Received: 31 May 2018 | Revised: 23 January 2019 | Accepted: 28 January 2019

DOI: 10.1002/acp.3530

SHORT PAPER

WILEY

A survey on adverse incidents in legal psychology studies: Reflections on ethics review

Melanie Sauerland | Anna Sagana | Harald Merckelbach

Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

Correspondence
Melanie Sauerland, Maastricht University, Section Forensic Psychology, Faculty of Psychology and Neuroscience, 616, 6200 MD Maastricht, The Netherlands.
Email: melanie.sauerland@maastrichtuniversity.nl

Summary

Ethics committees (ECs) regulate research activities to maintain research participants' autonomy and to protect them from harm and injury. No research to date attempted to establish how much risk is involved in social-science research. Using a survey approach, we set out to estimate the risk of being involved in an incident for research participants in legal psychology and assessed researchers' views of ECs. Fifty-nine of 188 respondents (31%) stated that they had experienced one or more incidents with a participant. The estimated risk of being involved in an incident was one to three per 10,000 participants, which according to biomedical standards defines a *rare* risk. Although some researchers were satisfied with their EC, the general tenor was one of discontent due to conservative decision-making, lacking expertise, and overstepping demands. Whether ECs succeed in protecting participants from loss of autonomy, harm and injury are unknown but are open to empirical research.

KEYWORDS

code of ethics, harm, randomized controlled trial, risk of participation, risk probability

1 | INTRODUCTION

The ethics of research with human subjects have fueled debates in the past half century. The introduction of ethics committees (ECs[1]) in research institutions is related to violations of patients' and participants' rights. Prominent examples of such violations in the biomedical field include studies in which treatment was denied to syphilis victims (Tuskegee, 1932–1973), and patients were injected with hepatitis (Willowbrook, 1956–1972) or cancer cells (Jewish Chronic Disease Hospital in the 1960s; Lemonick & Goldstein, 2002). In social science, instructing participants to punish another participant with potentially fatal electric shocks (Milgram, 1963) and failing to gain informed consent from participants in a study of impersonal public sexual behavior (thus undermining participants' autonomy; Humphreys, 1970) are notable examples that stimulated discussion on research ethics (e.g., Baumrind, 1964; Kaufmann, 1967; von Hoffman, 1970).

Since the establishment of ethics review (ER) in research institutions, researchers have questioned its legitimacy and its beneficial value (e.g., Burnham, 1966). Opponents object to the application of ER principles from the biomedical science to social science. More specifically, they challenge the idea that comparable risk is involved in social-science research, deeming ER unnecessary (Boden, Epstein, & Latimer, 2009; Dingwall, 2006; Hamilton, 2005; Schrag, 2011).

[1]Also termed ethics review boards, institutional review boards, review committees, or research ethics committees. Our article pertains to all of them. When referring to the reviewing activity, we use the term ethics review (ER). When referring to the institution conducting the review, we use the term ethics committee (EC).

Another point of critique is that costs of ER are often unattended or underestimated (Gray, 1977; Hammersley, 2006). Next to time and financial resources (e.g., Gunsalus, 2006), these costs include an impact on researchers' choices, which may lead, for example, to an increase in the collection of self-report rather than behavioral data (Baumeister, Vohs, & Funder, 2007), obstructive influence on some forms of research (Dingwall, 2008; Hamilton, 2005), hence challenging the degrees of freedom for academic research (Haggerty, 2004).

Proponents of ER hold that differences in the risks between social-science and biomedical research are often overstated (Jennings, 2012) and that potential harms for participants are not limited to the biomedical field (Hunter, 2014). They conclude that potential conflicts of interest and abuse justify ER (Edwards, 2009) and that the discussion should revolve around questions of practice (*how*) rather than principle (*if*; Nicholls, Brehaut, & Saginur, 2012).

Given this controversy, it is surprising that research so far has not attempted to assess to what extent ECs meet their main objectives (Grady, 2010; Resnik, 2015; Taylor & Ervin, 2017). These include the protection of human research participants from potential harm, threats to their autonomy (e.g., studying research participants without their consent and use of deception), and injury (e.g., Edwards, 2010; All European Academies, 2017; U.S. Department of Health & Human Services, 2018 [Common rule]). Rather, EC-related research has focused on ECs' mode of operation, showing that the review process can be inconsistent, inefficient, and cause delays (Abbott & Grady, 2011; Coleman & Bouësseau, 2008; Nicholls et al., 2015). It is similarly unclear how much protection participants in social science research actually need. This is because we know little about the potential risks of injury, harm, or loss of autonomy through participation in social-science research (Colnerud, 2015; Nicholls et al., 2015). At the same time, regulation exercised by ECs seems to be comparatively strong; some have argued even stronger than that of the criminal and civil law (Hammersley, 2009): Rather than setting principles to which researchers are held in cases of possible violations post facto, ECs operate *prospectively*.

In all, there seems to be a mismatch between the lack of evidence-based knowledge about (a) the risk of harm, injury, or loss of autonomy individuals face when participating in research in the social science and (b) in how far ERs warrant protection from such effects, on the one hand, and the rigor of research regulation by ECs, on the other hand. In an attempt to provide a first approximation of risks involved in social-science research, particularly the domain of legal psychology, we conducted a retrospective survey in which researchers indicated the frequency and nature of incidents in the field. To be able to add data on legal psychology researchers' experience with and opinion on ECs, respondents also answered a number of EC-related questions. We focused on one particular discipline, namely legal psychology. Legal psychology is a good starting point for building up an empirical knowledge base about risks of research participation, because studies potentially touch upon ethically sensitive issues (e.g., violence, offending, sexual abuse), experimental settings may involve deception (and hence threat to participants' autonomy), and participants might include vulnerable individuals (e.g., crime victims or offenders, including children and adolescents; González-Sala, Osca-Lluch, Gil, & Ortega, 2017). As such, legal psychology might be one of the high-risk disciplines within the social science.

2 | METHOD

2.1 | Survey respondents

Respondents were 188 researchers (61 male, 73 female, 54 nonidentified), aged between 24 and 78 years old (M = 41.3 years, SD = 13.2, n = 128) who had been active as researchers between 0.5 and 45 years (M = 12.0 years, SD = 9.8). No minimum or maximum sample size was determined beforehand. Researchers were invited to fill out the survey while attending a conference or via email. A broad definition of "researcher in legal psychology" was employed: Anyone visiting a psychology- and law-related conference or publishing in a Psychology and Law-related journal was considered a legal psychologist. In the beginning of the survey, however, we did indicate that we were "interested in the prevalence of stress-related incidents with participants taking part in lab or field experiments in our research domain (psychology & law)." This instruction was intended to attract researchers in the field of legal psychology and to deter psychologists who were primarily active on other research domains.

Emails were sent to members of the authors' networks and to first authors of articles published between 2014 and June 2016 in the following journals: *Applied Cognitive Psychology*; *Journal of Applied Research in Memory and Cognition*; *Law and Human Behavior*; *Legal and Criminological Psychology*; *Psychology, Crime & Law*. These journals were selected because they constitute the current and previous publishing organ of the *Society for Applied Research in Memory & Cognition*, the publishing journals of the *American Psychology-Law Society* (APA Division 41), the Psychology and Law journal of the *British Psychological Society*, and the publishing organ of the *European Association of Psychology and Law*, respectively. We reached out to researchers who recently published in Psychology and Law-related journals rather than members of the related associations because we wanted to ensure that respondents had recent research experience (i.e., had conducted and published at least one first author article in a relevant journal). To minimize the risk of double data inclusion, only first authors were approached. The total number of individuals approached via email was 660. We sent one reminder.

Most researchers (45%) had previously conducted between 11 and 50 studies. About one third (32%) had conducted between 1 and 10 studies, 14% between 51 and 100 studies, 6% between 101 and 150 studies, and 2% more than 150 studies thus far in their careers. They answered the survey online in 2016 (n = 150, corresponding to a response rate of 25% for individuals approached via email), by paper and pencil during the 2013 *Society of Applied Research in Memory and Cognition* conference (Rotterdam, n = 19), the *International investigative interviewing Research Group* conference (Maastricht, n = 15), or through email (n = 4). Four online-survey

TABLE 1 Survey overview and responses

Item	Results[a]		N	Survey version
1. For how many years have you been active as a researcher in the Psychology & Law domain? _____years	*M* = 12.4 *SD* = 10.1 Range 0.5–45		186	Both
2. How many experiments did you carry out in those years?			188	Both
1–10		32%		
11–50		45%		
51–100		14%		
101–150		6%		
____		2%		
3. When running an experiment, have you ever experienced an incident with a participant? (e.g., participant displayed serious levels of distress, started crying, later filed a complaint about the study, experienced long-term negative effects resulting from the study).			188	Both
No (continue with Question 16)		69%		
Yes (continue with Question 4)		31%		
4. If yes, how often *Open-answer format*	*M* = 4.1 *SD* = 6.9 Range 1–30		48	Both
Across how many studies? *Open-answer format*	*M* = 1.8 *SD* = 1.7 Range 1–10		48	Both (Yes)
5. Please describe each incident and the assumed cause(s) *Open-answer format*	See results section		48	Both (Yes)
6. In which year(s) did the study/studies take place? *Open-answer format*	1977–1985 1990–1999 2000–2009 2010–2016 *(% incidents)*	3% 7% 23% 67%	47	Both (Yes)
7. In what domain of study did the incident(s) occur (e.g., eyewitnesses, interrogations, lie detection etc.)? *Open-answer format, multiple answers possible*	Eyewitnesses Interrogation and confessions Lie detection and credibility Memory, misinformation, false-memory paradigm Psychopathology and mental health Stress, trauma, PTSD Juries and decision-making Victimology Psychopharmacology Other *(% incidents)*	27% 13% 11% 11% 7% 6% 6% 4% 4% 11%	48	Both (Yes)
8. Did it ever happen to you that participants reacted upset or angry as a reaction to certain elements in the study? (e.g., because they found questions insulting, manipulations unacceptable?)			34	Online (Yes)
Yes (if so, how many times) _____		29%		
No		71%		
9. Did it ever happen to you that respondents ignored your instructions?			33	Online (Yes)
Yes (if so, how many times) _____		55%		
No		46%		
10. Why do you think they ignored the instructions? Because they found (multiple answers possible)			18	Online (Yes)
Manipulation too stressful		6%		
Manipulation too boring		22%		
Unable to follow instructions		44%		
Other		78%		
11. Did you ever have to exclude participants, because they had mental health issues?			33	Online (Yes)
Yes (if so, how many times) _____		30%		
No		70%		
12. What kind of mental issues? *Open-answer format, multiple answers possible*	Substance abuse Medication Psychosis Mood disorder/distress Recent psychological/psychiatric issues Other	40% 20% 30% 20% 10% 20%	10	Online (Yes)

(Continues)

TABLE 1 (Continued)

Item	Results[a]		N	Survey version
Paper-and-pencil version:			14	Both (YesD)
13. Do you think that the incident was causally related to the study/ experiment per se?				
No		43%		
Yes		50%		
Do not know		7%		
Online version:				
13. Do you think that it were elements of the study itself that caused the incident(s) you experienced? (alternatively, incidents may have been related to the personal situation of the participant)			34	
No		71%		
Yes		29%		
Do not know		0%		
14. Did the study/studies have to pass a research ethics committee prior to running?			46	Both (Yes)
No		9%		
Yes		91%		
15. In which country was the study/were the studies run? *Open-answer format*	United States	31%	46	Both (Yes)
	United Kingdom	13%		
	Canada	13%		
	Germany	10%		
	Australia	10%		
	Scandinavia	8%		
	Netherlands	6%		
	Belgium	4%		
	Croatia	2%		
	Korea	2%		
Paper-and-pencil version:			36	BothD
16. Are research ethics committees in your opinion too liberal or lenient when evaluating Psychology & Law studies?				
No		92%		
Yes		0%		
Do not know		8%		
17. Are research ethics committees in your opinion too strict or conservative when evaluating Psychology & Law studies?			34	
No		47%		
Yes		41%		
Do not know		12%		
Online version:				
16. Are research ethics committees in your opinion too lenient or strict when evaluating Psychology & Law studies?			49	
Too lenient		2%		
Too strict		98%		
Neither		0%		
Do not know		0%		
17. Are research ethics committees in your opinion too liberal or conservative when evaluating Psychology & Law studies?			55	
Too liberal		0%		
Too conservative		100%		
Neither		0%		
Do not know		0%		
18. Did you ever refrain from carrying out an experiment because you anticipated too much critique from a research ethics committee?			162	Both
No		58%		
Yes		42%		
Do not know		0%		
19. Do you think that, in general, (stress-related) incidents in experiments prove the effectiveness of an experimental manipulation or a research approach?			128	BothD
No		56%		
Yes		34%		
Do not know		10%		

(Continues)

TABLE 1 (Continued)

Item	Results[a]	N	Survey version
20. Do research ethics committees, in your opinion, have sufficient expertise to evaluate the risks and benefits of Psychology & Law studies?		133	Both
No	56%		
Yes	41%		
Do not know	3%		
21. Is there anything on this topic that you would like to share with us?	See results section		Both
No			
Yes, namely,		47	

Note. Both: item presented to *all* respondents; Both (Yes): items presented to respondents who answered *yes* concerning incident (Item 3); Online (Yes): items presented to *online survey* respondents who answered *yes* concerning incident (Item 3); Both (YesD): items presented to respondents who answered *yes* concerning incident (Item 3) *but different versions* for paper-and-pencil versus online version; BothD: items presented to *all* respondents *but different versions* for paper-and-pencil versus online version.

[a]Percentages refer to the proportion of respondents who answered the referring question (i.e., the number of respondents for each question are mentioned in the second last column).

respondents answered the question whether they had filled out a similar survey in 2013 with "yes." Their paper-and-pencil forms could be matched with the online responses, and the paper-and-pencil responses were then dropped, so that every researcher would appear in the database only once. No compensation was provided. Ethical approval was granted by the standing local ethical committee post hoc.

2.2 | Materials

2.2.1 | Survey

The paper-and-pencil version of the survey contained 16 items. On the basis of the feedback received from respondents, we made some minor adjustments in the online version and five items were added. Table 1 shows a complete overview of the survey items.

Research experience, occurrence, and nature of incidents

Following two items about their research experience (years active in research and number of studies run), respondents were asked whether they had *ever experienced an incident with a participant* (e.g., *participant displayed serious levels of distress*, *started crying*, *later filed a complaint about the study*, *experienced long-term negative effects resulting from the study)*. In case of a *yes* answer on the incident item, seven items concerning the nature of incidents followed.

General opinion on ECs and incidents

Three items concerned researchers' opinion on ECs. One item addressed researchers' opinion on incidents in general, asking whether they thought that (stressrelated) incidents in experiments proved the effectiveness of an experimental manipulation or a research approach (i.e., in a study on interrogations and false confessions, does not a crying participant provide evidence to the idea that the interrogation was realistic?). In the final open question, respondents were given the opportunity to write down anything else they would like to share.

2.3 | Coding

2.3.1 | Incident descriptions (Item 5)

Four *incident* characteristics were coded based on researchers' free incident descriptions. These included observable *behavior* that was judged an incident (e.g., crying, anger, distress), the assumed incident *trigger* (e.g., participant's personal background, recalling autobiographic memories, deception), *type* of incident (e.g., psychological, physiological), and assumed *cause* (e.g., experiment, personal, combined). Combined causes referred to cases where a combination of both personal and experimental factors contributed to the incident (e.g., testing rape victims or young offenders; mock crime when participant recently witnessed a robbery). A complete list of the coding categories can be found in Tables 2 and 3.

2.3.2 | Final remarks (Item 21)

From researchers' final open *remarks*, we extracted *comments* about ECs in general (e.g., ECs are too strict, ECs lack knowledge, ECs are competent, ECs are important) and *recommendations and demands* for EC functioning and procedures (e.g., ECs should take a calculated risk: experts should be members of ECs, ECs should be evidence-based, ECs need to be flexible). A full list of the coding categories can be found in Table 4.

2.3.3 | Interrater reliability

The incident characteristics and final remarks were independently coded by the first and second author. Disagreement in the coding was resolved in discussion. Interrater reliability was substantial at $\kappa = .79$ (Landis & Koch, 1977).

3 | RESULTS

A complete overview of researchers' responses and the number of respondents for each item can be found in Table 1. In the following, the focus is on incident-related findings.

TABLE 2 Incident characteristics and triggers

Behavior judged an incident	Frequency	Assumed trigger	Frequency
Incidents related to research participants' well-being or discomfort			
Crying and emotional reaction	20	Personal background	12
Stress/distress	13	Recall autobiographic memories	9
Anger	12	Emotional stimuli	7
Dizziness and fainting	6	Unclear/other	6
Termination participation	5	Questions	5
Unclear/other	4	Performance pressure	5
Fear	3	Aversive stimuli	4
Nausea and indisposition	3	Cheating paradigm	4
File complaint	2	Experimental setting	4
		Mock crime	3
		Stress induction	3
		Deception, false memory paradigm, lying instructions	2
		Separation from parents	2
		Medical background	2
		Physical injection	2
		Had not eaten	2
Incidents related to research participants' autonomy			
Anger	4	Deception, false memory paradigm, lying instructions	6
Stress/distress	1	Cheating paradigm	1
Termination participation	1	Mock crime	1
File complaint	1		
Sum	75		80

Note. Multiple categories may apply for one incident.

TABLE 3 Incident types and assumed causes

Type of incident	Frequency	Assumed cause	Frequency
Incidents related to research participants' well-being or discomfort			
Psychological, for example, crying, anger, distress, and fear	45	Combined, for example, feeling nauseous in stress study and participant indicates she did not eat properly that day	34
Physiological, for example, dizziness, fainting, and nausea	7	Experiment, for example, reaction to mock crime or cheating instructions	12
Action, for example, terminate participation and seek counseling	2	Personal	4
Psychological and physiological	2	Unclear/other	6
Incidents related to research participants' autonomy			
Psychological	6	Experiment	5
Action	1	Combined	1
		Other	1
Sum	63		63

3.1 | Occurrence, nature, and probability of incidents

3.1.1 | Incident occurrence and incident-related studies

Of the 188 respondents, 59 (31%) answered the question on the occurrence of incidents (item 3) with *yes*. The reported incident-related studies were conducted between 1977 and 2016, most in the past 10 years (86%; 2007–2016), some between 1997 and 2006 (7%), and in the 20 years prior to 1997 (7%). Incident-related studies had been conducted on four different continents, with countries including the United States (33%), the United Kingdom (13%), Canada (13%), Germany (10%), Australia (10%), Scandinavian countries (8%), The Netherlands (7%), Belgium (4%), Croatia (2%), and South Korea (2%).

TABLE 4 Comments and recommendations

Comment (n = 47)[a,b]	%	Recommendation (n = 15)[a,c]	%
ECs are too strict/unreasonable	26	Should take calculated risk	31
Lack of knowledge and competence in ECs	17	Need for experts/competence	19
ECs are reasonable and competent	13	ECs should be evidence-based	13
ECs are important	11	ECs need to be flexible	13
Differences across ECs	11	Form advisory board	6
ECs overstep	9	Ethics should be processed in psychology faculty	6
ECs have wrong focus	4	ECs should allow for post-hoc explanations	6
ER is antiscience	4	EC's power should be limited	6
ECs underestimate researchers	4		
Other EC-related comments	4		
Non-EC-related comments	13		

Note. EC: ethics committee; ER: ethics review.

[a]Multiple comments possible. [b]Percentages refer to proportion from all comments (70 in total). [c]Percentages refer to proportion from all recommendations made (16 in total).

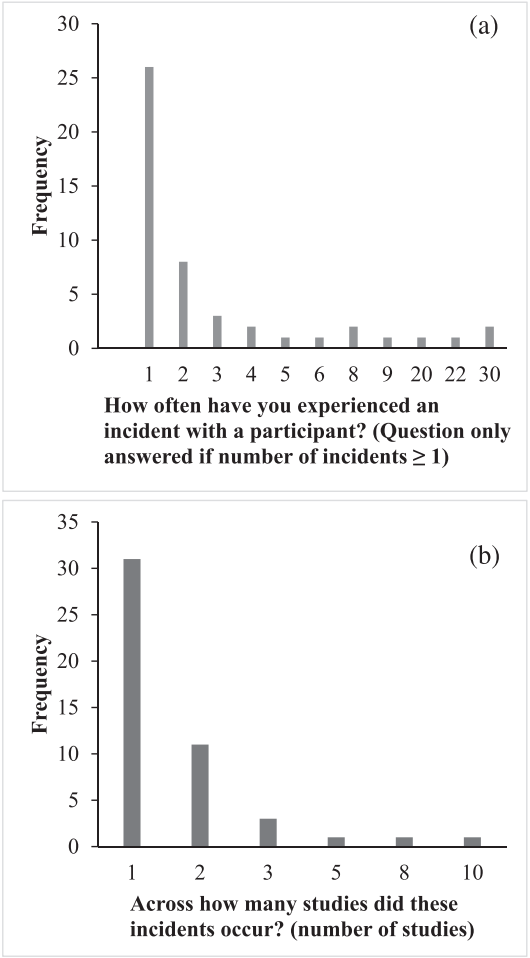

FIGURE 1 Frequency distributions for Item 4. (Phrasing adapted for Figure; for original phrasing, see Table 1)

Respondents who had experienced an incident with a participant (i.e., answered "yes" to Item 3) reported between one and 30 incident occurrences (M = 4.1, SD = 6.9; n = 48, Item 4) that had occurred across 1–10 studies (M = 1.8[2]; SD = 1.7; mode and median = 1). The resulting incident rate per study with an incident was M = 2.1 (SD = 2.7, mode and median = 1). Figure 1 depicts the distribution of the answers for the two parts of Item 4. Four respondents reported 20–30 incidents, of which many referred to children being separated from their parents (which is common in young children; e.g., Dallaire & Weinraub, 2005) and caffeine-induced paranoia (excessive caffeine consumption has been associated with development or aggravation of psychotic symptoms; Wang, Woo, & Bahk, 2015). Incident-related studies investigated eyewitnesses (27%); interrogation, interviewing, and false confessions (13%); lie detection and credibility (11%); memory, recognition, misinformation, and false memories (11%); psychopathology and mental health (7%); stress, trauma, and posttraumatic stress disorder (6%); jury behavior and legal decision-making (6%); child sexual abuse, sexual assault, and victimology (4%); and psychopharmacology (4%).

The majority of respondents reporting incidents had sought ethics approval for the incident-related studies (n = 42; 91%), whereas a minority had not (n = 4; 9%). Given the small number of studies conducted without ER, a statistical comparison between these two categories was not possible. What can be said on a purely descriptive level is that researchers who had sought ER did not report fewer incidents (M = 4.1; SD = 6.8, or M = 2.7; SD = 3.5, with two outliers excluded) than those who had not (M = 1.0; SD = 0.0).

[2]The second part of Item 4 (In how many of your studies did such incidents happen?) was misunderstood by eight respondents. That is, rather than indicating across how many studies the mentioned incidents had happened, they indicated how many studies they had ever conducted. Such a misunderstanding is evident when the number of studies noted is larger than the number of incidents (i.e., it is impossible that one incident happened across 80 studies). In such cases, we corrected the answer by setting the number of incidents to the number of studies. In six of the eight cases, the number of incidents was one. In the remaining two cases, the number or reported incidents were two and three. When setting the number of studies to one (rather than two or three) the mean incident rate per study with an incident was 1.7 (SD = 1.7; mode and median = 1).

3.1.2 | Incident probability

The sum of reported incident frequencies added up to an estimate[3] of 197 across all respondents. This rate can be qualified by the total number of studies conducted by respondents in our sample, namely, between 4,200 and 10,150 studies.[4] Dividing the number of incidents by the number of studies results in an estimate of 0.02 to 0.05 incidents per study or between two and five incidents per 100 studies. If one were to assume an average of 150 participants per study, this would be equivalent to one to three incidents per 10,000 participants.[5]

3.1.3 | Open-incident descriptions

In total, 63 incidents were described by 48 different researchers when invited to describe such incidents in an open-question format (Item 5; range: one to four incidents; *M* = 1.4; *SD* = 0.7). As said before, incidents were coded for (a) observable behavior judged an incident, (b) assumed trigger, (c) type of incident, and (d) assumed cause. The findings for each category can be found in Tables 2 and 3. Fifty-six incidents could be classified as referring to participants' well-being or discomfort (89%), seven to autonomy-related issues (11%).

Incidents related to participants' well-being or discomfort most commonly involved crying or emotional reactions (20 of 68); signs of stress and distress (13 of 68); and anger, being upset, or offended (12 of 68). Other behaviors included dizziness and fainting (six of 68), termination of participation (five of 68), fear (three of 68), nausea and indisposition (three of 68), and filing a complaint (two of 68). The most common incident triggers concerned the personal background of participants (12 of 72), having to recall autobiographic memories (nine of 72), and the use of emotional stimuli (seven of 72). Incident type was mostly psychological (45 of 56). A minority concerned physiological (seven of 56) or action type incidents (two of 56). Most incidents fell into the combined cause category (34 of 56). Few incidents were attributed to personal factors (four of 56), some to experimental features (12 of 56).

Incidents related to participants' autonomy mostly involved anger (four of seven). The primary trigger for such incidents was the use of deception, a false memory paradigm, or lying instructions (six of eight). The types of incidents were mostly psychological in nature (six of seven); the assumed causes being mostly the experiment (five of seven).

When asked specifically if they thought that the elements of the study itself caused the incidents, 19 researchers (of *n* = 35, 54%) did not, whereas 15 (43%) did think this was the case.

3.2 | ECs and incidents in general

Online survey respondents found ECs *too strict* (48 of *n* = 49, 98%) rather than *too lenient* (*n* = 1, 2%) and *too conservative* (55 of *n* = 55, 100%), rather than *too liberal* (*n* = 0). These items were phrased slightly different for paper-and-pencil respondents (see Table 1). In this group, 14 of 34 respondents (41%) assessed ECs as being *too strict or conservative*, whereas 16 (47%) disagreed. Thirty-three of 36 (92%) paper-and-pencil respondents disagreed with the statement that ECs were *too liberal or lenient*, whereas no one answered this question with yes.

Just above half of the respondents disagreed with the statement that ECs had sufficient expertise to evaluate the risks and benefits of legal psychology studies (75 of *n* = 133, 56%). A similar percentage of respondents stated they had never refrained from carrying out an experiment because they anticipated too much critique from the EC (94 of *n* = 162, 58%), whereas 68 researchers (36%) said they actually had.

Most respondents disagreed with the statement that incidents proved the effectiveness of an experimental manipulation or a research approach (72 of *n* = 128, 56%); 43 agreed (34%), and 13 (10%) answered *do not know*.

3.3 | Comments on ECs and recommendations

Forty-seven respondents came up with commentaries (70 comments in total) in response to the final open item, and 15 of them formulated recommendations (16 in total) or requirements for ECs' work. A complete overview can be found in Table 4. Researchers' comments most frequently referred to the notions that (their) ECs were too strict or unreasonable (26%), lacked knowledge or competence (17%), and overstepped (9%), on the one hand. On the other hand, researchers commented that (their) ECs were reasonable and competent (13%) and that ECs were important (11%).

Recommendations and demands referred to the notion that ECs should take a calculated risk (rather than handling a no-risk policy; 33%). Researchers also called for ECs to (partially) be constituted of experts in the field (20%), making ECs subject to science (i.e., evidence-based; 13%) and flexibility in ECs in the sense that assessment should be more tailored to individual research proposals (13%).

4 | DISCUSSION

This explorative study was a first attempt to estimate the risk of being involved in an incident when participating in legal psychology

[3]Researchers frequently indicated that their information were rough estimates. Thus, our calculations should be considered as rough approximations that should be seen in the light of limitations of self-report measures.

[4]These figures were computed by multiplying the extreme ends of the answer options provided for Item 2 (How many experiments did you carry out in those years?). A researcher who answered one to 10 studies would once be treated as having run one study and once as having run 10 studies.

[5]We base this assumption of 150 participants per study on two pieces of information: First, inspection of 17 recent meta-analyses in the field of legal psychology showed that the mean sample sizes in the included studies varied between 60 and 12,933 (e.g., Scott & Brown, 2018; Stewart, Woody, & Pulos, 2018; full references available from the first author). Even when excluding two meta-analyses with exceptionally large average samples of 2,300+ participants, the unweighted mean sample size per study was 192. Second, we inspected the sample sizes of the first six studies that we retrieved with search terms eyewitnesses, offenders, interrogation, or victims (in title and related to legal psychology) in a Web of Science search (i.e., 24 in total; references available from the first author). The average sample size across those studies was *N* = 297 (*SD* = 241.5, range 19–2,792). When excluding the three studies with larger average samples than 400, the mean was *N* = 148 (*SD* = 112.3).

research. About 30% of active researchers reported at least one time when such an incident occurred. About nine of 10 incidents described referred to instances of discomfort or impairment of participants' well-being, whereas one of 10 incidents concerned threats to participants' autonomy. The probability of being involved in an incident was estimated to be two to five incidents per 100 studies or one to three incidents per 10,000 participants. Resorting to the biomedical sciences has been much criticized when it comes to ER in the social sciences on the ground that adverse effects in biomedical research are potentially much more severe (Boden et al., 2009; Dingwall, 2006; Schrag, 2011). Still, it is an informative exercise to relate the results of the current study to established risk probabilities in the biomedical domain (cf. Nutt & Sharpe, 2008, for an application of these risk probabilities of psychological interventions during psychotherapy). In medical terms, the risk of a side effect that occurs with a probability of one to 10 in 10,000 cases is considered *rare* (Medical Dictionary for Regulatory Activities).[6] Alternatively, the reported incidents could be compared with the prevalence of similar incidents in real life. For example, crying—the most common incident in our sample—is a common expression of emotion among adults. Specifically, on average adults cry or tear-up between five to eight times a month (Bylsma, Croon, Vingerhoets, & Rottenberg, 2011; Frey, 1983). Our data, then, do suggest that, overall, the base rate of adverse effects for participants in legal psychology research is low—or rare. Following the argument that legal psychology is one of the disciplines in which intrusive manipulations (e.g., eliciting false confessions; exposing eyewitnesses to gruesome scenes) are not uncommon, the estimated hazards established for this particular domain might represent the higher end of the risk distribution within the social science, although this is open to empirical testing.

Unfortunately, the current data are silent on the crucial issue of whether ER or the mere presence of an EC affects hazards in a positive way, because the vast majority of incidents reported occurred in studies for which ethics approval had been obtained. This precluded a statistical comparison of the risk of experiencing an incident in a study that was versus was not subject to ER. Thus, we were unable to show that ER does or does not reduce the risk of participation in legal psychology research, and to the best of our knowledge, there is no study that looked into this. Our data are consistent with the idea that the presence of an EC (or the procedure of ER) succeeds in preventing researchers from conducting risky studies. Alternatively, they do not contradict the idea that ECs are too strict, thus preventing researchers from conducting their research in the most adequate way. A more controlled approach to risks and adverse effects of research is needed to disentangle these interpretations.

One more controlled approach to the question to what extent ECs succeed in protecting research participants would be to conduct a series of studies with and without ER in a randomized controlled trial (cf. Resnik, 2015). Given the considerable impact of ER on researchers' work and its substantial burden on financial and time resources (Grady, 2010; Gunsalus, 2006), it would be in the interest of researchers, research institutions, participants, and funding organizations to collect these type of data. Such venture, to be sure, comes with numerous challenges. One issue concerns the definition of outcome measures (e.g., Nicholls et al., 2015; Resnik, 2015; Scherzinger & Bobbert, 2017; Taylor & Ervin, 2017). The number of incidents, as researched here, can only be a starting point. Discussions on defining adverse outcomes can be found elsewhere (e.g., Coleman & Bouësseau, 2008; Resnik, 2015; Taylor & Ervin, 2017). Another important methodological issue concerns how to isolate ER impact on outcomes variables from the impact of other factors (i.e., participant factors; differences across studies and other confounding factors). Researchers would also have to consider legal issues. For example, a randomized controlled trial involving studies with and without ER can only take place in countries in which ER is not mandated by law (for all types of research; e.g., Italy, Spain, Sweden; Swedish Research Council, 2017). In these countries, researchers are held to the ethical code of conduct, but adherence is not inspected on an individual basis. To be sure, the argument that withholding one experimental group from ER is ethically questionable deserves careful consideration. If a no-ER condition were to be considered improper, outcomes of different types of ER (e.g., expedited, exempt, or full review) might be compared (Resnik, 2015).

The second part of the survey addressed researchers' view and opinion on ECs. Whereas some respondents were satisfied with their ECs, judging them as competent and important, the general tenor was one of discontent. This is in line with previous findings (e.g., Gray & Cooke, 1980; Gunsalus, 2004; Hammersley, 2009). Self-reported reasons for dissatisfaction included, amongst others, conservative and strict decision-making, lack of expertise, and overstepping demands. Although some of these assessments may be colored by bias (i.e., respondents might be motivated to convey an overly optimistic sketch of the risks associated with their research line), these data should give us pause. They may be taken as a thought-provoking message for all involved in ER—research institutions, ECs members and chairs, and researchers—to discuss their EC's mode of operation and composition.

The data presented here must be seen in the light of their limitations. Our study is based on self-report data and some questions refer to events that may have occurred many years ago, introducing effects of forgetting and memory distortion. As a result, the statistics provided here can only serve as a rough estimate or approximation of the number of incidents our respondents experienced during their careers. Future research might approach the number and frequency of incidents in a more controlled, prospective approach, for example, by using a diary method. Additionally, the number of complaints files at ECs may serve as a measure of participation risk. Another limitation concerns possible response bias. Researchers who had stress-related incidents with participants to report may have been more motivated to respond, given the phrasing of our invitation to participate. As a result, the proportion of respondents who reported an incident and the estimated incident probabilities per study and per participant might be inflated. Thus, the actual risk of being involved in an incident may be even lower than

[6]Standards for describing probability of side effects of medication range from *very common* (one in 10 cases), *common* (one to 10 in 100 cases), *uncommon* (one to 10 in 1,000 cases), *rare* (one to 10 in 10,000 cases) to *very rare* (<1 in 10,000 cases; Medical Dictionary for Regulatory Activities)

estimated. It is also possible that we did not succeed in reaching various groups of researchers in the field in a similar way, even though the research topics mentioned suggest that prominent and active areas such as the psychology of eyewitness testimony, information gathering from suspects, victimology, legal decision-making, and offender psychopathology and mental health were covered. For a more systematic approach, future studies might choose an inclusion strategy that allows inferences about the representativeness of the sample. A final limitation concerns the exclusive reliance on researchers' views in the current study. Future studies could be extended by including (potential) research participants, EC members, and chairs.

To conclude, the risk of harm and injury from research participation in legal psychology seems to be low by several standards. This relative low risk is accompanied by an overall skeptical attitude of researchers towards ER. We believe that the present survey contributes uniquely to the research on the efficacy of ethical committees in social sciences and paves the way for more evidence-based policies and practices in this field.

DISCLOSURE

The authors declare that they reported all measures, conditions, and data exclusions.

ACKNOWLEDGEMENTS

We thank all participating researchers for sharing their experiences and views with us. We thank Beyza Gürtekin for her help in coding the data.

ORCID

Melanie Sauerland https://orcid.org/0000-0002-7401-1365
Anna Sagana https://orcid.org/0000-0003-4776-8431
Harald Merckelbach https://orcid.org/0000-0002-5116-7826

REFERENCES

Abbott, L., & Grady, C. (2011). A systematic review of the empirical literature evaluating IRBs: What we know and what we still need to learn. *Journal of Empirical Research on Human Research Ethics*, *6*, 3–19. https://doi.org/10.1525/jer.2011.6.1.3

All European Academies. (2017). *The European code of conduct for research integrity*. Berlin, Germany: ALLEA.

Baumeister, R., Vohs, K., & Funder, D. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Baumrind, D. (1964). Some thoughts on ethics of research: After reading Milgram's "Behavioral Study of Obedience". *American Psychologist*, *19*, 421–423. https://doi.org/10.1037/h0040128

Boden, R., Epstein, D., & Latimer, J. (2009). Accounting for ethos or programmes for conduct? The brave new world of research ethics committees. *The Sociological Review*, *57*, 727–749. https://doi.org/10.1111/j.1467-954x.2009.01869.x

Burnham, P. (1966). Letters to the Editor. *Science*, *152*, 448–448. https://doi.org/10.1126/science.152.3721.448-a

Bylsma, L. M., Croon, M. A., Vingerhoets, A. J., & Rottenberg, J. (2011). When and for whom does crying improve mood? A daily diary study of 1004 crying episodes. *Journal of Research in Personality*, *45*, 385–392. https://doi.org/10.1016/j.jrp.2011.04.007

Coleman, C., & Bouësseau, M. (2008). How do we know that research ethics committees are really working? The neglected role of outcomes assessment in research ethics review. *BMC Medical Ethics*, *9*. https://doi.org/10.1186/1472-6939-9-6

Colnerud, D. (2015). Ethical dilemmas in research in relation to ethical review: An empirical study. *Research Ethics*, *10*, 238–253. https://doi.org/10.1177/1747016114552339

Dallaire, D. H., & Weinraub, M. (2005). Predicting children's separation anxiety at age 6: The contributions of infant–mother attachment security, maternal sensitivity, and maternal separation anxiety. *Attachment & Human Development*, *7*, 393–408. https://doi.org/10.1080/14616730500365894

Dingwall, R. (2006). Confronting the anti-democrats: The unethical nature of ethical regulation in social science. *Medical Sociology Online*, *1*, 51–58.

Dingwall, R. (2008). The ethical case against ethical regulations in humanities and social science research. *21st Society*, *3*, 1–12. https://doi.org/10.1080/1745640701749189

Edwards, S. (2009). The role, remit and function of the research ethics committee—1. The rationale for ethics review of research by committee. *Research Ethics*, *5*, 147–150. https://doi.org/10.1177/174701610900500405

Edwards, S. (2010). The role, remit and function of the research ethics committee—2. Science and society: The scope of ethics review. *Research Ethics*, *6*, 58–61. https://doi.org/10.1177/174701611000600207

Frey, W. H. (1983). Crying behavior in the human adult. *Integrative Psychiatry*, *1*, 94–98.

González-Sala, F., Osca-Lluch, J., Gil, F. T., & Ortega, M. P. (2017). Characterization of legal psychology through psychology journals included in criminology & penology and law categories of Web of Science. *Anales de Psicología/Annals of Psychology*, *33*, 411–416. https://doi.org/10.6018/analesps.33.2.262591

Grady, C. (2010). Do IRBs protect human research participants? *Journal of the American Medical Association*, *304*, 1122–1123. https://doi.org/10.1001/jama.2010.1304

Gray, B. (1977). The functions of human subjects review committees. *American Journal of Psychiatry*, *134*, 907–910. https://doi.org/10.1176/ajp.134.8.907

Gray, B., & Cooke, R. (1980). The impact of institutional review boards on research. *The Hastings Center Report*, *10*, 36–41. https://doi.org/10.2307/3560505

Gunsalus, C. (2006). Mission creep in the IRB world. *Science*, *312*, 1441–1441. https://doi.org/10.1126/science.1121479

Gunsalus, C. K. (2004). The nanny state meets the inner lawyer: Overregulating while underprotecting human participants in research. *Ethics & Behavior*, *14*, 369–382. https://doi.org/10.1207/s15327019eb1404_7

Haggerty, K. (2004). Ethics creep: Governing social science research in the name of ethics. *Qualitative Sociology*, *27*, 391–414. https://doi.org/10.1023/b:quas.0000049239.15922.a3

Hamilton, A. (2005). The development and operation of IRBs: Medical regulations and social science. *Journal of Applied Communication Research*, *33*, 189–203. https://doi.org/10.1080/00909880500149353

Hammersley, M. (2006). Are ethics committees ethical? *Qualitative Research*, *2*, 4–8.

Hammersley, M. (2009). Against the ethicists: On the evils of ethical regulation. *International Journal of Social Research Methodology*, *12*, 211–225. https://doi.org/10.1080/13645570802170288

Humphreys, L. (1970). *Tearoom trade. Impersonal sex in public places*. London, U.K.: Duckworth.

Hunter, D. (2014). Can significant differences in regulating medical and non-medical research be justified? *Monash Bioethics Review*, *32*, 254–267. https://doi.org/10.1007/s40592-015-0022-2

Jennings, S. (2012). Response to Schrag: What are ethics committees for anyway? A defence of social science research ethics review. *Research Ethics*, *8*, 87–96. https://doi.org/10.1177/1747016112445423

Kaufmann, B. (1967). Ethics by osmosis: A problem for school psychology. *Journal of School Psychology*, *5*, 310–315. https://doi.org/10.1016/0022-4405(67)90010-6

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310

Lemonick, M., & Goldstein, A. (2002). At your own risk. Some patients join clinical trials out of desperation, others to help medicine advance. Who is to blame if they get sick – or even die? *Time*, *159*, 46–56.

Medical Dictionary for Regulatory Activities. (2018). Retrieved from https://www.meddra.org/

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, *67*, 371–378. https://doi.org/10.1037/h0040525

Nicholls, S., Brehaut, J., & Saginur, R. (2012). Social science and ethics review: A question of practice not principle. *Research Ethics*, *8*, 71–78. https://doi.org/10.1177/1747016112445435

Nicholls, S., Hayes, T., Brehaut, J., McDonald, M., Weijer, C., Saginur, R., & Fergusson, D. (2015). A scoping review of empirical research relating to quality and effectiveness of research ethics review. *PLoS ONE*, *10*, e0133639. https://doi.org/10.1371/journal.pone.0133639

Nutt, D. J., & Sharpe, M. (2008). Uncritical positive regard? Issues in the efficacy and safety of psychotherapy. *Journal of Psychopharmacology*, *22*, 3–6. https://doi.org/10.1177/0269881107086283

Resnik, D. (2015). Some reflections on evaluating institutional review board effectiveness. *Contemporary Clinical Trials*, *45*, 261–264. https://doi.org/10.1016/j.cct.2015.09.018

Scherzinger, G., & Bobbert, M. (2017). Evaluation of research ethics committees: Criteria for the ethical quality of the review process. *Accountability in Research*, *24*, 152–176. https://doi.org/10.1080/08989621.2016.1273778

Schrag, Z. (2011). The case against ethics review in the social sciences. *Research Ethics*, *7*, 120–131. https://doi.org/10.1177/174701611100700402

Scott, T., & Brown, S. L. (2018). Risks, strengths, gender, and recidivism among justice-involved youth: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *86*, 931–945. https://doi.org/10.1037/ccp0000343

Stewart, J. M., Woody, D. W., & Pulos, S. (2018). The prevalence of false confessions in experimental laboratory simulations: A meta-analysis. *Behavioral Sciences & the Law*, *36*, 12–31. https://doi.org/10.1002/bsl.2327

Swedish Research Council (2017). *Good research practice*. Stockholm, Sweden: Swedish Research Council.

Taylor, H., & Ervin, A. (2017). A measure of effectiveness is key to the success of sIRB policy. *The American Journal of Bioethics*, *17*, 41–43. https://doi.org/10.1080/15265161.2017.1328537

U.S. Department of Health & Human Services. (2018). The federal policy for protection of human research subjects (45 CFR part 46). Washington, DC. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html

Von Hoffman, N. (1970). Sociological snoopers and journalistic moralizers. *Society*, *7*, 4–8. https://doi.org/10.1007/bf02804293

Wang, H. R., Woo, Y. S., & Bahk, W.-M. (2015). Caffeine-induced psychiatric manifestations: A review. *International Clinical Psychopharmacology*, *30*, 179–182. https://doi.org/10.1097/YIC.0000000000000076

How to cite this article: Sauerland M, Sagana A, Merckelbach H. A survey on adverse incidents in legal psychology studies: Reflections on ethics review. *Appl Cognit Psychol*. 2019;33:682–692. https://doi.org/10.1002/acp.3530