*, 30-39.
- McNally, R. J. (2003). Progress and controversy in the study of posttraumatic stress disorder. *Annual Review of Psychology, 54*, 1-24.
- Miller, L. (2006). Critical incident stress debriefing for law enforcement: Practical models and special applications. *International Journal of Emergency Mental Health, 8(3)*, 189-201.
- Mitchell, J. T. (1983). When disaster strikes... The critical incident stress debriefing process. *Journal of Emergency Medical Services, 8(1)*, 36-39.
- Mitchell, J. T. (2004). *Crisis intervention and critical incident stress management: A defense of the field*. Ellicott City, MD: International Critical Incident Stress Foundation.

- Mitchell, J. T., & Everly, G. S. (1996). *Critical incident stress debriefing: An operations manual for the prevention of traumatic stress among emergency services and disaster workers*. Ellicott City, MD: Chevron.
- Mitchell, J. T., & Everly, G. S. (1997). The scientific evidence for critical incident stress management. *Journal of Emergency Medical Services*, 22(1), 86-93.
- Mitchell, J. T., & Everly, G. S. (2000). Critical incident stress management and critical incident stress debriefings: Evolution, effects and outcomes. In B. Raphael, & J. P. Wilson (Eds.), *Psychological debriefing: Theory, practice and evidence* (pp. 71-90). Cambridge, UK: Cambridge University Press.
- Mitchell, J. T., & Everly, G. S. (2001). *Critical incident stress debriefing: An operational manual for CISD, defusing, and other group crisis intervention services* (3rd ed.). Maryland: Chevron Publishing corporation.
- Mitchell, J. T., & Levenson, R. L. Jr. (2006). Some thoughts on providing effective mental health critical care for police departments after line-of-duty deaths. *International Journal of Emergency Mental Health*, 8(1), 1-4.
- Moran, C. (1998). Individual differences and debriefing effectiveness. *Australasian Journal of Disaster and Trauma Studies*, 1. Retrieved February 15, 2007, from <http://www.massey.ac.nz/~trauma/issues/1998-1/moran1.htm>
- Morris, P. L. P. (2000). Is consensus about debriefing possible? In B. Raphael, & J. P. Wilson (Eds.), *Psychological debriefing: Theory, practice and evidence* (pp. 321-326). Cambridge, UK: Cambridge University Press.

- **Nurmi, L. A. (1999). The sinking of the Estonia: The effects of critical incident stress debriefing (CISD) on rescuers. *International Journal of Emergency Mental Health, 1(1)*, 23-31.
- Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper, & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 139-155). New York, NY: Russell Sage Foundation.
- Pender, D. A. (2006). *Critical incident stress debriefing: An examination of effectiveness in group processes*. Unpublished doctoral dissertation, Southern Illinois University, Illinois.
- Public Safety Canada. (2007). *Sexual offender treatment outcome research: CODC guidelines for evaluation*. Retrieved July 23, 2007, from <http://securitepublique.gc.ca/res/cor/rep/codc-en.asp>
- Raphael, B. (1984). Psychiatric consultancy in major disaster. *Australian and New Zealand Journal of Psychiatry, 18(4)*, 303-306.
- Raphael, B. (1986). *When disaster strikes*. London, UK: Hutchinson.
- Raphael, B., Meldrum, L., & McFarlane, A. (1995). Does debriefing after psychological trauma work? *British Medical Journal, 310(7018)*, 1479-1480.
- Regehr, C. (2001). Crisis debriefing groups for emergency responders: Reviewing the evidence. *Brief Treatment and Crisis Intervention, 1(2)*, 87-100.
- *Regehr, C. & Hill, J. (2000). Evaluating the efficacy of crisis debriefing groups. *Social Work with Groups, 23(3)*, 69-79.
- Renck, B., Weisaeth, L., & Skarbo, S. (2002). Stress reactions in police officers after a disaster rescue operation. *Nordic Journal of Psychiatry, 56(1)*, 7-14.

- Rick, J. (2000, January). *Trauma management vs. stress debriefing: What should responsible organizations do?* Paper presented at the meeting of the British Psychological Society Occupational Psychology Conference, London, UK.
- Robinson, R. C. (2000). Debriefing with emergency services: Critical incident stress management. In B. Raphael, & J. P. Wilson (Eds.), *Psychological debriefing: Theory, practice and evidence* (pp. 91-107). Cambridge, UK: Cambridge University Press.
- Robinson, R., Mitchell, J. T., & Murdock, P. (1995). Do psychological debriefings work? *Phoenix*, 54, 9-11.
- Rose, S., Bisson, J., & Wessely, S. (2003). A systematic review of single-session psychological interventions (debriefing) following trauma. *Psychotherapy and Psychosomatics*, 72(4), 176-184.
- *Rose, S., Brewin, C. R., Andrews, B., & Kirk, M. (1999). A randomized controlled trial of individual psychological debriefing for victims of violent crime. *Psychological Medicine*, 29(4), 793-799.
- Rose, S., & Tehrani, N. (2002). History, methods and development of psychological debriefing. Retrieved February 13, 2007 from <http://www.bps.org.uk>
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage Publications.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research: Applied social research methods series* (vol. 6). Thousand Oaks, CA: Sage Publications.

- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231-243). New York, NY: Russell Sage Foundation.
- Rosenthal, R. (1995). Writing meta-analytical reviews. *Psychological Bulletin*, *118*(2), 183-192.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59-82.
- Schopler, J., & Galinsky, M. (1993). Support groups as open systems: A model for practice and research. *Health and Social Work*, *18*(3), 195-208.
- **Sijbrandij, M., Olf, M., Reitsma, J. B., Carlier, I. V. E., & Gersons, B. P. R. (2006). Emotional or educational debriefing after psychological trauma: Randomised control trial. *The British Journal of Psychiatry*, *189*(2), 150-155.
- Simms-Ellis, R., & Madill, M. (2001). Financial services employees' experience of peer-led and clinician-led critical incident stress debriefing following armed robberies. *International Journal of Emergency Mental Health*, *3*(4), 219-228.
- Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology*, *59*, 623-664.
- Slaikeu, K. A. (1984). *Crisis intervention: A handbook for practice and research*. Boston, MA: Allyn and Bacon, Inc.

- Snook, B., Eastwood, J., Gendreau, P., Goggin, C., & Cullen, R. M. (2007). Taking stock of criminal profiling: A narrative review and meta-analysis. *Criminal Justice and Behavior, 34*(4), 437-453.
- *Stallard, P., Velleman, R., Salter, E., Howse, I., Yule, W., & Taylor, G. (2006). A randomised controlled trial to determine the effectiveness of an early psychological intervention with children involved in road traffic accidents. *Journal of Child Psychology and Psychiatry, 47*(2), 127-134.
- Stein, D. M., & Lambert, M. J. (1984). Telephone counselling and crisis intervention: A review. *American Journal of Psychology, 12*(1), 101-126.
- Taft, C. T., & Niles, B. L. (2004). Assessment and treatment of anger in combat-related PTSD (chap. 10). Retrieved July 26, 2007, from http://www.ncptsd.va.gov/ncmain/ncdocs/manuals/iraq_clinician_guide_ch_10.pdf
- Talbot, A. (1990). The importance of parallel process in debriefing crisis counsellors. *Journal of Traumatic Stress, 3*(2), 265-278.
- Tangri, R. P. (2003). *What stress costs*. Chrysalis Performance Strategies Inc.
- Tehrani, N., & Westlake, R. (1994). Debriefing individuals affected by violence. *Counselling Psychology Quarterly, 7*(3), 251-259.
- Uziel, L. (2007). Individual differences in the social facilitation effect: A review and meta-analysis. *Journal of Research in Personality, 41*(3), 579-601.
- van Emmerik, A. A. P., Kamphuis, J. H., Hulsbosch, A. M., & Emmelkamp, P. M. G. (2002). Single session debriefing after psychological trauma: A meta-analysis. *The Lancet, 360*(9335), 766-770.

- Veterans Affairs Canada. (2004). Treatment of PTSD. Retrieved July 27, 2007, from <http://www.vac-acc.gc.ca/clients/sub.cfm?source=mhealth/factsheets/treatment>
- Walker, G. (1990). Crisis-care in critical incident debriefing. *Death Studies, 14*, 121-133.
- *Warren, K. D. (1995). *Critical incident stress debriefing with emergency service workers*. Unpublished doctoral dissertation, California School of Professional Psychology, California.
- *Wee, D., Mills, D. M., & Koehler, G. (1999). The effects of critical incident stress debriefing (CISD) on emergency medical services personnel following the Los Angeles civil disturbance. *International Journal of Emergency Mental Health, 1(1)*, 33-37.
- Wilson, D. B. (2001). Meta-analytic methods for criminology. *Annals of the American Academy of Political and Social Science, 578*, 71-89.
- Wohlmuth, E. (2003). *The experiences of correctional workers to critical incident stress debriefings*. Unpublished master's thesis, Acadia University, NS.
- Wolf, F. (1986). *Meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Woody, R. H. (2005). The police culture: Research implications for psychological services. *Professional Psychology: Research and Practice, 36(5)*, 525-529.
- *Young, A. T. (2003). *An examination of the effectiveness of periodic stress debriefing with law enforcement personnel*. Unpublished doctoral dissertation, Texas Tech University, Texas.

Appendix A
List of Researchers Contacted for Studies for Inclusion

Researchers	Affiliations
Laurence Alison, Ph.D.	University of Liverpool
Keith Armstrong, Ph.D.	University of California, San Francisco
Jonathan Bisson, Ph.D.	University Hospital of Wales
Nancy Bohl, Ph.D.	The Counselling Team International, San Bernardino
Ingrid Carlier, Ph.D.	University of Amsterdam
Grant Devilly, Ph.D.	University of Queensland
George Everly, Ph.D.	Loyola College, Maryland
Colm Humphries, Ph.D.	St. John of God Menni Services, Dublin
Justin Kenardy, Ph.D.	University of Queensland
Jeffrey Lating, Ph.D.	Loyola College, Maryland
Linda Matthews, Ph.D.	University of Sydney
Jeffrey Mitchell, Ph.D.	President of the International Critical Incident Stress Foundation
Cheryl Regehr, Ph.D.	University of Toronto
David Richards, Ph.D.	University of Manchester
Suzanne Rose, Ph.D.	University of London
Arnold van Emmerik, Ph.D.	University of Amsterdam

Appendix B
Coding Manual

CODING MANUAL

Instructions: Please answer the following questions/statements regarding each study. The information provided in the study must be explicitly stated or very clear in order for you to record it, if it is not clear or there is no indication of what the question asks then please circle (when it is given as an option) or write NOT STATED after the numbered item. If the question is not applicable just leave it blank. Please provide any information that you think is necessary, important, or applicable. If more space is needed for a particular question please use the back of the subsequent page and label it using the question number.

1. Study ID # (found in right hand corner of title page)?
2. Authors (list first name, initials, and last name of all authors)?

PUBLICATION INFORMATION:

3. Year of publication? (If two separate reports are being used to code a single study, code the publication year of the most recent):
4. Source of results/study (or publication form):
 - 0 Published peer reviewed journal article
 - 1 Published non-peer reviewed journal article
 - 2 Ph.D. dissertation
 - 3 M.A. thesis
 - 4 Technical report
 - 5 Book or chapter in a book
 - 6 Other (please specify):
5. Country of origin of the study?

6. Affiliation of authors (>75% of the authors must be of the same affiliation or code them as multi-disciplinary, if >75% of the authors are of a particular affiliation not listed please choose other and specify):

- 0 Academic (e.g., college, university, etc.) (specify institution if possible):
- 1 Research firm/consultant (specify name if possible):
- 2 Government agency (e.g., National Research Council Canada, etc.) (specify agency if possible):
- 3 Multi-disciplinary (specify all disciplines mentioned):
- 4 Other (please specify):

SAMPLE CHARACTERISTICS:

7. Final total sample size (*N*) (please take into account any attrition of any participants throughout the study):

8. Provide the sample size of the control group (i.e., those not receiving any type of intervention) and the experimental group (i.e., the group that received only PD):

Control *n*:
 Experimental group *n*:
 List any the *n* of other groups (e.g., different types of interventions):

9. Gender of sample:

- 0 All male
- 1 Mixed (please specify the *n* for each): Females *n*: Males *n*:
- 2 Not stated

10. What is the age of the participants? (If the age is provided in another form, please specify below)

	Mean (SD)	Range	Median	Mode
Total sample:				
Control:				
Experimental group:				
Any additional groups:				

11. For those who received debriefing, was it mandatory?

- 0 No
- 1 Yes (please specify the conditions) (e.g., occupational requirements):
- 2 Not stated

12. For those who did not participate in debriefing, was the option given?

- 0 No
- 1 Yes
- 2 Not stated

13. Assignment of participants to conditions (circle all that apply):

- 0 Random
- 1 Non-random
- 2 Matched (participants assigned to each group are matched on one or more participant characteristic, e.g., age, length of service, etc. – if so, specify the criteria they are matched on):
- 3 Unmatched (assignment to groups is not based on the matching of participant characteristics)
- 4 Not stated/unclear
- 5 Other (please specify):

14. Were participants selected to participate in the study based on certain inclusion/exclusion criteria (e.g., victim of armed robbery who reported intense fear, victims of sexual assault over 18 years old, etc.)?

- 0 No
- 1 Yes (please specify):
- 2 Unknown

15. Were there significant differences between participants who were placed in the experimental group versus the control group?

- 0 No
- 1 Yes (please specify):
- 2 Unknown

16. Please specify the type of population that required debriefing. NOTE: for occupational participants (e.g., police, firefighters and medical personnel) who have also been victimized during the job-task please do not also categorize them as victims.

- 0 Police officers
- 1 Firefighters
- 2 Medical personnel (please specific the type of medical employee, e.g. nurse, psychiatric care worker, etc.):
- 3 Correctional officers
- 4 Soldiers
- 5 Victims of accidents (please specific the type of accident, e.g., road traffic, burns, etc.):
- 6 Victims of crime (please specific type of crime, e.g. armed robbery, sexual assault, etc.):
- 7 Victims of a natural disaster (please specific type of disaster, e.g. hurricane, earthquake, etc.)
- 8 Victims of unspecified trauma
- 9 Other (please specify):

17. Please specify the specific type of event to which the debriefed participants were exposed. NOTE: for occupational participants (e.g., police, firefighters and medical personnel) who have also been victimized during the job-task please do not also categorize them as victims.

- 0 Critical incidents for police officers (e.g., been wounded, killed a suspect, etc.)
- 1 Critical incidents for firefighters (e.g., involving death, injury, etc.)
- 2 War exposure (e.g., bombings, recovery of bodies, etc.)
- 3 Providing emergency services to shooting victims
- 4 Victims of violence (e.g., robbery, riots)
- 5 Victims of accidents (e.g., burns, traffic)
- 6 Victims of a natural disaster (e.g., earthquake, hurricane)
- 7 Exposure to multiple incidents/diverse types of populations experiencing diverse types of trauma
- 8 Unspecified experiences by occupational participants (e.g., police, correctional officers, psychiatric care workers, etc.)
- 9 Multiple types of events
- 10 Unspecified trauma
- 11 Other (please specify):

18. Please specify the individual(s) directly exposed to the critical incident.

- 0 A single individual involved (e.g., burn victim, road traffic accident)
- 1 A group of individuals involved (e.g., bombings, natural disaster)
- 2 Not stated

19. Please specify the form of debriefing received by the experimental group. NOTE: if number of stages is not given for Mitchell's model, assume 7-stages.

- 0 7-stage Mitchell's CISD model (group)
- 1 7-stage Mitchell's CISD model (individual)
- 2 6-stage Mitchell's CISD model (group)
- 3 Adapted from Mitchell's CISD model (group)
- 4 Adapted from Mitchell's CISD model (individual)
- 5 Dyregrov's model
- 6 Combination of Dyregrov & Mitchell's model
- 7 Educational and emotional debriefing
- 8 Connecticut's Critical Incident Stress Response Team approach
- 9 Intervention to promote emotional and cognitive functioning
- 10 Counselling session
- 11 Other (please specify):
- 12 Unknown

20. If indicated, how many participants were present in the debriefing group? (please also indicate if it was an individual session):

21. If indicated, how many debriefers (e.g., one mental health professional, peer, crisis team) were present in the debriefing session?

22. Please specify who gave the debriefing to the experimental group:

- 0 Trained peers (e.g., police, firefighters, etc.) (please specify the type of person):
- 1 Mental health professional (e.g., psychiatrist, clinical psychologist, etc.) (if stated, please specify the exact person):
- 2 Other type of health professional (e.g., social worker, nurse) (if stated, please specific the exact person):
- 3 Trained peers & some type of health professional
- 4 Principle researcher
- 5 Trained peers & principle researcher
- 6 Counsellor
- 7 Unspecified but trained/experienced debriefers
- 8 Other (please specify):
- 9 Unknown

23. If indicated, please specify at what time post-trauma the experimental group received the first debriefing session (e.g., 24-72 hours, 1 week, 6 months, etc.):

24. Please indicate how many debriefing sessions were provided for the experimental group:

25. If indicated, please specify the length of time of the debriefing session. If the experimental group had multiple sessions, please report the number of sessions and length of time for each individual session (e.g., 90 minutes, 30-120 minutes, two 1-hour sessions, etc.):

OUTCOME VARIABLES:

26. Were baseline measures taken of the experimental and control groups before the debriefing session?

- 0 No
- 1 Yes
- 2 Unknown

27. If yes, was there a difference between the experimental and control groups on any of these measures? (Please specify the difference between groups):

28. If clinical assessments were used as an outcome measure, was the clinician blind as to which group had received debriefing?

- 0 No
- 1 Yes
- 2 Unknown

29. Was the outcome variable:

- 0 Subjective (e.g., perceived helpfulness of PD by surveys, clinical assessments)
- 1 Objective (e.g., quantitative measures of symptoms from psychometric tests)
- 2 Both

30. Please list all of the specific types of outcome measure used to compare the experimental and control groups (e.g., psychological inventories, surveys, clinical assessments, etc.):

31. How many of the outcome measures were self-report inventories?

- 0 None
- 1 All
- 2 More than half
- 3 Less than half
- 4 Unknown

32. If appropriate, please indicate at what point a pre-debriefing assessment was made (e.g., 24 hours post-trauma):

33. Please indicate how many post-debriefing assessments were taken:

34. Please indicate at what point each of the post-debriefing assessments were taken (e.g., one week post-trauma, etc.):

ANALYSIS:

35. Please provide the level of statistical analysis used in the study:

- 0 Descriptive statistics only (M, SD, correlations, etc.)
- 1 Advanced statistics (t-tests, ANOVA, regression, etc.)

36. Were comparisons made between experimental and control groups conducted to check for equivalence on characteristics (e.g., age, gender, etc.)?

- 0 No
- 1 Yes (please specify):

37. Was the reliability, validity, or sensitivity of outcome measures given?

- 0 No
- 1 Yes (please specify):

38. Please provide the information that could be used to calculate the effect sizes (e.g., M , SD , n , t , F , p , z , X^2 , etc.). Also, make sure to indicate whether the difference/association was significant.

QUALITY OF STUDY:

39. Please assign a number to determine the quality of this study according to the number of quality variables included. These variables include: providing baseline measures pre-debriefing, having randomized experimental and control groups, having matched experimental and control groups, the use of objective (rather than subjective) outcome measures or having a total sample size over 100 (e.g., if a study has baseline measures and a sample size over 100, but no other quality variables it would receive a rating of 2)

- 0 None of these qualities
- 1 One of these qualities
- 2 Two of these qualities
- 3 Three of these qualities
- 4 Four of these qualities
- 5 Five of these qualities

Appendix C
Comprehensive List of Outcome Measures and Corresponding Aggregated Categories

Aggregate Category	Psychological Test (outcome measures)	Description
PTSD Symptoms	Children's Impact of Events Scale (CIES)	A modified version of the IES (3 scales: intrusion, arousal, avoidance) to quantify PTSD in children.
	Clinician-Administered PTSD Scale (CAPS-II)	A clinician-administered interview to assess PTSD symptoms.
	Everstine Trauma Response Index (ETRI)	A 6-item self report scale used to measure arousal level post-trauma.
	Frederick Reaction Index-Adult (FRI-A)	A 28-item self-report inventory to examine the presence and severity of PTSD symptoms on a 5-point scale.
	Impact of Events Scale (IES)	A 15-item self-report scale to measure intrusive re-experiencing of trauma and avoidance of trauma-related stimuli to assess PTSD.
	Impact of Events Scale-Revised (IES-R)	A 22-item self-report scale to measure intrusive re-experiencing of trauma and avoidance of trauma-related stimuli to assess PTSD and ASD.
	Peritraumatic Dissociation Experiences Questionnaire (PDEQ)	An 8-item self-report questionnaire to assess the presence or absence of 8 dissociative symptoms.
	Post-traumatic Stress Disorder Questionnaire	A 5-point self-report scale of 25 questions measuring PTSD symptoms based on the DSM-III.
	Post-traumatic Stress Disorder Symptoms Scale (PSS)	A 17-item self-report scale to assess the 17 PTSD symptoms according to the DSM-III-R rated on a 4-point scale.
Anxiety	Self-Rating Scale for PTSD (SRS-PTSD)	A self-report abridged version of the SI-PTSD with 17-items corresponding to the 17 PTSD symptoms.
	Structure Interview for PTSD (SI-PTSD)	Assesses the 17 PTSD symptoms in terms of their frequency and intensity from a current and lifetime perspective.
Anxiety	Brief Symptom Inventory (BSI) (anxiety scale)	A 53-item self-report shortened version of the SCL with an anxiety scale to detect psychological distress associated with anxiety.
	Hospital Anxiety Scale (HADS-A)	A self-report questionnaire used to establish the presence and severity of anxiety for in and out-patients.

	State-Trait Anxiety Inventory (STAI)	A self-report inventory with 20 items to assess state anxiety and 20 items to assess trait anxiety (frequency and intensity scales included)
	Symptom Checklist Revised (SCL-R-90) - Anxiety rating	A subscale of the SCL devoted to assessing anxiety.
	Revised Children's Manifest Anxiety Scale (RCMAS)	A 37-item self-report inventory to measure anxiety in children.
Depression	Beck Depression Inventory (BDI)	A 21-item self-report inventory measuring characteristic attitudes and symptoms of depression.
	Birleson Depression Inventory (BDI)	An 18-item self-report inventory to diagnose depressed children.
	Brief Symptom Inventory (BSI) (depression scale)	A 53-item self-report shortened version of the SCL with a depression scale to detect psychological distress associated with depression.
	Epidemiological Studies Depression Scale (CES-D)	A 20-item self-report questionnaire to assess depression symptoms within the past week.
	Hospital Depression Scale (HADS-D)	A self-report questionnaire used to establish the presence and severity of depression for in and out-patients.
	Symptom Checklist Revised (SCL-R-90) revised Depression rating	A subscale of the SCL devoted to assessing depression.
Anger	Novaco Provocation Inventory	An 80-item self-report test to explore the range of situations that evoke anger and the intensity of the anger experienced
	State-Trait Anger Expression Inventory (STAXI)	A 44-item self-report scale to measure the symptomatic outcome of the experience and expression of anger
General Health & Functioning	COPE Scale (adaptive and maladaptive)	A 13-item self report instrument used to determine one's coping ability
	CAGE questionnaire	A 4-item self-report questionnaire to detect a loss of control of alcohol use.
	General Health Questionnaire (GHQ-12 or 28)	A 12 or 28-item self-report questionnaire used to detect a wide-range of psychological disorders.
	Social Readjustment Rating Scale (SRRS)	A 42-item self-report scale to assess one's life stressors within the past year.
	Strengths and Difficulties Questionnaire (SDQ)	A self-report behavioural screening questionnaire about 25 positive and negative attributes.

Appendix D
Unweighted Mean d , Hedges' \hat{g} , Significance Levels and U_3 for Unaggregated Outcome Measures

Contrast (k)	N	$d(SD)$	CI_d	$\hat{g}(SD)$	CI_g	U_3
Debriefing vs. no debriefing (24)	2235	.11(.05)	.02 to .19	.11(.04)	.02 to .19	.54
<i>PTSD</i>						
Impact of Events Scale (9)	776	-.11(.07)	-.24 to .02	-.11(.07)	-.24 to .02	.54
Impact of Events Scale – Revised (4)	226	.78(.14)	.50 to 1.06	.77(.14)	.49 to 1.04	.77
Impact of Events Scale – Children (3)	158	.05(.17)	-.30 to .39	.05(.17)	-.30 to .39	.52
SCL-90-R (3)	189	.94(.16)	.64 to 1.24	.93(.15)	.63 to 1.23	.82
CAPS (3)	248	.00(.13)	-.26 to .25	.00(.13)	-.26 to .25	.50
Self-Rating Scale for PTSD - avoidance (1)	243	-.15(.16)	-.45 to .15	-.15(.15)	-.45 to .15	.56
Self-Rating Scale for PTSD - hyperarousal (1)	243	-.16(.16)	-.47 to .14	-.16(.15)	-.47 to .14	.56
Self-Rating Scale for PTSD - re-experiencing (1)	243	-.29(.16)	-.60 to .01	-.29(.15)	-.60 to .01	.62
Engelmann's CISTD survey – CES-D & IES (1)	111	-.54(.23)	-.98 to -.09	-.53(.23)	-.98 to -.09	.69
Avoidance scale (Matthews) (1)	63	.45(.33)	-.19 to 1.09	.44(.32)	-.18 to 1.07	.67
Intrusive thoughts scale (Matthews) (1)	63	.14(.32)	-.50 to .77	.13(.32)	-.49 to .75	.56
Hyperarousal scale (Matthews) (1)	63	.24(.32)	-.39 to .87	.24(.32)	-.39 to .86	.60
Structured Interview – PTSD (2)	317	-.11(.13)	-.35 to .14	-.10(.13)	-.35 to .14	.54
Frederick reaction index – adult (1)	65	.47(.26)	-.04 to .99	.47(.26)	-.04 to .98	.67
Penn Inventory (3)	189	1.04(.16)	.73 to 1.34	1.02(.154)	.722 to 1.33	.84
PTSD Symptoms Scale (1)	157	-.06(.12)	-.30 to .17	-.06(.12)	-.30 to .17	.52
Level of distress questionnaire (1)	80	.32(.43)	-.53 to 1.16	.31(.41)	-.51 to 1.12	.62
<i>Anxiety</i>						
Hospital Anxiety and Depression Scale – Anxiety subscale (4)	533	-.10(.09)	-.28 to .08	-.10(.09)	-.28 to .08	.54
State-Trait Anxiety Inventory (2)	136	1.00(.19)	.62 to 1.38	.99(.19)	.61 to 1.37	.84
Brief Symptom Inventory – Anxiety (1)	61	-.04(.21)	-.45 to .38	-.04(.21)	-.45 to .37	.50

Manifest Anxiety Scale-R (children) (1)	158	.08(.17)	-.27 to .42	.08(.17)	-.26 to .42	.54
<i>Depression</i>						
Hospital Anxiety and Depression Scale – Depression subscale (4)	533	-.05(.09)	-.23 to .13	-.05(.09)	-.23 to .13	.50
Birleson Depression Inventory (1)	158	.16(.18)	-.18 to .51	.16(.17)	-.18 to .50	.56
Brief Symptom Inventory – Depression (1)	61	.06(.21)	-.35 to .48	.06(.21)	-.35 to .47	.50
Beck Depression Inventory (5)	457	.35(.09)	.18 to .53	.35(.09)	.18 to .52	.64
<i>Anger</i>						
Novaco Anger Inventory (2)	136	.83(.18)	.47 to 1.18	.82(.18)	.47 to 1.17	.79
STAXI – anger control (1)	60	.41(.26)	-.10 to .92	.36(.26)	-.15 to .86	.64
STAXI – anger expression (1)	60	.36(.26)	-.15 to .87	.40(.26)	-.10 to .91	.66
STAXI – anger in (1)	60	.21(.26)	-.30 to .71	.20(.26)	-.30 to .70	.58
STAXI – anger out (1)	60	.17(.30)	-.34 to .67	.16(.26)	-.34 to .67	.52
STAXI – angry temperament (1)	60	.67(.27)	.15 to 1.19	.66(.26)	.15 to 1.17	.74
STAXI – angry reaction (1)	60	.48(.26)	-.03 to .99	.48(.26)	-.03 to .99	.69
STAXI – state anger (1)	60	.60(.26)	.08 to 1.11	.59(.26)	.08 to 1.10	.72
STAXI – trait anger (1)	60	.59(.26)	.07 to 1.11	.58(.26)	.07 to 1.09	.71
Brief Symptom Inventory – Hostility (1)	61	-.30(.21)	-.71 to .12	-.30(.21)	-.71 to -1.41	.61
<i>General Health & Functioning</i>						
CAGE (1)	106	.00(.19)	-.38 to .38	.00(.19)	-.38 to .38	.50
Adaptive coping (1)	60	.48(.26)	-.04 to .99	.47(.26)	-.03 to .98	.67
Maladaptive coping (1)	60	.11(.26)	-.40 to .62	.11(.26)	-.39 to .61	.54
Bohl's Stress Survey (2)	136	1.50(.20)	1.12 to 1.89	1.49(.19)	1.11 to 1.87	.93
General Health Questionnaire (1)	60	-.04(.27)	-.58 to .50	-.04(.27)	-.57 to .49	.50
General Symptom Inventory (1)	61	-.39(.21)	-.81 to .03	-.39(.21)	-.80 to .03	.65
Strengths & Difficulties Questionnaire (1)	158	.28 (.18)	-.06 to 0.63	.28(.17)	-.06 to 0.62	.60

Appendix E
Hedges' \hat{g} , Q Statistics, and Significance Levels for Standard Moderators of all Outcome Measures

Hedges' \hat{g} , Q Statistics, and Significance Levels for Publication Source of all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Combined outcome measures							
Publication (peer-reviewed)	20	.07(.05)	-.03 to .16	6.38(1)	.012	114.34(22)	.000
Non-peer reviewed publication	4	.42(.13)	.16 to .68				
PTSD symptoms							
Publication (peer-reviewed)	19	.03(.05)	-.06 to .12	7.40(1)	.007	73.98(19)	.000
Non-peer reviewed publication	2	-.49(.19)	-.86 to -.13				
Anxiety							
Publication (peer-reviewed)	6	-.06(.08)	-.21 to .09	25.93(1)	.000	32.24(6)	.000
Non-peer reviewed publication	2	.99(.19)	.62 to 1.4				
Depression							
Publication (peer-reviewed)	8	.02(.06)	-.10 to .14	42.60(1)	.000	20.49(9)	.015
Non-peer reviewed publication	3	1.17(.17)	.85 to 1.50				
Anger							
Publication (peer-reviewed)	2	.00(.16)	-.33 to .31	11.58(1)	.001	6.60(2)	.037
Non-peer reviewed publication	2	.82(.18)	.47 to 1.17				
General health & functioning							
Publication (peer-reviewed)	5	.04(.10)	-.15 to .23	45.25(1)	.000	7.52(5)	.185
Non-peer reviewed publication	2	1.49(.19)	1.11 to 1.86				

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Affiliation of Authors Moderator of all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
Academic	11	.20(.07)	.06 to .33	45.52(3)	.000	75.20(20)	.000
Private	3	-.04(.13)	-.29 to .21				
Multi-disciplinary	8	-.09(.07)	-.23 to .04				
Other	2	1.22(.19)	.84 to 1.59				
PTSD symptoms							
Academic	11	.17(.07)	.03 to .30	11.64(2)	.003	69.75(18)	.000
Private	3	-.10(.13)	-.34 to .15				
Multi-disciplinary	7	-.16(.07)	-.30 to -.02				
Anxiety							
Private	2	.03(.14)	-.25 to .31	26.47(2)	.000	31.70(5)	.000
Multi-disciplinary	4	.09(.01)	-.27 to .08				
Other	2	.99(.19)	.62 to 1.37				
Depression							
Academic	2	.19(.11)	-.04 to .41	51.04(3)	.000	12.05(7)	.099
Private	2	.22(.14)	-.06 to .50				
Multi-disciplinary	5	-.01(.08)	-.26 to .06				
Other	2	1.38 (.19)	1.01 to 1.76				
General health & functioning							
Private	3	-.15(.13)	-.39 to .10	50.31(2)	.000	2.45(4)	.653
Multi-disciplinary	2	.28(.14)	.00 to .57				
Other	2	1.49(.19)	1.11 to 1.86				

Hedges' \hat{g} , Q Statistics, and Significance Levels for Assignment to Conditions Moderator of all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
Random	9	-.07(.06)	-.18 to .05	8.128(1)	.000	74.779(18)	.004
Non-random	11	.20(.07)	.06 to .34				
PTSD symptoms							
Random	9	-.10(.06)	-.22 to .02	4.13(1)	.042	70.21(17)	.000
Non-random	10	.10(.08)	-.05 to .25				
Depression							
Random	7	.04(.06)	-.08 to .17	2.70(1)	.100	23.50(8)	.003
Non-random	3	.30(.14)	.02 to .57				
General health & functioning							
Random	3	.01(.11)	-.21 to .22	12.16(1)	.000	24.49(3)	.000
Non-random	2	.78(.19)	.40 to 1.15				

Hedges' \hat{g} , Q Statistics, and Significance Levels for the Quality of Study Moderator of all Outcome Measures

Moderating Variables	k	$\hat{g}(SD)$	$CI_{\hat{g}}$	$Q_b (df_b)$	p	$Q_w (df_w)$	p
Overall outcome measure							
0	3	.37(.18)	.02 to .73	39.29(4)	.000	81.43(19)	.000
1	3	.78(.16)	.47 to 1.08				
2	7	.36(.10)	.15 to .52				
3	7	-.12(.07)	-.26 to .01				
4	4	.00(.09)	-.18 to .17				
PTSD symptoms							
0	3	.37(.18)	.02 to .73	19.31(3)	.000	61.68(16)	.000
2	6	.31(.10)	.11 to .51				
3	7	-.17(.07)	-.31 to -.04				
4	4	-.03(.09)	-.21 to .14				
Anxiety							
1	2	.99(.19)	.62 to 1.37	26.35(2)	.000	31.81(5)	.000
3	2	-.14(.14)	-.42 to .14				
4	4	-.03(.09)	-.20 to .15				
Depression							
1	2	1.38(.19)	1.01 to 1.76	42.80(2)	.000	16.37(7)	.022
3	4	.08(.09)	-.10 to .25				
4	4	.05(.09)	-.13 to .22				
General health & functioning							
1	3	.97(.16)	.66 to 1.28	16.07(1)	.000	22.72(3)	.000
4	2	.16(.13)	-.10 to .41				