Archival Analyses of Eyewitness Identification Test Outcomes: What Can They Tell Us About Eyewitness Memory?

Ruth Horry, Paul Halford, Neil Brewer*

Flinders University, Adelaide, Australia

Rebecca Milne

Portsmouth University

Ray Bull

Portsmouth University, University of Derby

Accepted for publication in *Law and Human Behavior*. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

* Corresponding author.

Neil Brewer, School of Psychology, Flinders University, GPO Box 2100, Adelaide, 5001, South Australia. Neil.brewer@flinders.edu.au.

Abstract

Several archival studies of eyewitness identification have been conducted, but the results have been inconsistent and contradictory. We identify some avoidable pitfalls that have been present in previous analyses, and we present new data in which we address these pitfalls. We explored associations among various estimator variables and lineup outcomes for 833 'real life' lineups, including 588 lineups where corroborating evidence of the suspect's guilt existed. Suspect identifications were associated with exposure duration, viewing distance, and the age of the witness. Non-identifications were associated with the number of perpetrators. We also consider some of the inherent, unavoidable limitations with archival studies and consider what such studies can really tell researchers. We conclude that differences in sampling prohibit sensible comparisons between the results of laboratory and archival studies, and that the informative value of archival studies is actually rather limited.

Keywords: Eyewitness identification, estimator variables, archival data.

Archival Analyses of Eyewitness Identification Test Outcomes: What Can They Tell Us About Eyewitness Memory?

Hundreds of laboratory experiments have investigated the effects of many variables on the ability of witnesses (typically young adult students) to identify perpetrators from lineups and to reject lineups that do not contain the perpetrator. Following Wells (1978), researchers have categorized forensically-relevant variables into two categories: *system* variables and *estimator* variables. System variables are within the control of the legal system, and include factors such as pre-lineup instructions (Malpass & Devine, 1981), selection of lineup foils (Wells, Rydell, & Seelau, 1993), and the presentation format of the lineup (Lindsay & Wells, 1985). Estimator variables are not in the control of the legal system, and so their impact on eyewitness reliability can only be estimated. Estimator variables include encoding conditions (exposure duration, viewing distance, visibility), retention interval, characteristics of the witness and suspect (e.g., sex, age, ethnic background), and factors relating to stress and arousal (e.g., presence of a weapon, threat or use of violence, witness' level of involvement).

Archival studies involve exploring associations between various predictor variables (usually estimator variables) and lineup identification outcomes in real cases. Several such studies have been conducted, both in the United States (Behrman & Davey, 2001; Tollestrup, Turtle, & Yuille, 1994) and in the United Kingdom (Horry, Memon, Wright, & Milne, 2012; Memon, Havard, Clifford, Gabbert, & Watt, 2011; Pike, Brace, & Kynan, 2002; Valentine, Pickering, & Darling, 2003; Wright & McDaid, 1996). The results of previous archival studies have been inconsistent and, at times, contradictory. In this paper, we consider some of the limitations that have contributed to these inconsistencies. Some of these limitations are avoidable, while others are not. We first consider four limitations that can be addressed, at least to some degree, in archival studies. These limitations include: 1) absence of ground truth regarding suspect guilt; 2) use of inappropriate statistical methods; 3) small or biased samples; and 4) inclusion of cases in which the culprit was familiar to the witness. We present a new archival dataset in which we addressed each of these limitations. In the Discussion, we discuss some of the fundamental difficulties with archival studies and what such studies are really able to tell us about eyewitness memory. First, however, we briefly review the results of previous field studies. We then consider how some aspects of the methods and analyses of these studies may undermine our ability to draw strong conclusions from them.

Previous archival studies of eyewitness identification

Archival studies have principally focused on the influence of estimator variables. Tollestrup et al. (1994) published the first archival study of eyewitness identification, sampling all cases of robbery and fraud that took place within a given time period in a single Canadian police force. The authors were particularly interested in the impact of arousal and threat, focusing on the involvement of the witness (victim vs. bystander), the type of crime (robbery vs. fraud) and the presence of a weapon. They found that victims of robbery (who probably experienced the highest arousal levels in this sample) were most likely to identify the police suspect whereas witnesses to fraudulent activities (who likely experienced the lowest arousal levels in this sample) were least likely to identify the suspect. Thus, in contrast to laboratory research (Deffenbacher, Bornstein, Penrod, & McGorty, 2004), these data seemed to suggest that arousal could be beneficial for eyewitness identification performance. Tollestrup et al. also reported that the presence of a weapon in robbery cases was associated with fewer suspect identifications than the absence of a weapon, supporting laboratory research on the ‘weapon focus’ effect (e.g., Fawcett, Russell, Peace, & Christie, 2013; E. F. Loftus, G. R. Loftus, & Messo, 1987). However, both of these predictor variables (weapon presence and involvement of witness) were confounded with retention interval within their sample. As we argue later, failing to account for the multicollinearity among factors may have created spurious effects in their study. At the very least, such multicollinearity makes it difficult to draw any firm conclusions.

One other North American study (Behrman & Davey,2001) included a wider range of estimator variables in the analysis: characteristics of the witness and offender (e.g., sex, age, ethnicity), and factors associated with likely arousal and stress (weapon presence, victims vs. bystanders). In contrast to Tollestrup et al.'s (1994) data, Behrman and Davey found no significant associations between suspect identifications and weapon presence or victimization. They did, however, report that retention interval was negatively associated with suspect identification rates.

In a British archival study, Wright and McDaid (1996) sampled a large number of lineups from several locations in London. They were interested in comparing outcomes from lineups conducted at specialist identification suites with those conducted in ordinary police stations. However, this was complicated by the fact that the types of cases handled by the identification suites differed in several ways from the lineups conducted in police stations. For example, violent crimes and minority-ethnicity suspects were more likely to be dealt with by identification suites than by police stations. Wright and McDaid only explored two estimator variables (use of violence and ethnicity of suspect), but they found that minority-ethnicity suspects were more likely to be identified than suspects from European backgrounds. Use of violence was not significantly associated with lineup outcomes.

Pike et al. (2002) analyzed data from 2,628 lineups conducted in England. They reported that robbery produced lower suspect identification rates than assault and theft cases, speculating that robberies may typically involve shorter exposure durations than other types of crimes. They also found that older adult witnesses (over 60) and child and adolescent witnesses (under 17) identified suspects less frequently than young adult witnesses. Several variables were not significantly associated with identification outcomes, including ethnicity of suspect and witness, use of violence, and presence of a weapon.

Valentine et al. (2003) explored the widest range of estimator variables of any archival study to date. They included factors associated with quality of encoding (e.g., exposure duration, lighting conditions, viewing distance, obstructed view) as well as characteristics of the witness and suspect (sex, age, ethnicity) and crime factors (type of crime, weapon presence). When all of the factors were entered in a global regression model, witness age was negatively associated with suspect identifications. The association between exposure duration and suspect identifications was marginally significant, with witnesses who viewed the perpetrator for more than 60 seconds more likely to identify the police suspect than witnesses who viewed the perpetrator for less than 60 seconds. No other variables, including weapon presence, made significant independent contributions to the global regression model.

Memon et al. (2011) sampled all lineups conducted in Scotland in 2008. The authors focused on witness factors such as age, sex, and level of involvement (bystander vs. victim), as well as crime type and retention interval. Consistent with Valentine et al. (2003), they found that younger witnesses were more likely to identify suspects than older witnesses. However, in contrast with Tollestrup et al. (1994), bystanders were more likely to identify the police suspect than victims. Violent crimes were associated with a higher suspect identification rate than crimes of dishonesty, and suspect identifications generally decreased as retention interval increased.

The most recent archival study was reported by Horry et al. (2012). The authors focused on estimator variables such as characteristics of the witness and suspect (sex, age, and ethnicity), the type of crime, and the retention interval. Retention interval and the age difference between the witness and suspect were negatively associated with suspect identifications. Additionally, majority-ethnicity suspects were more likely to be identified than minority-ethnicity suspects and robberies were associated with fewer suspect identifications than any other type of crime.

Across the archival studies described above, the inconsistences outweigh the consistencies. Regarding consistencies, four studies (Behrman & Davey, 2001; Horry et al., 2012; Memon et al., 2011; Tollestrup et al., 1994) reported that suspects were less likely to be identified after longer retention intervals. Four studies found that older witnesses were less likely to identify suspects than younger witnesses (Horry et al.; Memon et al.; Pike et al., 2002; Valentine et al., 2003). Regarding inconsistences, weapon presence was associated with suspect identifications in just one study (Tollestrup et al.). Other factors have even produced opposite results in different studies. For example, Tollestrup et al. found that victims were more likely than bystanders to identify suspects, whereas Memon et al. found that bystanders were more likely than victims to identify suspects. Finally, Wright et al. found that minority-ethnicity suspects were more likely to be identified than majority-ethnicity suspects, while Horry et al. reported the reverse and Pike et al. found no association.

Why have results from archival studies been so inconsistent? When conducting archival studies, researchers have no experimental control over the data. Participants cannot be assigned to conditions and many factors will be intercorrelated, adding additional noise to the data. This noise may overwhelm genuine associations, leading to Type 2 errors. Additionally, inconsistencies in results may be produced by inconsistencies in methods. We argue below that each published archival study suffers from at least one serious, but avoidable, limitation. In the following section, we outline four such limitations and argue how each of them may have contributed to Type 1 and Type 2 errors in the literature.

Limitations of previous archival studies

Absence of ground truth. In the laboratory, researchers know the identity of the perpetrator. Thus, lineups can be designated as target-present and target-absent and identifications can be categorized as accurate or inaccurate. In a criminal investigation, the identity of the perpetrator is not known. Indeed, if the identity of the perpetrator were known, there would be no need to conduct a lineup. In archival research, therefore, it is not possible to know whether suspects are guilty or innocent – and, in turn, whether any given suspect identification is accurate or inaccurate. If we were able to assume that the base-rate of suspect guilt was approximately equal over all levels of our predictors (i.e., subject only to random error), we could dismiss this problem as one of statistical noise in the data set. However, is it really a safe assumption that suspect guilt is uncorrelated with various other predictors? Below we will argue that this assumption is not safe, and that selection effects are likely to influence the culprit-present base rate across different levels of our predictors.

Investigators have finite resources that must be managed efficiently in order to meet goals of public safety and not all crimes will be allocated the same amount of resources. Serious crime investigations (i.e., those involving weapons and violence) are likely to be allocated more resources than less serious crimes, which may produce differences in culprit-present base rates in lineups. The amount of evidence that police officers require in order to construct a lineup may also vary from crime to crime. In some cases, a history of similar offences may be sufficient to warrant a lineup, whereas in others, corroborating physical evidence may be required. Again, the standards of corroborating evidence are likely to vary with other aspects of the event.

If the base-rate of culprit presence does indeed covary with various predictor variables, this could produce both Type 1 and Type 2 errors. Which type of error is more likely will depend upon the direction of the association. For example, if the condition that we would expect to produce the highest suspect identification rate is associated with a higher culprit present base rate (e.g., if suspects are more likely to be guilty when arrested after a shorter delay), then the apparent association between the predictor and the outcome may be grossly inflated, increasing the risk of a Type 1 error. In other words, suspects might be identified more frequently after a short retention interval because they are, in fact, more likely to be guilty. On the flip side, if the condition that we would expect to produce the highest suspect identification rate was associated with a lower culprit-present base rate (e.g., if suspects for weapon-present crimes were more likely to be guilty than suspects in weapon-absent crimes), then any genuine association between the predictor and lineup outcomes may be obscured, increasing the risk of a Type 2 error. The effect of the predictor might be counteracted by the difference in culprit-present base rates.

There is no sure-fire way to protect against this problem in archival research. However, North American studies have taken the approach of categorizing cases according to the amounts and types of corroborating evidence available. For example, Behrman and Davey (2001) coded corroborating evidence as being of minimal probative value (e.g., similarities with previous offences, anonymous tips, information provided by a co-felon) or substantive probative value (e.g., video evidence, possession of stolen items, confession). Suspect identification rates were much higher in cases with corroborating evidence (92% for ‘substantive probative value’ and 71% for ‘minimal probative value’) than in cases without any corroborating evidence (57%), presumably because culprit-present base rates were higher in cases with corroborating evidence. Tollestrup et al. (1994) categorized cases as including no corroborating evidence, some corroborating evidence (e.g., fingerprints, fraudulent cheques including the suspect’s driver’s license number), or a confession. In line with Behrman and Davey’s findings, suspects were more likely to be identified if the suspect confessed (48%) or if there was corroborating evidence (42%) than if there was no corroborating evidence (18%). Once again, it is likely that the culprit-present base rate was higher in cases with confessions or other evidence than in cases without any corroborating evidence.

British archival studies have generally been unable to use such an approach because they have included ongoing cases in which evidence is still being gathered. Due to confidentiality concerns around ongoing investigations, any information on corroborating evidence has generally not been shared with researchers. This potentially leaves the British studies vulnerable to the types of errors discussed above. An important and innovative feature of our archival analysis was the use of case files from past lineups, allowing us to collect critical information on corroborating evidence. We conducted separate analyses on our corroborated and uncorroborated cases, allowing us to explore cases in which suspect guilt is more or less likely.

There is an important caveat with this approach: suspect guilt is not certain, even in cases with corroborating evidence that appears quite strong (e.g., confessions, Kassin & Kiechel, 1996). A burgeoning literature on investigative decision-making has shown that obtaining one piece of evidence can influence the likelihood of further evidence being obtained (see Kassin, Dror, & Kukucka, 2013, for a review). This is a form of confirmation bias, which has been named ‘corroboration inflation’ (Kassin, 2012). We consider this issue at greater length in the Discussion.

Use of inappropriate statistical methods. There are two distinct difficulties posed by archival data that are not faced by laboratory researchers in typical eyewitness identification experiments: multicollinearity among variables and non-independence of data points. There are statistical methods that can be used to deal with these problems, but many archival studies have failed to use them, relying instead on chi-square tests of association (e.g., Behrman & Davey, 2001; Memon et al., 2011; Tollestrup et al., 1994).

Unlike in the laboratory, witnesses in archival studies cannot be randomly allocated to groups or distributed evenly across cells of a design. Rather, some variables will tend to co-occur quite frequently, creating multicollinearity. If not properly accounted for, such multicollinearity can lead to spurious associations and can exaggerate effect sizes. Consider the following examples from Tollestrup et al. (1994). In cases with a confession, victims of robbery were most likely to identify the suspect (84.6%), followed by bystanders to robberies (55.5%) and witnesses of fraudulent activities (22.7%). These authors argued that victims of robberies were likely to experience the highest arousal levels in the sample, while witnesses of fraud were likely to experience the lowest arousal levels. However, Tollestrup et al. noted that witness type was confounded with retention interval within their sample. (The average retention intervals for victims of robberies, bystanders to robberies, and witnesses to fraudulent activities were 31 days, 44 days, and 108 days, respectively.) Furthermore, retention interval was strongly associated with suspect identifications for both types of crimes. It is therefore impossible to determine whether the nature of the crime or the retention interval (or some other variable) was responsible for the association with suspect identifications.

Tollestrup et al. (1994) also examined associations between weapon presence and identifications (for robbery cases only). They found witnesses to weapon-absent crimes identified the suspect in 73.3% of lineups, while witnesses to weapon-present crimes identified the suspect in only 30.6% of lineups, seemingly providing support for the “weapon focus effect” (Fawcett et al., 2013; E. F. Loftus et al., 1987). However, once again, Tollestrup et al. noted a confound with retention interval. The average retention interval in weapon-absent crimes was 7 days compared to 41 days for weapon-present crimes. When retention interval was covaried out, weapon presence was no longer significantly associated with suspect identifications.

Regression techniques can be used to partial out the effects of other variables. In a regression model that includes all of the predictors, we can see which predictors independently explain significant portions of the variance in the outcome variables in the context of the whole dataset. Regression techniques have been used in some previous archival studies (e.g., Horry et al., 2012; Valentine et al., 2003; Wright & McDaid, 1996) while others have taken a piecemeal approach to their analyses, conducting independent chi-square tests for each predictor (e.g., Behrman & Davey, 2001; Memon et al., 2011; Tollestrup et al., 1994). It should be noted, of course, that regression techniques cannot solve problems of multicollinearity with 'hidden' variables. Some of these may relate to aspects of the crime that are not routinely included in witness statements, or that are difficult to capture. Others may relate to the way the investigation unfolds (which leads are chased and which are not; whether evidence is considered probative or not; which witnesses are selected to participate in lineups, and so on). In any archival study, the researcher decides which variables to record according to theoretical interest and the available information. There will always be a number of hidden variables that may be confounded with the predictors. Neither regression nor chi-square analyses can deal with this problem.

In addition to checking for multicollinearity, associations among predictors can be useful and interesting in their own right. By examining the structure of a dataset in detail, we can ask questions about how actual crimes typically differ from the staged crimes widely used in the laboratory. For example, are weapon-present crimes typically associated with a different range of viewing conditions than weapon-absent crimes? Importantly, such knowledge will help us when we wish to generalize from the laboratory to the field.

The second statistical issue that must be addressed in any archival study is the non-independence of data points. Actual criminal events often involve multiple witnesses, any number of whom may be asked to view a lineup (Skagerberg & Wright, 2008). Many actual crimes include multiple perpetrators, in which case the same witness may view more than one lineup (rather rarely studied in the laboratory, but see Kask & Bull, 2009). The data, therefore, are not independent. Rather, individual lineups should be thought of as nested within suspects, witnesses, and criminal cases (Wright & McDaid, 1996). Multilevel modelling (also known as mixed effects modelling, or hierarchical linear modelling) is well-suited for such data, as grouping variables (witness, suspect, crime) can be included as random effects. We take the multilevel approach here, which was advocated by Wright and McDaid and used in some subsequent archival studies (Horry et al., 2012; Valentine et al., 2003). The majority of field studies, however, have used more traditional statistical techniques that rely on assumptions of independence, even though this assumption is clearly violated by the data (e.g., Berhman & Davey, 2001; Memon et al., 2011; Tollestrup et al., 1994).

In this paper, we addressed both of the issues above by using multilevel regression analyses. All predictor variables were entered in a global model so that independent contributions of each predictor could be assessed in the context of the whole dataset. Individual lineups were nested at the level of the criminal event. In addition, we explored all pairwise associations between predictors. This will aid our understanding of why some predictors may or may not produce statistically significant associations, and provide valuable information about some of the ways in which real criminal events differ from laboratory events.

Small or non-representative samples. In the laboratory, the importance of sampling is taken for granted. Researchers appreciate that larger samples allow them to draw stronger conclusions and reduce the risk of Type 2 errors. They also appreciate that, if they wish to generalize from a sample to a population, the sample must be randomly sampled from that population. Any systematic bias in sampling will undermine the generalizability of the results. Both of these considerations are just as important for archival data as laboratory data. Because field data are being analyzed does not mean that researchers can afford to dismiss these fundamentals of good science (Wright, Memon, Dalton, Milne, & Horry, 2013).

British archival studies have generally included large numbers of lineups (in the range of 600 to over 1000) sampled representatively from a given police force or forces in a specified period of time (e.g., Horry et al., 2012; Memon et al., 2011; Pike et al., 2002; Valentine et al., 2003; Wright & McDaid, 1996). North American studies, however, have included much smaller samples (170 identification attempts in Tollestrup et al., 1994; 284 photographic lineups in Behrman & Davey, 2001). In addition, Behrman and Davey did not use a representative sample. In almost half of the cases in their sample, the first author acted as a consultant, suggesting that there was something anomalous about the case or about the identification procedure (Wright et al., 2013). This may explain the extremely high suspect identification rates in some conditions. Small or non-representative samples should be treated with extreme caution as results are less likely to be reliable or generalizable.

In this paper, we included a large number of identification attempts sampled from a single but relatively large police force in England. The area covered by the police force has a total population of around 1.25 million, and is largely rural. However, the area also contains two cities and other large towns, producing wide variations in socio-economic status. We note, however, that the region is not ethnically diverse, with 98% of the population self-identifying as being of European descent (compared to 91% for England and Wales as a whole at the time of the data collection). Thus, though our sample can be considered to be representative of the population of cases in the region, care must be taken when generalizing the results to areas with very different demographics (for example, areas with much higher ethnic diversity).

Relying on police files stored in archives, all cases in which a lineup was conducted were coded for all relevant predictor variables. After excluding cases with some prior familiarity between the witness and perpetrator (see below), we were left with 833 lineups that should be representative of all lineups conducted within the region.

Prior familiarity between the witness and perpetrator. Laboratory studies almost always focus on a situation in which the perpetrator is completely unfamiliar to the witness. Memory for the target therefore becomes memory for the 'once-seen face' (Deffenbacher, Bornstein, McGorty, & Penrod, 2008) and pre-experiment familiarity is controlled. However, in archival cases, there may be some prior familiarity between the witness and the perpetrator. For example, the perpetrator may live locally, and the witness may report seeing the perpetrator on several prior occasions. While some archival studies have addressed this issue, restricting analyses only to cases in which the perpetrator was unknown to the witness (Horry et al., 2012; Memon et al., 2011; Valentine et al., 2003), others have failed to record such information (Behrman & Davey, 2001; Pike et al., 2002; Tollestrup et al., 1994; Wright & McDaid, 1996). Where prior familiarity has been recorded, the proportion of cases in which the perpetrator was known to the witness has varied from 10% (Valentine et al.) to 40% (Memon et al.). Thus, familiar-perpetrator cases can form a fairly substantial minority of the overall dataset.

Including familiar-perpetrator cases is problematic because it may increase the risk of a Type 2 error and provide an unrealistically high estimate of witness accuracy. For example, Memon et al. (2011) reported that suspect identification rates dropped from 92.5% when the witness was familiar with the suspect to 43.6% when the witness was unfamiliar with the suspect. The risk of a Type 2 error is increased because we would expect all estimator variables to have a smaller effect in any cases where the witness' memory of the perpetrator is relatively strong due to pre-existing familiarity. Thus, familiar-perpetrator cases may reduce the size of associations between estimator variables and lineup identifications. Here, we follow the lead of Valentine et al. (2003) in excluding from the analyses all cases in which the witness reported some prior familiarity with the perpetrator.

Summary and overview of dataset

Prior archival studies have provided valuable information on eyewitness accuracy in real cases. However, results across studies have been inconsistent, with opposite associations even found for some variables (e.g., role of witness; ethnicity of suspect). Some of these inconsistencies may be due to methodological problems such as an absence of ground truth (Horry et al., 2012; Memon et al., 2011; Pike et al., 2002; Valentine et al., 2003; Wright & McDaid, 1996), use of inappropriate statistical methods (Behrman & Davey, 2001; Memon et al., Tollestrup et al., 1994), small or non-representative samples (Behrman & Davey; Tollestrup et al.), and inclusion of familiar-perpetrator crimes (Behrman & Davey; Pike et al.; Tollestrup et al.; Wright & McDaid). Here, we addressed each of these problems by i) coding the lineups for corroborating evidence, analysing separately corroborated and uncorroborated cases; ii) using multilevel regression to account for multicollinearity among predictors and non-independence of data; iii) sampling all lineups conducted in a given time period by a large police force; and iv) restricting our analyses to cases in which the perpetrator was unfamiliar to the witness. In the Discussion, we consider whether there are additional, unfixable, problems inherent to archival research, and what such problems mean for interpreting the results of archival studies.

We included ten estimator variables in our regression model. These included type of criminal offence, witness characteristics (sex and age), likely quality of encoding (exposure duration, distance, number of perpetrators), retention interval, and factors that may well relate to arousal (use of violence, presence of a weapon, and involvement of witness – bystander vs. victim). Additional variables were considered (e.g., suspect age, suspect sex, ethnicity of witness and suspect) but were excluded because the sample was too homogeneous (i.e., most suspects were young males of European descent and most witnesses were also of European descent).

Method

Sample

The data were collected from the archives of a single relatively large police force in England over a period of one year from 2001-2002 and included cases from 1992 to 2000. Over 240,000 case files were examined by the first author (then a senior police officer who had privileged and one-off access to the files) and cases were included if they met the following criteria: i) the case file included at least one lineup identification; and ii) the witness had no prior familiarity with the suspect. Two hundred and ninety five case files met the above criteria. These included 833 lineups seen by 709 unique witnesses. These lineups were predominantly live lineups, which have since been replaced by video lineups in the UK. It is important to note that witnesses were only ever given one opportunity to identify a suspect. Where case files contained multiple identification attempts, it was because the case included multiple witnesses and/or multiple suspects.

The collection of eyewitness evidence in the UK is governed by a mandatory set of guidelines outlined in the Police and Criminal Evidence (PACE) Act of 1984. The particular police force sampled in this study had produced a booklet based on the guidelines, to ensure uniformity in the conduct of the lineup. The booklet included instructions to the witness, intended to be read verbatim (including an instruction that the offender “may or may not be present”). The officers were also required to write down, verbatim, any response made by the witness upon viewing the lineup.

Under the PACE Act, lineups must include eight foils in addition to the suspect, who must “so far as possible, resemble the suspect in age, height, general appearance and position in life" (PACE Code D, Annex B). Though this particular guideline has been criticized for being somewhat vague (Horry, Memon, Milne, Wright, & Dalton, in press), it should, at least, ensure that the lineups were reasonably fair (a conclusion drawn by Valentine and Heaton, 1999, in their evaluative survey of British lineups). We are unable to verify the proportion of cases in which the construction and conduct of lineups had been followed.

The majority of the cases (70.59%) included corroborating evidence. The most common corroborating type of evidence was very strong circumstantial evidence (70.41%). An example is a case in which a suspect was found hiding in a bush outside of an office block that had been broken into. The suspect matched the description of the perpetrator provided by a security guard, both in terms of appearance and clothing. Circumstantial evidence alone was available in 170 cases (28.91% of all corroborated cases). The next most common types of corroborating evidence were guilty pleas (47.96%) and possession of stolen property (25.34%). Other indicators of guilt included confessions (10.71%), video evidence (10.37%), fingerprints (2.57%), evidence against a co-accused (4.59%), bodily samples including DNA (2.55%), and other forensic evidence (2.72%). Note that these figures sum to more than 100% because many of the cases included more than one type of corroborating evidence against the suspect. In fact, two types of corroborating evidence were available in 36.56% of the cases, three types of evidence were available in 20.41% of the cases, and four types of evidence were available in a small number of cases (0.85%).

An important caveat is that we were unable to record accurate information about the temporal order in which the corroborating evidence was collected. For most categories of evidence, we are therefore unable to determine whether the evidence was obtained before or after the identification procedure. The two major exceptions to this are circumstantial evidence, which was always related to the circumstances of the arrest (and therefore always preceded the lineup), and guilty pleas/confessions, which were always obtained after the lineup (though 94% of guilty pleas and confessions were accompanied by at least one other category of corroborating evidence). We consider this issue at more length in the Discussion.

Coding of variables

The second author recorded information on a number of forensically relevant variables from the witness statements included in each case file. Police officers in the UK are required to record information on several key factors when taking witness statements, including exposure duration, distance, and any prior familiarity with the perpetrator. These statements are taken shortly after the crime is reported, so they are likely to precede the lineup identification (although, again, this order of events cannot be verified). However, in some cases, the relevant data were missing. In these cases, the second author provided a best estimate based on the known details of the crime. For example, a crime in which the witness was physically assaulted was likely to involve face-to-face contact with the assailant. Alternatively, if a witness viewed a crime from across the street, the viewing distance was likely to be more than five meters.

Note there is clearly margin for error in witness statements, as witnesses may not be able to accurately estimate factors such as exposure duration and distance retrospectively (Lindsay, Semmler, Weber, Brewer, & Lindsay, 2008; Wells & Quinlivan, 2009). For this reason, we coded the data using relatively broad categories, rather than attempting to create fine-grained categories or treating variables as continuous. We extracted information for each lineup on the following factors:

1) Type of crime: crimes of dishonesty: (*N* = 281, 33.73%); assault (*N* = 266, 31.93%); robbery (*N* = 137, 16.45%); rape/indecent assault (*N* = 103, 12.36%); other crimes (*N* = 46, 5.52%).
2) Sex of the witness: male (*N* = 465, 55.82%); female (*N* = 368, 44.18%).

3) Age of the witness: 10-15 years old (*N* = 95, 11.40%); 16-20 years old (*N* = 142, 17.05%); 21-50 years old (*N* = 496, 59.54%); over 50 years old (*N* = 100, 12.00%).
4) Exposure duration: less than 60 seconds (*N* = 249, 29.89%); more than 60 seconds (*N* = 584, 70.11%).
5) Distance between witness and perpetrator: face-to-face contact (*N* = 466, 55.94%); 0 – 5 meters (*N* = 228, 27.37%); more than 5 meters (*N* = 139, 16.69%).
6) Number of perpetrators: single perpetrator (*N* = 321, 38.54%); multiple perpetrators (*N* = 512, 61.46%).
7) Delay between the crime and the lineup: less than a month (*N* = 183, 21.97%); one to three months (*N* = 460, 55.22%); more than three months (*N* = 190, 22.81%).
8) Use of violence: violence used (*N* = 209, 25.09%); violence not used (*N* = 624, 74.91%).
9) Presence of a weapon: weapon present (*N* = 119, 14.29%); weapon absent (*N* = 714, 85.71%).
10) Role of witness: victim (*N* = 386, 46.34%); bystander (*N* = 447, 53.66%).

Results

Relationships among predictors

Before beginning our main analyses, we explored the structure of our dataset. We calculated Cramer's *V* for each pair of variables in the dataset (see Table 1). Our first question concerned whether corroborated and uncorroborated cases differed in their characteristics. The top row of Table 1 shows that corroboration was significantly associated with several of our predictor variables. Robberies were the most likely crimes to be corroborated (91% corroborated), while assaults were the least likely crimes to be corroborated (61% corroboration). Corroborated crimes were more likely to involve longer exposure durations, shorter retention intervals, and face-to-face contact between the witness and the perpetrator than uncorroborated crimes. Generally, then, corroborated crimes were associated with more optimal encoding and retrieval conditions than uncorroborated crimes. We can only speculate on why such associations were found. It may be that witnesses who got a good view of the suspect provided more useful information to the police at the scene, increasing the possibility that a suspect would be apprehended and corroborating evidence collected. For example, following a street robbery, an accurate and detailed description from a witness might help investigators to locate a suspect rapidly, while he is likely to still be in possession of any stolen property. Alternatively, it may be that optimal viewing conditions lead to suspect identifications, which in turn fuel a search for corroborating evidence, creating a kind of reverse causation between lineup outcomes and corroborating evidence.

As Table 1 shows, many of the other predictor variables were inter-correlated. For example, comparing crimes in which violence was used with crimes in which violence was not used, we note the following differences: Witnesses to violent crimes were more likely to be young (under 20) and male; witnesses to violent crimes were almost always victims rather than bystanders; violent crimes almost always involved face-to-face contact between the witness and perpetrator; and violent crimes were more likely to involve weapons and multiple perpetrators. Victim witnesses were more likely than bystanders to be young (under 20); they were more likely to be in close proximity to the perpetrator during the crime; and they were more likely to be exposed to a weapon. Exposure duration was also related to several other factors. Crimes with longer exposure durations were more likely to involve shorter viewing distances and multiple perpetrators than crimes with short exposure durations.

Identification outcomes

We analyzed the data using multilevel logistic regression, which is ideal for datasets that violate assumptions of independence (see Wright & London, 2009). Separate regression models were created for suspect identifications, foil identifications, and non-identifications.

For suspect identifications, the outcome variable was whether or not the witness identified the suspect; for foil identifications, the outcome variable was whether or not the witness identified a foil; for non-identifications, the outcome variables was whether or not the witness identified no-one. Because the same suspect was sometimes viewed by more than one witness, we nested lineups within criminal cases.

We created a model that contained all main effects, which served as our baseline model. To assess the independent contribution of each of the predictors, we then removed one predictor from the model and compared the resulting model to the baseline. We tested the change in the goodness-of-fit using chi-squared tests. A significant chi-square test indicates predictor was significantly contributing to the model fit. For every chi-square test with one degree of freedom, the natural log-odds ratio (*ln*OR) is reported as a measure of effect size, with 95% confidence intervals. A *ln*OR of 0 indicates that the outcome was equally likely across the two conditions. For an additional measure of effect size, unstandardized coefficients with standard errors are reported in Tables 2 to 4.

Before beginning our main analyses, we tested whether lineup outcomes varied for corroborated and uncorroborated crimes. Corroboration was significantly associated with suspect identification rates, $\chi^2(1, N= 833) = 17.49$, $p < .001$, *ln*OR = 0.97 (95% CI = 0.67, 1.31), and with non-identification rates, $\chi^2(1, N= 833) = 10.19$, $p < .001$, *ln*OR = 0.69 (95% CI = 0.39, 0.99). Suspects were more likely to be identified if there was corroborating evidence of guilt (52.72%) than if there was no corroborating evidence of guilt (29.39%). Non-identifications was more common if there was no corroborating evidence of guilt (47.76%) than if there was corroborating evidence of guilt (31.46%). These results are in line with what we would expect if the corroborated cases were more likely to be target-present than the uncorroborated cases. However, a similar result could also be obtained if the presence of corroborating evidence caused investigators to behave in ways that increased the likelihood of securing suspect identifications (possibly through such routes as foil selection and verbal/non-verbal communication with the witness). Note that the suspect identification rate for the uncorroborated crimes is much higher than chance (11% for nine-person lineups, the minimum permitted size in the UK). This suggests that: 1) A reasonable proportion of the uncorroborated crimes were target-present; and/or 2) the lineups were biased towards the police suspect. Foil identification rates did not significantly vary with corroboration (15.82% for corroborated cases, 22.86% for uncorroborated cases), $\chi^2(1, N= 833) = 2.08$, $p = .15$.

For our main analyses, we created separate regression models for corroborated cases and for uncorroborated cases. Tables 2, 3, and 4 show the baseline (global) regression models for suspect identifications, foil identifications, and non-identifications, respectively.

Exposure duration was significantly associated with suspect identification rates for both the corroborated cases $\chi^2(1, N= 588) = 11.63$, $p < .001$, *ln*OR = 0.77 (95% CI = 0.39, 1.14), and the uncorroborated cases, $\chi^2(1, N= 245) = 3.89$, $p = .049$, *ln*OR = 0.53 (95% CI = -0.06, 1.09). For both subsets of cases, the association was positive; longer exposure durations (57.53% for corroborated cases; 33.56% for uncorroborated cases) were associated with higher suspect identification rates than shorter exposure durations (38.67% for corroborated cases; 23.23% for uncorroborated cases). Unsurprisingly, therefore, non-identifications were also significantly associated with exposure duration, though only for the corroborated cases, $\chi^2(1, N= 588) = 19.14$, $p < .001$, *ln*OR = 1.02 (95% CI = 0.63, 1.40). Non-identifications were more common if the exposure duration was less than 60 seconds (48.67%) than if the exposure duration was more than 60 seconds (25.57%). These results suggest that longer exposure durations were generally beneficial, creating stronger memory traces for the perpetrator's face.

Viewing distance was also significantly associated with suspect identifications, but only for the corroborated cases, $\chi^2(2, N= 588) = 17.74$, $p < .001$. Suspects were less likely to be identified if the perpetrator was seen from a distance of greater than five meters (25.93%) than if seen in a face-to-face situation (59.77%; *ln*OR = -1.43 (95% CI = -1.90, -0.97) or from a distance of less than five meters (55.91%; *ln*OR = -1.28, 95% CI = -1.83, -0.73). These results suggest that closer viewing distances allow for more optimal encoding of the perpetrator's face, creating a more robust memorial representation.

For the uncorroborated cases, witness age was significantly associated with suspect identification rates, $\chi^2(3, N = 245) = 8.40$, $p = .04$. Child witnesses (10-15 years old) were less likely to identify the suspect (16.22%) than the 16-20 year olds (43.75%; *ln*OR = -1.56, 95% CI = -2.66, -0.43). For the corroborated cases, age of the witness was not significantly associated with either suspect identifications or non-identifications

Though the non-identification rates usually mirrored the suspect identification rates, there were some exceptions. The number of perpetrators was significantly associated with non-identification rates for the corroborated cases, $\chi^2(1, N= 588) = 7.18$, $p = .007$, *ln*OR = 0.59 (95% CI = 0.22, 0.96), though not for the uncorroborated cases. Multiple-perpetrator crimes were associated with higher non-identification rates than single-perpetrator crimes (36.07% vs. 23.87% for corroborated lineups). These results suggest that dividing attention across multiple targets at encoding reduces the strength of the memory traces of the perpetrators' faces. Retention interval was not significantly associated with non-identification rates for either corroborated or uncorroborated crimes.

For foil identifications there was a significant association with weapon presence for the corroborated cases, $\chi^2(1, N= 588) = 5.23$, $p = .02$, *ln*OR = -0.62 (95% CI = -1.35, 0.11). Weapon-present crimes were associated with fewer foil identifications (9.68%) than weapon-absent crimes (16.56%). But for these cases, the association between weapon presence and suspect identifications was not significant. For the uncorroborated cases, the the role of the witness was significantly associated with foil identifications, $\chi^2(1, N= 245) = 3.92$, $p = .048$, *ln*OR = 0.37 (95% CI = -0.24, 0.97). Bystanders were more likely to identify foils (25.98%) than were victims (19.49%). Use of violence was also significantly associated with foil identifications, $\chi^2(1, N= 245) = 6.23$, $p = .01$, *ln*OR = -0.05 (95% CI = -0.71, 0.61). The direction of the association indicated that violent crimes were associated with fewer foil identifications than non-violent crimes. However, the association is not clearly apparent in the descriptive statistics, with similar foil identification rates for violent (22.22%) and non-violent (23.12%) crimes. Thus, this association only becomes apparent when the effects of one or more of the predictor variables are partialled out. Indeed, if violence is entered as the first predictor in the model, it does not significantly improve the model fit, $\chi^2(1, N= 245) = 0.04$, $p = .85$. Note that for all of the significant associations with foil identifications, the confidence intervals around the effect sizes included zero, indicating that these effects are, at best, small.

Discussion

We have argued that methodological and statistical flaws have created inconsistencies among previous archival studies, making results difficult to interpret. We presented a large archival dataset in which we attempted to address these flaws. Below, we consider the predictor variables that were most strongly associated with lineup outcomes: exposure duration, distance, witness age and the number of perpetrators. We then consider, in detail, the general limitations inherent to all archival studies. Finally, we ask the question of what, if anything, archival studies can tell us about eyewitness memory.

In the current study, suspects were more likely to be identified, and non-identifications were more common, if the exposure duration exceeded 60 seconds than if the exposure duration was less than 60 seconds. Though few experimental studies have systematically manipulated exposure duration using eyewitness identification tasks (see Memon, Hope, & Bull, 2003; Palmer, Brewer, Weber, & Nagesh, 2013), the recognition memory literature provides a firm basis for predicting that longer exposure durations should increase recognition accuracy (e.g., Laughery, Alexander, & Lane, 1971; Reynolds & Pezdek, 1992; Shapiro & Penrod, 1986). So intuitive is this hypothesis that exposure duration is one of the factors relied on by courts both in the UK and in the US to decide whether a witness is likely to be reliable (*State v. Larry R. Henderson,* 2011; *R v Turnbull,* 1976). Our results are in line with Valentine et al.'s (2003) archival study, as longer exposure durations were associated with higher suspect identification rates. To our knowledge, no other archival studies have included exposure duration as a predictor variable.

Our results, in conjunction with similar results from the literature, suggest that exposure duration may be a valid criterion for courts to use when assessing witness reliability. However, relying on witness estimates of exposure duration can be problematic, as people often overestimate the duration of events (E. F. Loftus, Schooler, Boone, & Kline, 1987; Pedersen & Wright, 2002; Yarmey & Yarmey, 1997). Duration estimates may become even less reliable if they are made following positive post-identification feedback (Wells & Quinlivan, 2009; Wright & Skagerberg, 2007). Thus, duration should only be relied upon when it can be objectively verified.

The present study also found that, for corroborated cases, suspects were more likely to be identified if the perpetrator had been seen at a distance of less than 5 meters than if they had been seen from a distance of greater than 5 meters. Laboratory work on face perception suggests that our ability to recognize *familiar* faces drops sharply after distances of about 7.5 meters, approaching zero by around 33.5 meters (G. R. Loftus & Harley, 2005). Furthermore, McKone (2009) found that the ability to extract holistic information from faces (believed to underpin face recognition; Maurer, Le Grand, & Mondloch, 2002) actually peaks between 2 and 10 meters, declining sharply at distances of less than 2 meters and more slowly at distances of greater than 10 meters. The range of distances in our long distance category was wide, with distance exceeding 20 meters in some cases. At such long distances, accuracy for a once-seen face is likely to be very poor (Lindsay et al., 2008). Our results suggest that distance may be a valid indicator of witness reliability. However, subjective estimates of distance from witnesses are unreliable, and the reliability decreases with increased distances and retention interval (Lindsay et al., 2008). Thus, distance may only serve as a reliable indicator of witness reliability when the distance can be objectively verified.

Though it is often assumed that factors such as exposure duration and distance will have large impacts on eyewitness identification accuracy, direct tests of these assumptions within the eyewitness literature are scarce. Furthermore, only one other archival study has included these factors (Valentine et al., 2003). Currently, we have little knowledge about the shape of the functions linking many factors (such as exposure duration and distance) to eyewitness memory. We know very little about the size of these effects and whether they manifest more strongly for correct identifications or for false identifications. It is very important, therefore, to resist the temptation to search for convenient cut-off points above which a witness would be considered reliable and below which a witness would be considered unreliable. For example, a ‘rule of thumb’ was suggested by Wagenaar and van der Schrier (1996), who proposed that distances of greater than 15 meters should be considered as too far to support an eyewitness identification beyond reasonable doubt. However, Lindsay et al. (2008) argued that the arbitrary nature of this cut-off point renders it invalid and potentially dangerous. In a field experiment of eyewitness identification, they noted that increasing distance was associated with decreased correct identification rates both within and outside of the 15 meter zone. Furthermore, the rate of decline was similar above and below 15 meters, suggesting that there is no clear point at which eyewitness identification drops catastrophically. Thus, while we may be able to conclude that witness reliability generally increases as exposure duration increases and distance increases, we make no attempt here to identify specific points at which reliability becomes too low for an identification decision to have probative value.

Witness age was associated with suspect identifications, but only for uncorroborated cases. This association was driven by the child witnesses (under 16) choosing the suspect less frequently than 16-20 year olds. The uncorroborated cases probably contain a higher proportion of target-absent lineups than the corroborated cases – but they are unlikely to all be target-absent. Interpreting these results is therefore somewhat complicated. Perhaps the child witnesses were missing guilty perpetrators when they were present, or perhaps they were being more conservative in their choosing, and were therefore resisting choosing innocent suspects. Either way, these results do not map neatly onto laboratory findings. In the laboratory, children have been found to be less accurate than adults at eyewitness identification tests (e.g., Keast, Brewer, & Wells, 2007; Pezdek, Blandon-Gitlin, & Moore, 2003). However, children usually adopt more lenient decision criteria than adults, making them prone to 'guessing' from target-absent lineups (Parker & Carranza, 1989; Parker & Ryan, 1993; Pozzulo & Lindsay, 1998). If this were the case in our sample, we might expect to see *increased* suspect identification rates for uncorroborated crimes, possibly accompanied by increased foil identification rates. Our results are thus difficult to reconcile with the patterns of behaviour observed in the laboratory.

The presence of multiple perpetrators at a crime was associated with higher non-identification rates than the presence of just a single perpetrator. Divided attention has been shown to reduce recollection of unfamiliar faces in old/new recognition tasks (e.g., Palmer, Brewer, & Horry, 2013; Parkin, Gardiner, & Rosser, 1995; Reinitz, Morrissey, & Demb, 1994), and to reduce correct suspect identification rates in an eyewitness identification task (Palmer, Brewer, McKinnon, & Weber, 2010). Witnesses in multiple-perpetrator cases would have had to divide their spotlight of attention across more than one unfamiliar face, which seemingly reduced their willingness to choose. However, we should note that the increase in non-identification rates under divided attention was not accompanied by a significant drop in suspect identification rates, as would be expected if witnesses simply shifted their decision criterion.

To sum up, we found that suspect identifications were associated with factors related to the quality of the encoding episode (exposure duration, distance), as well as the age of the witness. Non-identifications were additionally associated with the number of perpetrators. But what do these results mean? In the following section, we consider the general limitations of archival studies. In light of these limitations, we then ask what archival studies are able to tell us about eyewitness identification.

Limitations of archival studies

No matter how representative the sample, or how sophisticated the analysis, there are some unavoidable limitations with archival studies. These include hidden variables stemming from police behaviour leading up to and during the lineup, complex inter-correlations among predictors, use of arbitrary categorizations on continuous variables, a lack of ground truth, and selection effects stemming from police investigations. We consider each of these in turn below.

First, it is simply not possible in any archival study to code for every way in which the behaviour of investigators could potentially influence the outcome of the lineup. Consider, for example, the multiple biases potentially associated with non-blinded lineups. These biases include the way in which foils are selected, the verbal and non-verbal communication between the investigator and the witness (see Wright, Carlucci, Evans, & Schreiber-Compo, 2010, for a discussion of these cues), and the way in which responses are recorded (see Wells, 2008, for a discussion of how foil identifications are sometimes recorded as non-identifications). These biases would be almost impossible to record in any archival study, yet they would clearly exert a sizeable influence over the outcomes of the lineups.

In this particular sample, lineups were not blind (as blind administration is not required under UK guidelines), leaving opportunities for investigators to influence lineup outcomes. However, we should acknowledge that the police force had taken efforts to minimize some of these sources of bias. All instructions were to be read verbatim from a manual, and all comments made by the witness were required to be written verbatim. If properly adhered to, this should have reduced the opportunity for overt verbal cues to the witness, as well as the opportunity for selective recording of identification outcomes. However, the instructions would have done little to reduce potential biases in lineup construction, nor would they have eliminated non-verbal cues during the lineup (Horry et al., in press).

The biases described above are just a few specific examples of the many ways in which police investigators could influence witnesses. No archival study could ever hope to code for all of them. Furthermore, these behaviours may be intercorrelated with other variables. For example, in cases where there is corroborating evidence of guilt, police may behave more strongly in ways that bias the witness's decision. Thus, the influence of these biases is likely to be complex, and virtually impossible to capture in any single study.

Correlations between predictors form another major problem for archival researchers. We explored some two-way associations between our predictors, showing, for example, that violent crimes were more likely to involve weapons, multiple perpetrators, and face-to-face contact between the witness and perpetrator than non-violent crimes. The chances of detecting a significant association between a predictor and a lineup outcome will depend to a large degree on how that variable is associated with other predictors. For example, in our sample, highly arousing crimes (i.e., those involving violence and weapons) were generally associated with more optimal encoding conditions (shorter viewing distances, longer exposure durations). The benefits afforded by the encoding conditions may have off-set the negative effects of arousal that we would expect based upon psychological theory (Deffenbacher et al., 2004) and based on experimental findings (e.g., Morgan et al., 2004), potentially explaining why the associations with arousal factors so weak in this study. With so many predictor variables (as well as unmeasured, ‘hidden’ variables), the complexity of inter-correlations in archival datasets is difficult to capture. We explored two-way associations, but there are also likely to be higher-order associations. These associations will be running in many directions, sometimes compounding effects and at other times obscuring them. No matter how sophisticated the analysis, it will never be possible to eliminate these effects in any archival study. We briefly note, however, that there are statistical methods in widespread usage in fields such as economics and epidemiology that might allow researchers to address some of these limitations (e.g,, the method of instrumental variables; see, for example, Angrist & Kreuger, 2001).

Many estimator variables are continuous in nature, yet archival researchers often have to convert them into categorical variables. Witness estimates of factors such as exposure duration and distance are often lacking the precision required for continuous scaling. Even if they could be scaled, there would be sufficient reason to doubt their veracity, as people are quite poor at estimating such factors (see Lindsay et al., 2008; E. F. Loftus et al., 1987). Thus, archival researchers are often faced with ballpark estimates that must be treated in a categorical manner. However, categorizing continuous variables reduces statistical power and can distort relationships between inter-correlated variables (Irwin & McClelland, 2003). Furthermore, the results are obviously influenced to a large extent by the placement of the categorical boundaries. Ideally, theory would be used to guide the placement of the boundaries. In practice, however, practical concerns such as ensuring adequate cell sizes often make the placement much more arbitrary. As a consequence, different categories are used in different studies, making it difficult to compare results.

In theory, as sample size increases, the number of categories that could be used for each variable increases, allowing for more fine-grained distinctions. However, no matter how large the sample, there are some variables for which the crucial part of the range is likely to be missed, such as retention interval. There is strong empirical evidence that memory degrades rapidly after an event and that the rate of decay slows as more times passes (Wickelgren, 1972, 1974). Deffenbacher et al. (2008) estimated that approximately 15% of the original memory trace is lost within ten minutes, and that by one week, only around 50% of the memory trace remains. Thus, the greatest chance of detecting a retention interval effect would be to compare lineups conducted within minutes or hours of the crime with lineups conducted after longer retention intervals. While this is possible to do in an experiment (e.g., Sauer, Brewer, Zweck, & Weber, 2010), archival studies are constrained by the speed at which police investigations move. In our sample, just five lineups (0.6% of the entire sample) were conducted within 48 hours of the crime. Even with a ten-fold increase in sample size, we would still likely have too few cases at extremely short retention intervals to make meaningful comparisons. Though new technologies have substantially shortened the average retention interval in the UK (Kemp, Pike, & Brace, 2001), there are still likely to be days or weeks between a crime and a lineup. Thus, the effects of retention interval on eyewitness identification accuracy have likely been greatly underestimated in archival studies.

When trying to establish what actually happened in a criminal event, witness reports must often be relied upon. However, witness reports can be unreliable, with errors stemming from two sources. First, witnesses have difficulty estimating factors such as distance and time reliably (Lindsay et al., 2008; E. F. Loftus et al., 1987). Second, information that the witness receives over the course of the investigation can distort these already unreliable estimates.

Wright and Skagerberg (2007), for example, found that witnesses who were told that they had identified the suspect from a police lineup estimated that they had seen the culprit for longer and that they had had a better view than witnesses who were told that they had identified a foil. In archival studies, it would be ideal for the details of the crime to be taken from objectively verifiable sources (for example, known details of the crime scene, surveillance footage). If objective sources are not available and witness reports must be used instead, the reports should be those taken at the earliest stages of the investigation. We took the details of the crime from witness statements, all of which should have been taken shortly after the crime was reported in accordance with the Turnbull rulings (*R v. Turnbull and others*, 1976). Thus, while there is scope for some unreliability in the estimates, they should not have been distorted by feedback from the identification procedure or any other part of the investigation. Archival researchers should be very wary of witness descriptions of the event that were taken a long time after the event, especially if they followed an identification attempt.

An additional ground truth problem in archival studies is that we do not know the identity of the perpetrator. Any archival sample likely contains some guilty suspects and some innocent suspects, but we do not know which are which. There is no clear way of dealing with this issue in archival studies. Most researchers have largely ignored the issue, treating it as one of unpredictable noise in the data. We took the approach of relying on corroborating evidence to separate out cases in which the suspect was more likely to be guilty from cases in which the suspect was less likely to be guilty. A similar approach has been used by other researchers (Behrman and Davey, 2001; Tollestrup et al., 1994). The logic of this approach is as follows. Corroborating evidence directly increases the likelihood that the suspect is guilty; suspect guilt directly increases the likelihood that the suspect will be identified; therefore, corroborating evidence can be used as a proxy for suspect guilt.

Crucially, however, corroborating evidence can act as a proxy for suspect guilt *only if* the following conditions are met: 1) the only influence of corroborating evidence on lineup outcomes is indirect, via suspect guilt; and 2) lineup outcomes exert no influence on corroborating evidence. However, research on investigative decision-making suggests that these assumptions are almost certainly violated. Corroborating evidence could exert a *direct* influence over lineup outcomes by influencing police behaviour (a particular problem with non-blind lineups, as previously discussed). Furthermore, lineup outcomes can directly influence the likelihood of further corroborating evidence being obtained (Hasel & Kassin, 2009). This could create a loop of *corroboration inflation* (Kassin, 2012), in which corroborating evidence influences lineup outcomes, and lineup outcomes fuel the search for further corroborating evidence. Any archival study that wished to eliminate this problem would therefore need to very carefully code when each piece of evidence was collected. Even then, however, it would be difficult to remove any biasing effect of corroborating evidence on police behaviour before and during the lineup.

Of course, even if we were able to rule out the possibility of corroboration inflation, the presence of corroborating evidence does not guarantee that the suspect is guilty. The collection and interpretation of forensic evidence can be guided by the expectations of the forensic examiners (Kassin et al., 2013), and flawed forensic evidence has been present in many of the DNA exoneration cases documented by the Innocence Project (www.innocenceproject.org). Equally, the absence of corroborating evidence does not necessarily imply that the suspect is innocent. Certain types of crime may lend themselves more easily to the collection of physical evidence. For example, robberies provide an opportunity for the police to find stolen property in the possession of the suspect, and sexual assaults may leave behind DNA evidence. Other crimes may be much more reliant on witness reports, but that does not necessarily mean that the suspects in such cases are not guilty.

The final major limitation we wish to discuss is the selection effects that operate on archival data. There are many more witnesses to crimes than there are witnesses who view lineups. The police may choose not to conduct a lineup if a witness is perceived as unreliable, uncooperative, or hostile. Some witnesses may never make themselves known to police. And of course, there any many crimes for which a viable suspect is never found. It is impossible to know what proportion of all witnesses go on to view lineups. However, the proportion may be relatively small. Over 240,000 case files were initially examined for the current study, of which around 0.001% contained an identification attempt. Of course, not all of these cases would have included an eyewitness, so this figure is almost certainly an underestimation. Nevertheless, it seems likely that the vast majority of potential witnesses are lost to the archival researcher before the study even begins. In contrast, very little selection occurs in the laboratory. With the exception of some rare circumstances in which a researcher may choose to exclude a participant or in which a participant chooses to withdraw their participation, all witnesses who view the crime will view a lineup. Archival studies and laboratory studies therefore take very different samples from the universe of all possible witnesses.

Given these differences in sampling, can we meaningfully compare a laboratory population with an archival population? And is it realistic to expect results from the laboratory to generalize to an archival study? Unfortunately, the answer to these questions may well be “no”. Even if it is true that variable X affects memory, and we are able to observe this effect in the laboratory, we may not see the same effect operating in the population of witnesses who view police lineups. Consider retention interval as an example. With the more time that passes since a crime, investigators may become more selective in terms of the evidence required to place a suspect in a lineup, and in terms of the witnesses that they select to view a lineup. Witnesses may also self-select, declining to take part at longer retention intervals if they feel that they would be unable to identify the suspect. Thus, at longer retention intervals, the witnesses viewing the lineups may be those who had a relatively good look at the offender and who are reasonably confident in their ability to recognize the offender. For almost any variable that has been consistently associated with identification accuracy in the laboratory, one could probably generate a plausible hypothesis for how selection effects in archival studies might severely limit the chances of detecting an effect in the field.

It is likely that each of the limitations discussed above contributed to the inconsistent findings of the archival studies that have been conducted to date. Furthermore, none of them is easily addressed, suggesting that future archival studies will continue to produce inconsistent results. We now ask ourselves, what questions can archival studies actually answer?

What can we conclude from archival studies?

In this paper, we have identified some of the fixable problems that have been present in previous archival studies and we have attempted to address them. These problems include small or biased samples, lack of ground truth, inclusion of familiar-perpetrator cases, and inappropriate statistical methods. But should we consider the problems with archival studies now fixed? Should we discard the results of previous archival studies in favour of these new findings? No. For every problem that we were able to address, there were additional problems that became apparent and that do not seem to have an obvious solution. Not least of these is the problem of the different population of witnesses that are included in archival studies versus laboratory studies. So what questions can archival studies answer and what questions can archival studies *not* answer?

Archival studies are often considered to be tests of whether effects found in the laboratory generalize to the field. The generalizability of laboratory results to real witnesses has been questioned by some legal professionals due to the many differences between the experiences of a typical participant and a typical eyewitness (e.g., Egeth, 1993; Elliott, 1993). Given the ethical and pragmatic constraints on laboratory research studies, it is not possible to reconcile all of these differences between laboratory participants' experiences and those of real life witnesses (though some researchers have attempted to increase the realism of their experiments in various ways; see, for example, Hope, Lewinski, Dixon, Blocksbridge, & Gabbert, 2012; Hosch, Leippe, Marchioni, & Cooper, 1984; Morgan et al., 2004). But can archival studies really answer questions about generalizability from the lab to the field? We have argued above that archival studies and laboratory experiments sample from very different populations of witnesses. It is simply not possible to make sensible comparisons between them. Thus, we would argue that archival studies cannot, in fact, speak to the generalizability of results from the laboratory to the field.

Perhaps archival studies can tell us whether a witness in the courtroom is likely to be accurate? Unfortunately, this is probably not true either. Only a small percentage of all cases in which an identification is attempted will proceed to trial. Many suspects will plead guilty, while other suspects will not be prosecuted (through lack of evidence or because the crime is seen as trivial). Once again, there is a sampling discrepancy between witnesses who view lineups and those that testify at trial, with trial witnesses being a highly selected sample from a much, much wider universe of witnesses. Furthermore, to draw any conclusions about accuracy would require ground truth, which is almost never going to be established with certainty (and if someone could only sample those cases in which the ground truth was certain, that would be an even more selective population of cases). Well-designed and appropriately analyzed archival studies can possibly tell us something about the factors that are related to a specific population of cases: those in which the police choose to test an eyewitness and in which the identity of the perpetrator is unknown. Whether this knowledge is particularly valuable is unclear. Researchers should consider each of these limitations carefully before embarking on archival research, and readers should bear these limitations in mind when interpreting the results of any archival study.

References

Angrist, J. D., & Kreuger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15 (4),* 69-85.

Behrman, B. W. & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law & Human Behavior, 25*, 475-491. doi: 10.1023/A:1012840831846.

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied, 14*, 139-150. doi: 10.1037/1076-898X.14.2.139.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law & Human Behavior,* 28, 687-706. doi: 10.1007/s10979-004-0565-x.

Egeth, H. E. (1993). What do we *not* know about eyewitness identification? *American Psychologist*, *48*, 577-580. doi: 10.1037/0003-066X.48.5.553.

Elliott, R. (1993). Expert testimony about eyewitness identification: A critique. *Law and Human Behavior, 17,* 423-437. doi: 10.1007/BF01044376.

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2013). Of guns and geese: A meta-analytic review of the 'weapon focus' literature. *Psychology, Crime, & Law, 19,* 35-66. doi: 10.1080/1068316X.2011.599325.

Hasel, L. E. & Kassin, S. M. (2009). On the presumption of evidentiary independence: Can confessions corrupt eyewitness identifications? *Psychological Science, 20,* 122-126. doi: 10.1111/j.1467-9280.2008.02262.x.

Hope, L., Lewinski, W., Dixon, J., Blocksidge, D., & Gabbert, F. (2012). Witnesses in action: The effect of physical exertion on recall and recognition. *Psychological Science, 23,* 386-390. doi: 10.1177/0956797611431463.

Horry, R., Memon, A., Milne, R., Wright, D. B., & Dalton, G. (in press). Video identification of suspects: A discussion of current practice and policy in the United Kingdom. *Policing: A Journal of Policy and Practice.* doi: 10.1093/police/pat008.

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law & Human Behavior, 36,* 257-265. doi: 10.1007/s10979-011-9279-z.

Hosch, H. M., Leippe, M. R., Marchioni, P. M., & Cooper, D. S. (1984). Victimization, self-monitoring, and eyewitness identification. *Journal of Applied Psychology, 69,* 280-288. doi: 10.1037//0021-9010.69.2.280.

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research, 40,* 366-371. doi: 10.1509/jmkr.10.3.366.19237.

Kask, K., & Bull, R. (2009). The effects of different presentation methods on multi-ethnicity face recognition. *Psychology, Crime and Law, 15*, 73-89. doi: 10.1080/10683160802131131.

Kassin, S. M. (2012). Why confessions trump innocence. *Amercian Psychologist, 67,* 431-445. doi: 10.1037/a0028212.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition, 2,* 42-52. doi: 10.1016/j.jarmac.2013.01.001.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science, 7,* 125-128. doi: 10.1111/j.1467-9280.1996.tb00344.x.

Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology, 97,* 286-314. doi: 10.1016/j.jecp.2007.01.007.

Kemp, R. I., Pike, G. E., & Brace, N. A. (2001). Video-based identification procedures: Combining best practice and practical requirements when designing identification systems. *Psychology, Public Policy, & Law,* 7, 802-807. doi: 10.1037//1076-8971.7.4.802.

Laughery, K. R., Alexander, J. F., & Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology, 55,* 477-483. doi: 10.1037/h0031646.

Lindsay, R. C. L., Semmler, C., Weber, N., Brewer, N., & Lindsay, M. R. (2008). How variations in distance affect eyewitness reports and identification accuracy. *Law & Human Behavior, 32,* 526-535. doi: 10.1007/s10979-008-9128-x.

Lindsay, R. C. L. & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556-564. doi: 10.1037/0021-9010.70.3.556.

Loftus, G. R. & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review, 12,* 43-65. doi: 10.3758/BF03196348.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". *Law and Human Behavior, 11,* 55-62. doi: 10.1007/BF01044839.

Loftus, E. F., Schooler, J. W., Boone, S. M., & Kline, D. (1987). Time went by so slowly: Overestimation of event duration by males and females. *Applied Cognitive Psychology, 1,* 3-13. doi: 10.1002/acp.2350010103.

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482-489. doi: 10.1037/0021-9010.66.4.482.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255-260. doi: 10.1016/S1364-6613(02)01903-4.

McKone, E. (2009). Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. *Vision Research, 49,* 268-283. doi: 10.1016/j.visres.2008.10.020.

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness evidence. *Psychology, Crime, & Law, 17,* 711-729. doi: 10.1080/10683160903524333.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94,* 339-354. doi: 10.1348/000712603767876262.

Morgan, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., Baranoski, M., & Southwick, S. M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law & Psychiatry, 27,* 265-279. doi: 10.1016/j.ijlp.2004.03.004.

Palmer, M. A., Brewer, N., & Horry, R. (2013). Understanding gender bias in face recognition: Effects of divided attention at encoding. *Acta Psychologica, 142,* 362-369. doi: 10.1016/j.actpsy.2013.01.009.

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica,* 133, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Palmer, M. A., Brewer, N., Weber, N. & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19,* 55-71. doi: 10.1037/a0031602.

Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in target-present and target-absent lineups. *Law and Human Behavior, 13,* 133-149. doi: 10.1007/BF01055920.

Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adult' eyewitness identifications. *Law and Human Behavior, 17,* 11-26. doi: 10.1007/BF01044534.

Parkin, A. J., Gardiner, J. M., & Rosser, R. (1995). Functional aspects of recollective experience in face recognition. *Consciousness & Cognition*, 4, 387-398. doi: 10.1006/ccog.1995.1046.

Pedersen, A. C. I., & Wright, D. B. (2002). Do differences in event descriptions cause differences in duration estimates? *Applied Cogntive Psychology, 16,* 769-783. doi: 10.1002/acp.827.

Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition: More evidence for the cross-race effect. *Journal of Applied Psychology*, 88, 760-763. doi: 10.1037/0021-9010.88.4.760.

Pike, G., Brace, N., & Kynan, S. (2002). *The visual identification of suspects: Procedures and practice.* London: Policing and Reducing Crime Unit, Home Office Research, Development and Statistics Directorate.

Pozzulo, J. D. & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law & Human Behavior, 22,* 549-570. doi: 10.1023/A:1025739514042.

R v. Turnbull. (1976). 98 Cr. App. R. 313.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 20, 161-168. doi: 10.1037/0278-7393.20.1.161.

Reynolds, J. K. & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology, 6,* 279-292. doi: 10.1002/acp.2350060402.

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law & Human Behavior, 34,* 337-347. doi: 10.1007/s10979-009-9192-x.

Shapiro, P. N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139-156. doi: 10.1037/0033-2909.100.2.139.

Skagerberg, E. M. & Wright, D. B. (2008). The prevalence of co-witnesses and co-witness discussions in real eyewitnesses. *Psychology, Crime, & Law, 14,* 513-521. doi: 10.1080/10683160801948980.

State of New Jersey v. Larry R. Henderson (2011). 27 A.3d 872.

Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D., Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (p. 144-162). Cambridge, UK: Cambridge University Press.

Valentine, T. & Heaton, P. (1999). An evaluation of the fairness of police line-ups and video identifications. *Applied Cognitive Psychology,* 13, S59-S72. doi: 10.1002/(SICI)1099-0720(199911)13:1+<S59::AID-ACP679>3.0.CO;2-Y.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17,* 969-993. doi: 10.1002/acp.939.

Wagenaar, W. A., & van der Schrier, J. H. (1996). Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime, & Law, 2,* 321-332. doi: 10.1080/10683169608409787.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology, 36*, 1546-1557. doi: 10.1037/0022-3514.36.12.1546.

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law & Human Behavior, 32,* 6-10. doi: 10.1007/s10979-007-9098-4.

Wells, G. L. & Quinlivan, D. S. (2009). Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law & Human Behavior, 33,* 1-24. doi: 10.1007/s10979-008-9130-3.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78,* 835-844. doi: 10.1037/0021-9010.78.5.835.

Wickelgren, W. A. (1972). Time resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9,* 418-455. doi: 10.1016/0022-2496(72)90015-6.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition, 2,* 775-780. doi: 10.3758/BF03198154.

Wright, D. B., Carlucci, M. E., Evans, J. R., & Compo, N. S. (2010). Turning a blind eye to double-blind line-ups. *Applied Cognitive Psychology, 24*, 849-867. doi: 10.1002/acp.1592.

Wright, D. B. & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical & Statistical Psychology*, 62, 439-456. doi: 10.1348/000711008X327632.

Wright, D. B. & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology*, *10*, 75-84. doi: 10.1002/(SICI)1099-0720(199602)10:1<75::AID-ACP364>3.0.CO;2-E.

Wright, D. B., Memon, A., Dalton, G., Milne, R., & Horry, R. (2013). Field studies of eyewitness memory. In B. L. Cutler (Ed.), *Reform of Eyewitness Identification Procedures* (p. 179-201). Washington, DC: APA Publications.

Wright, D. B. & Skagerberg, E. M. (2007). Postidentification feedback affects real eyewitnesses. *Psychological Science*, 18, 172-178. doi: 10.1111/j.1467-9280.2007.01868.x.

Yarmey, A. D., & Yarmey, M. J. (1997). Eyewitness recall and duration estimates in field settings. *Journal of Applied Social Psychology, 27,* 330-344. doi: 10.1111/j.1559-1816.1997.tb00635.x.

Table 1

Associations among predictor variables (Cramer's V).

	Corroboration	Crime type	Witness sex	Witness age	Exposure duration	Distance	Number of perpetrators	Dela
Corroboration	---	.251***	.031	.096	.148***	.201***	.025	.171*
Crime type		---	.407***	.216***	.231***	.217***	.440***	.212*
Witness sex			---	.088	.069	.090	.200***	.162*
Witness age				---	.132**	.077	.052	.124*
Exposure duration					---	.266***	.194**	.027
Distance						---	.085*	.078
Number of perpetrators							---	.141*
Delay								---
Use of violence								
Weapon presence								
Role of witness								

Note: Significance levels: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 2

Global regression model for predicting suspect identifications for corroborated cases (left) and for uncorrob

Parameter	Level 1	Level 2		Corroborated cases (N = 588) Estimate	SE	z	*p*	Uncorrob Estimate
Fixed effects								
Intercept			β_0	-1.05	0.57	-1.84	.07	-2.35
Crime type	Rape/indecent assault	Robbery	$\beta_{1(a)}$	-0.32	0.45	-0.71	.48	1.18
	Rape/indecent assault	Other crime	$\beta_{1(b)}$	0.19	0.62	0.31	.76	-0.19
	Rape/indecent assault	Assault	$\beta_{1(c)}$	-0.81	0.45	-1.79	.07	0.22
	Rape/indecent assault	Dishonesty	$\beta_{1(d)}$	-0.42	0.42	-1.01	.31	-0.26
Witness sex	Male	Female	β_2	-0.06	0.22	-0.26	.79	0.52
Witness age	10-15	16-20	$\beta_{3(a)}$	0.58	0.43	1.36	.17	2.20
	10-15	21-50	$\beta_{3(b)}$	0.10	0.38	0.27	.79	1.44
	10-15	Over 50	$\beta_{3(c)}$	-0.46	0.48	-0.96	.34	1.87
Exposure duration	< 60 seconds	> 60 seconds	β_4	0.88	0.25	3.46	<.001	0.90
Distance	> 5 meters	< 5 meters	$\beta_{5(a)}$	1.28	0.32	3.86	<.001	-0.12
	> 5 meters	Face-to-face	$\beta_{5(b)}$	1.22	0.32	3.84	<.001	-0.66
Number of perpetrators	One	More than one	β_6	-0.39	0.25	-1.60	.11	-0.66
Delay	< 1 month	1-3 months	$\beta_{7(a)}$	-0.08	0.28	-0.29	.78	-0.33
	< 1 month	> 3 months	$\beta_{7(b)}$	0.09	0.34	0.27	.79	-0.76
Use of Violence	Yes	No	β_8	0.31	0.33	0.91	.36	-0.21
Weapon	Present	Absent	β_9	0.59	0.33	1.80	.07	-0.94
Role of witness	Victim	Bystander	β_{11}	-0.00	.29	-0.01	.99	0.53
Random effects								
Case number			σ	0.56	0.75			1.28

Table 3

Global regression model for predicting foil identifications for corroborated cases (left) and for uncorroborat

				Corroborated cases (N = 588)				Uncorrob
Parameter	Level 1	Level 2		Estimate	SE	z	*p*	Estimate
Fixed effects								
Intercept			β_0	-3.35	1.05	-3.20	.001	-1.07
Crime type	Rape/indecent assault	Robbery	$\beta_{1(a)}$	0.91	0.82	1.10	.27	-2.12
	Rape/indecent assault	Other crime	$\beta_{1(b)}$	1.13	0.99	1.13	.26	0.57
	Rape/indecent assault	Assault	$\beta_{1(c)}$	1.81	0.80	2.27	.02	-0.76
	Rape/indecent assault	Dishonesty	$\beta_{1(d)}$	0.84	0.76	1.11	.27	0.44
Witness sex	Male	Female	β_2	-0.37	0.34	-1.11	.27	-0.25
Witness age	10-15	16-20	$\beta_{3(a)}$	0.55	0.80	0.69	.49	-1.48
	10-15	21-50	$\beta_{3(b)}$	1.00	0.72	1.39	.17	-0.71
	10-15	Over 50	$\beta_{3(c)}$	1.30	0.83	1.57	.12	-0.98
Exposure duration	< 60 seconds	> 60 seconds	β_4	0.31	0.40	0.77	.44	-0.19
Distance	> 5 meters	< 5 meters	$\beta_{5(a)}$	-0.68	0.47	-1.43	.15	0.67
	> 5 meters	Face-to-face	$\beta_{5(b)}$	-0.87	0.45	-1.95	.050	0.49
Number of perpetrators	One	More than one	β_6	-0.39	0.39	-1.01	.32	-0.35
Delay	< 1 month	1-3 months	$\beta_{7(a)}$	-0.07	0.46	-0.16	.87	0.85
	< 1 month	> 3 months	$\beta_{7(b)}$	-0.14	0.57	-0.25	.81	0.07
Use of Violence	Yes	No	β_8	-0.71	0.55	-1.29	.20	1.74
Weapon	Present	Absent	β_9	-1.25	0.60	-2.08	.04	0.99
Role of witness	Victim	Bystander	β_{11}	0.72	0.46	1.57	.12	-1.06
Random effects								
Case number			σ	2.04	1.43			0.39

Table 3

Global regression model for predicting non-identifications for corroborated cases (left) and for uncorrobora

Parameter	Level 1	Level 2		Corroborated cases (N = 588)				Uncorrob
				Estimate	SE	z	*p*	Estimate
Fixed effects								
Intercept			β_0	0.63	0.59	1.08	.28	0.26
Crime type	Rape/indecent assault	Robbery	$\beta_{1(a)}$	-0.12	0.48	-0.26	-.80	-0.02
	Rape/indecent assault	Other crime	$\beta_{1(b)}$	-0.88	0.73	-1.21	.23	-0.54
	Rape/indecent assault	Assault	$\beta_{1(c)}$	-0.04	0.49	-0.08	.94	0.05
	Rape/indecent assault	Dishonesty	$\beta_{1(d)}$	0.06	0.45	0.13	.90	-0.13
Witness sex	Male	Female	β_2	0.28	0.24	1.18	.24	-0.25
Witness age	10-15	16-20	$\beta_{3(a)}$	-1.02	0.46	-2.23	.03	-0.50
	10-15	21-50	$\beta_{3(b)}$	-0.69	0.40	-1.74	.08	-0.41
	10-15	Over 50	$\beta_{3(c)}$	-0.26	0.49	-0.54	.59	-0.32
Exposure duration	< 60 seconds	> 60 seconds	β_4	-1.14	0.26	-4.46	<.001	-0.47
Distance	> 5 meters	< 5 meters	$\beta_{5(a)}$	-0.81	0.33	-2.42	.02	-0.09
	> 5 meters	Face-to-face	$\beta_{5(b)}$	-0.58	0.32	-1.84	.07	0.10
Number of perpetrators	One	More than one	β_6	0.74	0.27	2.74	.006	0.71
Delay	< 1 month	1-3 months	$\beta_{7(a)}$	0.13	0.29	0.44	.66	-0.18
	< 1 month	> 3 months	$\beta_{7(b)}$	-0.08	0.37	-0.22	.83	0.67
Use of Violence	Yes	No	β_8	-0.03	0.38	-0.08	.94	-0.60
Weapon	Present	Absent	β_9	-0.03	0.35	-0.09	.93	0.07
Role of witness	Victim	Bystander	β_{11}	-0.47	0.33	-1.41	.16	0.22
Random effects								
Case number			σ	0.53	0.73			0.19