Forensic Science International: Digital Investigation 35 (2020) 301062

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

Part 1: The need for peer review in digital forensics

Graeme Horsman [a], Nina Sunde [b, *]

[a] Teesside University, Middlesbrough, Tees Valley, TS1 3BX, UK
[b] Norwegian Police University College, Pb. 2109 Vika, 0125, Oslo, Norway

ARTICLE INFO

Article history:
Received 21 February 2020
Received in revised form
3 July 2020
Accepted 9 August 2020
Available online 23 September 2020

Keywords:
Digital forensics
Peer review
Digital evidence
Quality assurance
Quality control
Forensic science

ABSTRACT

The importance of peer review in the field of digital forensics cannot be underestimated as it often forms the primary, and sometimes only form of quality assurance process an organisation will apply to their practitioners' casework. Whilst there is clear value in the peer review process, it remains an area which is arguably undervalued and under-researched, where little academic and industrial commentary can be found describing best practice approaches. This work forms the first of a two part series discussing why the digital forensics discipline and its organisations should conduct peer review in their laboratories, what it should review as part of this process, and how this should be undertaken. Here in part one, a critical review of the need to peer review is offered along with a discussion of the limitations of existing peer review mechanisms. Finally, the 'Peer Review Hierarchy' is offered, outlining the seven levels of peer review available for reviewing practitioner findings.

1. Introduction

It goes without saying that any form of evidence entering a criminal justice system must be reliable if it is to be used as part of any legal decision making (Ivanovic, 2019). Whilst this may seem a simple concept, in reality achieving this is a challenge, which has led to recent scrutiny of all forensic science sub-disciplines (Hamer and Edmond, 2019; Roux, 2019). Forensic science as a whole is littered with examples of miscarriages of justice (Hamer and Edmond, 2019) and examples of unsafe evidence making it to trial (see Smit et al.'s., 2018 review of cases in England and Wales). Given the importance and power of forensic evidence in criminal cases, this remains a deeply unsatisfactory position that forensic science finds itself in, and in England and Wales the Forensic Science Regulator has voiced such concerns (Tully, 2019, 2020).

Whilst the wider forensic community maintains its own set of issues, focus here is maintained on the digital forensic (DF) arena. Whilst there are arguably fewer high-profile documented examples of the poor practices leading to miscarriages, as digital evidence continues to feature prominently in many criminal investigations, entering legal systems from a range of investigative authorities and a diverse set of devices, it is arguably only a matter of time before similar is witnessed. As a result, it is important that the field of DF does not mimic the mistakes of other forensic sub-disciplines and fail to assess the reliability of the work its practitioners produce to a sufficient, robust standard (Page et al., 2019). Whilst it remains easy to make such a statement regarding the need to ensure measures for determining reliability are in place, how we validate the quality of this evidence has and continues to remain one of the greatest challenges that the DF field will face, and is arguably yet to be satisfactorily addressed Horsman, 2020.

ISO 9000:2015 (EN) (International Organization of Standardization, 2015) section 3.6.2 defines quality as the "degree to which a set of inherent characteristics of an object fulfils requirements", where object means anything perceivable or conceivable - for example a product, service, process, person, organisation, system or resources. It is also useful to distinguish between two different dimensions of quality management: Quality Assurance (QA), and Quality Control (QC) QA is a preventative enterprise, which is focused on ensuring that quality requirements will be fulfilled (Doyle, 2019). QC is reactive, and focused on controlling that quality requirements have been fulfilled (Doyle, 2019). In DF, essential to the assessment of evidence 'quality' are the dimensions of validity and reliability associated with the human or tool/technology based methods used, and the end result. Hand (2004) describes reliability in the scientific sense as how different observers measure or score the same phenomenon differently, and

* Corresponding author.
E-mail addresses: g.horsman@tees.ac.uk (G. Horsman), nina.sunde@phs.no (N. Sunde).

https://doi.org/10.1016/j.fsidi.2020.301062

there are a variety of statistical measures to quantify reliability depending on the type of observation. Validity could be described as "the overall probability of reaching the correct conclusion, given a specific method and data" (Christensen et al., 2014, p. 124). In turn, reliability is a multi-factored construct, despite the term is often used as an umbrella kite-mark. In reality, reliability covers the overall worth of a piece of evidence but this overall worth consists of three sub-areas; the reliability of the investigatory processes undertaken; the reliability of practitioner interpretation and the reliability of the practitioner's ability to convey this information to their target audience. Whilst the reliability of evidence will be assessed at its point of entry into any legal process, it should also be assessed when in the laboratory environment. The advantages of reliability checks at this point are as follows:

1. *Cost saving*: Problems with findings which are established at this stage mean that the costing implications with initiating legal proceedings is prevented.
2. *Damage limitation*: Limiting potential damage to the practitioner and organisational reputations.
3. *Safeguard the rule of law:* Preventing misleading or erroneous information being used in trial proceedings.

Ensuring evidence reliability and validity is the job of the investigating practitioner, supported by their organisation. Whilst this may seem straightforward, the fact that there are currently few formalised and standardised methods for assessing the reliability of digital forensic evidence in the laboratory environment means that achieving this is difficult. For most (if not all), lab-based peer-review is the primary source of quality control and evidence reliability checks for practitioner work, and it rightly should be as peer reviews are a known technique for reducing errors (Keim, 2003). Yet the problem remains that there are few, if any available resources discussing how to conduct a DF peer review and what the process should involve. As a result, it is likely that peer review approaches are divergent between organisations suggesting that the quality and robustness of such peer reviews may also vary. Ineffective peer review in DF has been noted for over 15 years (Meyers and Rogers, 2004) and if the field is to secure its status as an important and reliable forensic science, it must take steps towards rigorous peer review procedural development.

Based on information collected during the assessment of ISO/IEC 17025 accredited DF units in England and Wales undertaken by United Kingdom Accreditation Service, Tully et al. (2020) conclude that standards are no absolute guarantee for quality, noted below.

> "A wide range of technical, administrative and management findings were raised, demonstrating that the accreditation process is addressing the provision of digital forensics services as a system and not concentrating on a single element, such as the technical examination of an item. No matter how skilled the examiner, if he or she is not supported with the requisite equipment, resources and ongoing training, the system will not function as it should." (Tully et al., 2020 p. 9).

This article forms the first in a two part series discussing peer review in the field of DF. These pieces will provide the following narratives:

- *Part 1*: Why the DF field should peer review, discussing sources of error and peer review concepts, strengths and weaknesses.
- *Part 2*: What and how the DF field should peer review, where the Phase-oriented Advice and Review Structure (PARS) for digital forensic investigations is offered and described. The available options to ease the implementation into existing practices are also discussed (Sunde and Horsman, 2020).

Peer review for DF investigations is an under researched area and is yet to be formally addressed and evaluated by the community for scrutiny. Through this series of papers covering this topic area we provide the first formalised set of focused discussions on this topic and the first documented peer review model for use in DF investigations. Part 1 commences with a discussion of the two sources-types of error that can occur in a DF investigation, technical and non-technical. The need for peer review, Problems with it and the forms it can take are then described in Section 3 along with the 'peer review hierarchy', which denotes seven levels of peer review. Finally, we posit the requirement for peer review as part of all DF investigation processes, paving the way for the PARS model to be outlined and discussed in Part 2.

2. Error in DF investigations

No tool, process, technique or practitioner can guarantee error free results over a course of time, hence the importance of processes designed to check such work. Yet before considering the details of peer review and its importance, it is first necessary to identify the sources of error which can exist in a DF investigation. Error sources can be separated into two main categories: 'technical sources of error' and 'non-technical sources of error', with each described in turn below. It is important to highlight that when discussions around peer review occur, there is an assumption that it is the human who is subject to review, but such a view is too narrow. Error sources include erroneous tool outputs, potential bias, oversight and pressure from external influences (Welner et al., 2014) and it is necessary to address each in turn and ensure that such checks incorporate this into a review. We acknowledge that errors may happen as a result of a deliberate action made with bad intent, however, have no reason to believe that this is a major problem area in DF and will focus on unintended errors in the further sections.

2.1. Technical sources of error

In line with the Scientific Working Group on Digital Evidence (SWGDE, 2018), errors linked to DF tools and technology usage can be placed into three categories: techniques, implementation error, and tool usage and interpretation error (SWGDE, 2018):

1. *Techniques*: Errors which can be attributed to techniques (algorithms) are those which may occur randomly where the potential exists to calculate their occurrence via error rates (SWGDE, 2018). However, the use and value of error rates in relation to DF tools and techniques is not straightforward. Calculating an error rate requires a 'population' (a set of known documented data depicting a specific technique's behaviour), and due to the rapid technology development, as well as the complexity and diversity of devices and technology this is not always viable to achieve, and therefore in many cases may not actually be possible.
2. *Implementation error:* Implementation of the algorithms in tools may lead to systematic errors, since they are logical flaws that will repeat themselves. Implementation errors are those which occur due to issues during the tool or techniques development. It is perhaps in this category easier to consider the example of a forensic software package. During the development and testing phase, it is possible that any given tool does not leave the developer being wholly error-free, and this is often raised in the tool's end-user license agreement (EULA) (Horsman, 2019,

2018). These are Implementation errors broadly fall into three sub-categories:

i. *Incompleteness*: All available information that should be identified by the tool (taking into account what the tool states it does) is not. It is important here to make sure that errors caused by misuse of a tool are not confused with implementation errors.
ii. *Inaccuracy*: The tool should not report artifacts that do not exist, group unrelated items together or alter data in a way that changes its meaning.
iii. *Misinterpretation*: The results have been incorrectly understood by the tool.

3. *Tool usage and interpretation of results*: This category also relates to errors associated with human–technology interaction. Even in cases where a tool is functioning correctly (having been tested), usage errors can still occur where a practitioner incorrectly operates a tool, or when using the tool for a task it was not designed to handle. Interpretation error may occur because of ambiguities in the way the tool presents the data.

Technical sources of error are important to consider and guard against, as there can be a perceived consensus that the tools practitioners pay for should not, and therefore are not, prone to error. As a result, under this belief it may be easy to see how errors occurring from such a source may not be considered by review systems.

2.2. Non-technical sources of error

In addition to those issues noted above, random mistakes and systematic errors are a risk factor of any process and should also be of concern. Errors of this type are difficult to guard against as they may appear in any part of the investigatory process and potentially in areas where there is a perceived expectation that the practitioner would perform error free, making the need for review processes great. As a result, it should not be assumed that just because a task may be viewed as 'simple', that it should not also be scrutinised for the potential of errors. Systematic errors should be a bigger concern as they inherently repeat themselves and may be an indication of flawed investigatory or organisational procedures, or thought processes. The following sources of non-technical error are proposed:

Knowledge: As in any discipline, knowledge is often involved in the presence of errors in work. Whilst knowledgeable practitioners are not infallible to error, those which have wide encompassing knowledge of their craft are arguably in a better position to prevent errors occurring. Conversely, those practitioners who lack knowledge in specific areas may be more at risk of errors. This raises the issue of ensuring that practitioners know and understand the remit of their expertise and no not supersede this without specific training and support. The problem this causes is that due to the pace of technological change, it is common for practitioners to encounter data types that they have not seen before but are expected to interpret accurately.

Experience: Linked closely with knowledge, experience is an important aspect for error mitigation. Practitioners who can call upon previous experiences and knowledge may be in a position to do this in order to prevent future error. Where a practitioner has seen a previous comparable investigative scenario and learned from this, specifically if previous errors were made, then repetition of erroneous practice is arguably less likely. This typically takes into consideration the old adage of 'learning from one's past experiences'. In comparison, inexperienced practitioners may not be in a position to draw upon such experiences.

Lack of procedures: Digital Forensics is characterized by a set of well known principles, such as maintaining evidence integrity and chain of custody (Casey, 2011; Årnes, 2018), or as stated in ISO/IEC 27037:2012 (EN) - auditability, justifiability and repeatability/reproducibility (International Organization of Standardization, 2012). However, these principles do not give any guidance in how the tasks in the DF process should be performed. Research in expert performance has suggested that procedures and checklists are important tools for error mitigation. Research from the medical domain showed that by introducing checklists on a few and well-known steps for prevention of infections (e.g. wash hands, cleaning the skin with chlorhexidine, removing unnecessary catheters) had a profound effect on the infection rate - it was nearly eliminated during the period of the study (Pronovost et al., 2006). What should be noticed is that the errors were not due to lack of knowledge. The medical personnel knew very well that they should wash their hands and sterilize the skin before inserting a catheter. The checklist was to avoid mistakes that could happen for a number of reasons, such as time pressure or inattention due to distractions - regardless of level of expertise. Checklists should therefore not be a substitute for knowledge - but should be aimed at underpinning expert performance and reducing the risk of mistakes.

Subjectivity and interpretation: Every process that involves a human is prone to human error, and particularly when the process involves subjectivity, interpretation, judgements and decisions. These factors are relevant to every phase of the digital forensic process. Examples of human errors include finding false negatives - *where a practitioner does not find what is actually there*, or finding false positives - *where a practitioner finds/sees something that is actually not present*. Every practitioner maintains the potential to misinterpret the meaning of a given set of evidential data. When concluding or presenting the evidence there remains a risk of overstating (or understating) the relevance or reliability of any given evidential findings.

Cognitive bias: Systematic errors caused by cognitive bias are identified as a problem across forensic science as a whole, influencing observations and decisions. Nakhaeizadeh et al (2014, p. 208) state "research has shown that decision-making can be influenced by cognitive processes and cause forensic experts to modify their judgement". The field of DF is not immune to such issues and therefore issues around cognitive biases errors are of equal importance in this domain. Dror (2017) described a seven level taxonomy of factors that could lead to bias, and recently added an additional eight level (Dror, 2020). The taxonomy involves factors related to human nature, environment, culture and experience and case specific factors.

It is important to notice that this is about cognitive bias, which is not about errors related to bad ethics or bad intentions or incompetence. These are biases which are mostly unconscious processes, which happens with experts as well as novices. A global survey among trained professionals in several forensic science disciplines suggests a bias blind spot, which entails that they tend to recognise biases in others, but deny the existence of those same biases in themselves (Kukucka et al., 2017). This survey also showed a limited understanding for how to effectively mitigate cognitive bias. 71% believed that bias could be reduced by simply trying to ignore their expectations.

While bias in forensic science work has been subject to research, the problem of cognitive bias in DF investigations is not yet explored with empirical research. Drawing on findings from other forensic science domains, identified that all the levels, in what was then a seven level taxonomy, are relevant in DF. While there is subjectivity, choice and interpretation in all the phases of the DF process, none of the phases should be considered secure from human error caused by cognitive bias.

In addition, there are several situational factors that may lead to bias, such as time pressure, emotions, perceived responsibility, and the concern for future consequences (Ask, 2013).

1. *Pressure*: The pressures of the DF practitioner's role can, and often are, great. In most cases a practitioner will manage multiple cases at once, dealing with the competing demands of their clients and their employing organisation. Often, time requirements will be placed upon any given examination, whether that be costed-time (the amount of time a practitioner is 'billed' to work on a case) or if in the criminal sphere, schedules court dates may dictate approaches. In either case, the practitioner does not have an infinite amount of time to complete their investigation allowing an examination of every file present (Pollitt, 2008), instead they will be faced with pressures to complete work within specific time frames, which can be pressure-inducing. Time-pressure is also a recognised risk factor for cognitive bias (Ask, 2006).
2. *Emotions:* DF practitioners often handle material that may affect them emotionally. Carrying out examinations with for example child sexual exploitation and abuse (CSEA) material may lead to frustration, sorrow and anger, or a feeling of personal responsibility for the outcome of the investigation. These emotions may bias the observations and conclusions of the practitioner (Ask, 2013).

3. Peer review

The importance of peer review in forensic science cannot be understated (Heilbrun et al., 2004), and the peer review of practitioner case work should be sought in every instance (Cohen, 2010). When discussing what science is, Doyle (2019 p.2) argues that "review by an expert of at least equal standing is a minimum requirement for the enquiry to qualify as scientific". It forms part of the expert evidence admissibility tests seen in *Daubert v. Merrell Dow Pharms., 516 U S. 869 (U S. 1995)* (sometimes referred to as the *Daubert Test* or *Daubert Standard*) in the United States (Jonakait, 1993; Meyers and Rogers, 2005). Peer review is one of the main mechanisms for vetting work for errors and to correct these where present (Sims et al., 2013) indicating that 'inadequate performance in forensic work can be monitored and corrected by implementation of a program of peer review' (Appelbaum, 1992). Peer review is also one of the best ways to overcome and prevent potential biases in judgment which a practitioner may have made (Budowle et al., 2009) and facilitate continual 'quality improvement' (Obenson and Wright, 2013). Therefore through the integration of peer reviewers', complementary expertise and experience enhances the sophistication and overall quality of assessment (Welner et al., 2014)'. Perhaps the need for peer review is best offered by Welner et al (2012, p.1)

> "Peer-reviewed forensic science is an important methodological solution for ineffective and unethical forensic science assessment, promoting integrity, quality, and confidence in justice" (Welner et al., 2012, p.1).

Regardless of the discipline, the aims of peer review, which are threefold, largely remains the same:

1. To uncover and correct errors in any current case subject to review.
2. To assess the reliability of any given investigation and results. Where errors are not found to be present, peer review acts as a quality 'kite-mark' and reassurance that no errors are present, not just because they have been missed, but that the work is confirmed as being error free.
3. To improve the quality of future work by offering learning experiences. Errors may be systematic or random, and peer review is relevant to address both, highlighting weaknesses in laboratory processes. Identifying bad practice means that it can be addressed before it compromises further activities.

In relation to the DF field, the need for peer review is no different to traditional forensic science types; the mechanism must be viewed as a mandatory element of the DF investigatory process. In part, this is due to the diversity and complexity of digital evidence types which may occur, where drawing reference to other investigators sciences may provide an illustrative example. Horsman (2019) compares fingerprint analysis to that of a DF investigation to highlight the differences in both investigative procedures, where this example can be taken further here. Where fingerprints are lifted from a source, this may in some cases be carried out using a single process (examples include gel lifting or powder and photograph with light source) followed by a 'match-analysis' carried out by the practitioner and any supporting match databases to link the print to a suspect. The point is that the peer review of the methodology aspect of the investigation is somewhat simple due to the linear approach - `identify print -> capture and lift -> analyse`. Comparatively, a DF practitioner may utilise multiple tools within any single examination, followed by any number of features within that tool - each capable of being configured incorrectly, misunderstood by the practitioner or performing erroneously. Therefore any one DF examination may contain multiple processes being run by a practitioner, each requiring review and validation and as a result, the review process is potentially exponentially greater in size as the amount of tasks where errors may occur are greater in number. It is argued that the scope for undetected error is greater in DF due to the amount of components that are involved in a typical process.

When considering the need for peer review, it would seem to be an obvious solution yet it is not without its issues.

3.1. Peer review and its problems

Generally, criticisms of peer review often stem from its use in academic publishing contexts (Smith, 1988; Daniel at al., 2007; Henderson, 2010), and many of these Problems overlap with the use of peer review in the context of forensic science. The following are avenues often raised as weaknesses of peer review, placed in the context of its implementation in DF.

We don't know if it is being done at all in organisations: Perhaps one of the biggest issues we have is that the DF field currently does not know what every organisation is doing as part of their quality management processes. Whilst it is assumed that peer review is currently the primary method for detection unreliable work, at this moment there is no consensus as to whether peer review is even carried out in every organisation. If we assume that any form of peer review is better than none, an initial position requires a declaration from all organisations to state that as part of case review processes, work undergoes review before leaving the confines of their laboratory.

We don't know how it is being done: If an assumption is made that peer review is taking place within organisations, then it is likely that different organisations will be approaching this task differently. The lack of a defined peer review standard means that it is currently unknown as to what shape and form a peer review in DF currently takes.

We don't know at what stage a review is occurring: The investigatory process in DF is multifaceted meaning that any peer review process should arguably take this into account. It is assumed that if peer review takes place at present, it is likely that this task is one which is complete at the close of all work. This may appear a logical approach at first glance, but the problem this raises if that Problems which are detected at this point may be substantial to address and earlier errors may compound issues later down the line. This raises the question as to when the best time to initiate a peer review is and whether the review itself is not singular in nature, but a processes with multiple steps to it.

We don't know what a review 'contains': One of the Problems with peer review in this context lies with what the review itself should contain. In essence, what questions must the reviewer ask and what issues must the investigating practitioner address in order to convey reliance in their given evidence. Whilst arguably the content of the review will be dependent on the context of a given case, the goal in all cases is the same; to demonstrate reliability. How this is achieved through review is yet to be defined.

We don't know who is reviewing: A key task in any peer review is the allocation of an effective peer reviewer. This is even more challenging in the laboratory environment which likely maintains a relatively small demographic of people to choose from who most likely vary in experience and knowledge. Effective peer review hinges on the ability of the reviewer and therefore the review of complex casework by a junior colleague who has never witnessed such work before may not achieve the required level of rigor. There remains a real risk that in some organisations and laboratories, the current mix of staffing abilities may prevent effective peer review from taking place.

We don't know how practitioners respond to a review: Similar to issues relating to what an effective peer review looks like, we also maintain Problems as what an effective peer review response should be. The obvious answer is a response, which passes all of the challenges to evidence reliability offered as part of the review, yet this may not always be straight forward. Essentially a review process must consider when to determine an effective response has been received from the peer review process, what an effective response is, and, most problematic, what to do when an issue as part of peer review is raised but has not or cannot be addressed.

We don't know the level of depth involved with a review: Consideration must be given as to the depth of analysis undertaken as part of the review process. Approaches such as 'fact checking' vs. procedural repeating for validation purposes both carry different advantages and disadvantages. This issue is discussed further in Section 3. Interviews in a study with Norwegian DF practitioners sheds some light to this challenge (Sunde, 2017). The DF practitioners were asked about how quality checks of reports were done, and told that they would routinely read and assess each other's reports, but they had no template or list of criteria to refer to. Their feedback to the author would therefore depend on the reviewers ability, motivation and knowledge.

Peer review training: Given the importance of peer review, this process should be treated with equal importance to all other aspects of a forensic investigation. Sufficient training is required in order to become a competent peer reviewer and to recognise what one looks like. Training ensures peer reviews are appropriate where a non-effective peer review risks being of even greater detriment by giving a stronger perception that work is reliable just because it has been checked. It is worth noting that being taught to peer review may not be commonly part of forensic science practitioner training and therefore given the importance of peer review, it is suggested that this become part of syllabuses going forward.

Peer reviewing must be efficient: Whilst this may seem contrary to the purpose and nature of a peer review, it is important that the peer review process is efficient. Peer review mechanisms that are over-burdensome risk disengagement from staff in the laboratory environment. There is a trade-off to be struck between a robust review and one that fits the needs of the organisation.

Peer reviewer experience: The experience of the peer reviewer has been previously mentioned and this raises questions as to what happens when there is not the experience in a lab setup to undertake a peer review and how this can be overcome.

Evaluating a peer review: Whilst the peer review is designed to evaluate a set of results, the peer review itself should be evaluated for quality. Any peer review process should have mechanisms in place to identify poor peer reviewing.

Problems *with self-regulating*: The problem most laboratories will have lies with self-regulation - peer review of their staff will take place using their staff. While each individual is their own entity, they will be shaped by common internal procedures and processes. Therefore, the risk of complacency and bias as part of a review may be an issue.

Emotional side of peer review: Peer review is designed to scrutinise and evaluate. However, we do not know the impact on the review when a reviewer does not want to criticize colleagues, or feel uncomfortable to do so?

A meaningful task: Whilst the concept of a peer review should be viewed as useful by anyone operating in any discipline of forensic science, in reality it may not always be fully engaged with. Given that peer reviewing is an additional workload burden, how do we make it a task which does not allow a practitioner to just 'go through the motions' and tick boxes. In essence, the peer review must be a meaningful task.

Evidence too complex to review, requiring re-examination: This issue is more likely to arise in complex, multi-exhibit investigations where the type and volume of evidence may be such a level of complexity and volume that it becomes difficult to validate practitioners interpretation and procedures without having undertaken the same work again. Whilst controversial to suggest, peer review may not be suited to every investigatory instance where depending on those factors noted above and the severity of a suspected offence. It may be of less risk to repeat the work where the impact of error may be seen as too severe to risk standard peer reviews as the only method of detecting them.

Acceptance of review results by the reviewee: In essence, the principle concern here is 'will the recipient of a peer review be receptive to the results of a peer review, and if not - what happens when a dispute arises?' Therefore it is important that consideration must not just be given to developing a system of peer review, but how it is managed taking into account the need for dispute resolution in some instances.

We don't know the effect of the evidential status 'peer reviewed': Peer review is concerned with verifying the results of an investigation. Yet those involved later in the judicial chain do not have the same prerequisite to assess the quality of the evidence. They often lack the competency to ask the right questions concerning the reliability of the process of producing the evidence and the validity of the result - and would probably trust the outcome of a peer review. The problem is that if the peer review is by now a non-standardized process, which may vary from reading through the report looking for bad spelling and grammar errors, to a full re-examination of the evidence files. This entails that the report status 'peer reviewed' may be associated with verification of all the findings through a high quality process, while the truth sometimes may be a brief check of the report. A non-standardized and non-transparent peer review process may therefore add another layer of trust to the evidence, instead of providing insight into the actual quality and probative value of the evidence.

3.2. Forms of peer review in DF

Peer review is mentioned or described in several standards relevant to DF, such as ISO/IEC 17020:2012, ISO/IEC 17025:2017 and

ISO/IEC 27001:2013 (International Organization for Standardization, 2012, 2013, 2017). Watson and Jones (2013) present a list of questions to support an administrative review (p. 192) as well as a technical peer review of the case (p. 406–407). The questions are general, and referring to expressions, which may be open to interpretation if not specified. For example “is the deliverable (report, statements, depositions, etc.) accurate” and “were the methods used appropriate” (Watson and Jones, 2013 p. 407).

While standards may provide useful guidance for peer review, they are often behind paywalls. We have searched for detailed descriptions of peer review in digital forensics that are publicly available. ENFSI (2015 p. 16) underlines the importance of peer review “to ensure the strength and provenance of the details being assessed”, and highlight the importance of independence and sufficient competence when assigning a reviewer. A template for peer review of reports is provided (ENFSI, 2015, Appendix G, G.3), and, similar to the lists provided by Watson and Jones (2013) contains quite general descriptions of what the technical peer review should focus on, for example,“have the appropriate examination been carried out?” and “was the analysis proportionate?” (ENFSI, 2015 p. 62).

In 2019, Page et al. considered the structure and content of an effective peer review for a DF organisation, noting down five levels, notably the ‘proof check’, ‘sense review’, ‘conceptual peer review’, ‘verification review’ and ‘re-examination’. This work absorbs these levels as a foundation for discussion and elaborates further on each of these review types, whilst also extending the forms of peer review which can be opted for in DF. A proposed extension is made to Page et al.'s (2019) work, which is formalised in our ‘*Peer Review Hierarchy*’, shown in Fig. 1, and discussed below.

Fig. 1 provides for seven levels of peer review, organized hierarchically where it is considered levels 1–3 are less likely to detect critical errors in any investigation, but the implementation of these peer review-types is less resource intensive as they focus on evaluating the presentation of reported work. In comparison, peer reviews at levels 4–7 begin to evaluate and interpret the findings of reported work, inevitably a more resource intensive process due to the level of required scrutiny, naturally increasing the chance of critical error detection. We consider the boundary between level 3 and 4 to be the point in the peer review hierarchy where the review moves from evaluating the appearance of work, to evaluating its substance.

The first point of note and deviation from Page et al (2019, p. 89) is the inclusion of an ‘administrative check’, considered here as a level 1. Second, Page et al.'s (2019, p. 89) ‘verification review’ has been split into two separate review types, a ‘sampled verification review’, and a ‘full verification review’. As a result, there are seven proposed peer review types, which a DF organisation may choose to implement.

The hierarchical nature of the types of peer review shown in Fig. 1 are likely to have an unsurprising impact on both their chance of detecting error and the amount of time and effort it takes to conduct the review itself. In essence, the distinguishing factors between each review type is the amount of scrutiny applied at each level, and the amount of organisational resources it will take to carry out this task. Perhaps the best example to demonstrate this is a direct comparison between a level 2 ‘proof check’ which may identify grammatical and typographical errors in a practitioners report, and a level 7 ‘re-examination’, which could discover procedural flaws and misinterpretation of evidence (critical errors). The obvious problem here is that whilst a proof check may take an hour of the reviewer's time, a re-examination could take days. As a result, there is a trade-off which an organisation must make between the pursuit of robust quality control, and the feasibility of carrying out a peer review.

A divide is placed between levels 1–3 and levels 4–7 distinguishing reviews which focus purely on the presentation of a practitioners report (levels 1–3) and those which begin to scrutinise the investigatory work which has been undertaken (levels 4–7). It should also be noted that as the review types are structured hierarchically, tasks at lower points in the hierarchy are absorbed by reviews higher up. For example, those carrying out a level 4 conceptual review, as part of the review process would be expected to incorporate all of those tasks which would occur at levels 1–3. Each of the seven review types are discussed below.

3.2.1. The ‘administrative review’ (level 1)

Acknowledgement of the need for an administrative review

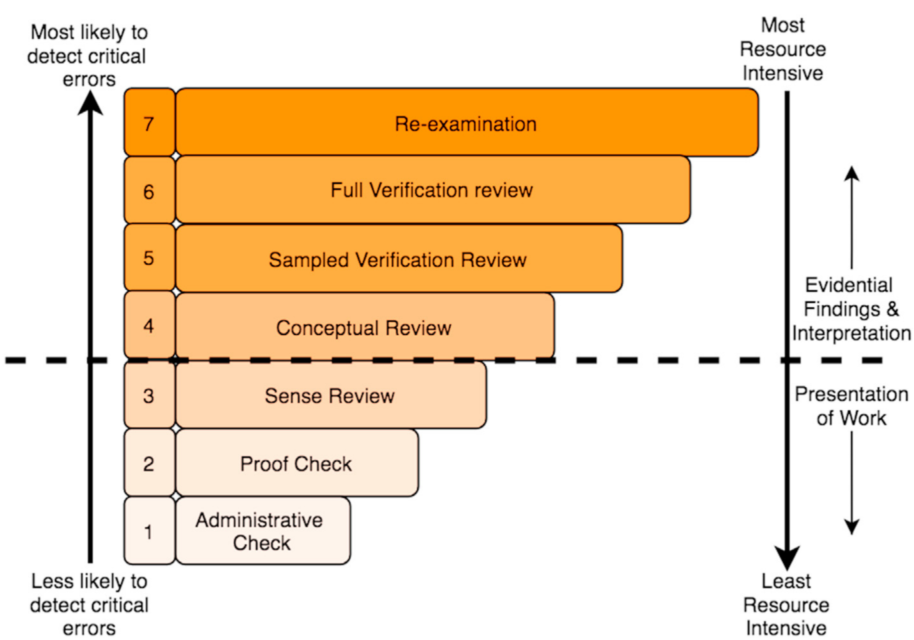

Fig. 1. The ‘peer review hierarchy’ for DF.

stems from Ballantyne et al.'s (2017) overview of peer review in forensic science, and Watson and Jones (2013) description on how to meet the requirements of ISO/IEC standards relevant to DF. An administrative review focuses on whether the practitioner has undertaken the correct investigation, followed client's requirements and ultimately completed those tasks on the relevant exhibits which were required. In most cases, this review is a simple administrative task and likely to be the least resource-intensive to implement and quickest to perform.

3.2.2. The ‘proof check’ and ‘sense review’ (levels 2 & 3)

The ‘proof check’ and ‘sense reviews’ are low-labor-intensive approaches to review, and arguably require limited technical knowledge to carry out as the reviewer is simply ensuring that the report is written in a manner which is acceptable. Page et al (2019, p. 89) define these review types as follows; and this work sees no reason to expand these formal definitions:

> “Sense review: A check of work to ensure it makes sense as a piece of deliverable evidence. Technically, no checks of evidence or evaluation take place.”

> “Proof check: Here, the work is proofread purely for grammatical and spelling issues.” (Page et al., 2019, p.89, p.89)

These approaches to peer review may be more appealing to organisations as they come at a relative low cost in terms of resourcing, but the issue here lies with a potential misunderstanding of the value and need for peer review. Whilst an expectation of peer review would be to assess the quality of the ‘written word’, it is at its heart a process of error detection and correction. Organisations should not consider peer review to be a process that simply identifies and corrects grammatical errors and poorly written text (albeit part of the process), the ability to communicate effectively is a core and essential skill of all DF practitioners therefore errors at this level should only be minor. The problem is that ‘proof checks’ and ‘sense reviews’ will not effectively scrutinise work to determine scientific errors, arguably rendering these review types when used in isolation, of limited value to an organisation who seeks to use peer review for quality assurance. Instead, it is argued that ‘proof checks’ and ‘sense reviews’ should be absorbed into more rigorous review approaches and form part of a peer review, but not be considered as the peer review itself. Despite the fact that ‘proof checks’ and ‘sense reviews’ are easy to implement, organisations should resist the temptation to consider them as their primary error checking mechanism, where arguably peer review mechanism should, as a baseline standard, be at least reviewing at a ‘conceptual’ (Level 4) peer review level.

3.2.3. The ‘conceptual review’ (level 4)

Moving up the review hierarchy, Page et al (2019, p. 89) define a ‘conceptual peer review’ as follows:

> “Conceptual peer review: This type of review focuses on an individual's description of evidence artefact types, ensuring that they have documented the true interpretation of what they do. Here, there is no check on quantification of evidence, the collection process is assumed to be correct. Instead, it is a check and evaluation of whether the practitioner has conveyed the true understanding of specific evidence types.”

Page et al.’s (2019) ‘conceptual peer review’ definition indicates this review type begins to question the science underpinning the report, a positive step towards developing a review process which improves an organisation's chances of error mitigation in practitioner work. However, this review also starts to provide an organisational burden. To assess evidence descriptions and interpretations requires understanding, knowledge and expertise. Whilst the conceptual peer review description offered by Page et al. (2019) does not state it, it appears implied (and this work would like to provide backing to this position) that those conducting conceptual peer reviews will almost certainly have to be more experienced (or in a worst case scenario, of equal experience) than the principal examiner in order to be in a position to provide effective scrutiny. As a result, this form of peer review is potentially more costly in time and resources as it requires senior staff to conduct the review and additional time to assess the report's content at a ‘technical meaning’ level. Despite improving the thoroughness of the review process, arguably the conceptual peer review does not go far enough in terms of being an effective peer review. Conceptual peer review is an extensive peer review based on the practitioner's documentation derived from the analysis. This means that the reviewer cannot replicate experimental methods or data, and must use their professional and scientific expertise to determine if the documented experimental design, methods, results and conclusions appear valid.

3.2.4. The ‘sampled & full verification review’ (levels 5 & 6)

At the verification review stage, this work proposes a deviation from that noted by Page et al (2019, p. 89) who define a ‘verification review’ as follows:

> “Verification review: A full validation review is in essence a second examination carried out across the principal examiners data set only, not all of the acquired digital data, with prior knowledge of the primary practitioners' results set. It allows steps to be re-traced, hypotheses to be confirmed and findings to be validated. In doing so, the chance of misinterpretation is arguably lessened.”

The described ‘verification’ peer review focuses on an evaluation of the practitioner data set (data and files recovered or parsed and deemed evidential). In essence, the verification review examines the practitioner's dataset that has been identified as part of the investigation. In doing so the reviewer can scrutinise a practitioner's interpretation of data to a greater degree by being able to look back at the original data and form their own judgement. This very process helps to identify errors arising from a ‘non-technical’ source (essentially, the practitioner - all forms noted previously in Section 2.2), but may do very little in the way of identifying technical errors as such errors will have occurred prior to the generation of the dataset which a reviewer is reviewing. Nevertheless, verification reviews do expose the practitioners work to a greater degree of scrutiny, but their implementation requires more effort. To review actual investigative work carried out by the practitioner (albeit likely a smaller subset of data than the total size of all exhibited data), this does require time and effort. There will be an inevitable slowing of the turnover of work carried out by practitioners within the lab setting, where a bottleneck at the reviewing stage may begin to be witnessed, particularly in smaller organisations who may be reliant on a small number of employees who are qualified enough to carry out a verification review (or which this itself may cause issues if there are not enough staff members in an organisation for the distribution of reviewer responsibilities).

This article proposes to split the verification stage into two, a ‘Sampled’ and a ‘Full Verification Review’. The difference here lies with the resources it takes to carry out this process, where sampled review is arguably less burdensome, but not as robust in terms of a peer review method.

3.2.5. The 're-examination' (level 7)

Finally, the proposal of a re-examination review is possibly considered the 'gold standard' and is described by Page et al (2019, p. 89) as:

> "Re-examination: A complete re-examination of the original case and cross-referencing of results. The process is completed blind, followed by a comparative review and discussion of findings between the primary and reviewer's investigation. Such a process is utilised in other forensic science disciplines to combat forms of cognitive bias and therefore offers potential benefit to DF. Yet the resource overheads in terms of time and cost may make this option unfeasible."

A blind re-examination by a suitably qualified practitioner followed by a 'comparative review and discussion of findings between the primary and reviewer's investigation' is possibly one of the most robust mechanisms of peer review. Yet, the costing impact for organisations means that in the field of DF, this approach is likely to be infeasible to consider. It may be argued that such an approach is not required as the opposing party (prosecution/defense) carries out such scrutiny, yet it is difficult to support this stance. In addition, in some jurisdictions those operating in a defense capacity are not permitted to have access to all of the original case data, therefore effectively removing the option for carrying out a re-examination, and in essence are limited to a verification-styled examination of the practitioner's findings. Whilst some investigations in more traditional forms of forensic science may involve tests, which are quick to run and analyse results, in DF, examinations of data can take multiple days or weeks. The re-running of such processes followed by the time it takes a second practitioner to examine and interpret this data will in almost all cases be too expensive to implement, leading to many organisations compromising from a review which falls in line with those previously noted as further down the review hierarchy.

The DF investigation would never operate on its own. The work would be commissioned by the prosecution/investigation team, and the starting point would be a mandate or task description. This should also be the starting point when deciding the scope of the peer review.

All of the stated peer review forms above (except re-examination) focus on an end goal, that of the reported findings. Whilst in some sense this appears logical and efficient to approach peer review in this way, it is argued that this is not the ideal process. Such a process is reactive, and whilst this is somewhat the goal of peer review (to react and evaluate work post-completion), it is argued that it is better to prevent errors than simply detect them.

An acknowledgement must also be made with regards to the natural peer review, which takes place in many jurisdictions in regards to the prosecution/defence relationship. Prosecution evidence is often scrutinized by defence experts to ensure a fair legal process. In England and Wales, initiatives such as Streamlined Forensic Reporting (SFR) encourage discourse between both prosecution and defence experts in an effort to agree findings or identify points of conflict. This process carries an inherent level of scrutiny and review of the investigatory process where legal obligations exist on both parties to support the court in their pursuance of justice. In essence, in absence of any formalised peer review initiative, a fallback position may be considered to be this relationship between parties. However, it is argued that this is not wholly an acceptable position to be in where defence experts may not always be best positioned or effectively resourced to carry out effective scrutiny. As a result, we argue for the purpose of quality assurance and control that internal peer review processes should also exist, without placing the sole burden on the defence to discover errors.

4. Taking things one step further

During the investigative journey of a typical case, the practitioners usually drive the examination process on their own up until the point where they believe they have completed their examination. This work argues that this reactive approach to peer review may not provide the most efficient approach. Therefore, it is proposed that peer review structures must extend into the investigative process in order to support the practitioner from the start of their casework. Fig. 2 provides an overview over the DF investigation process where the supervision and advice is included prior to the subsequent review stage.

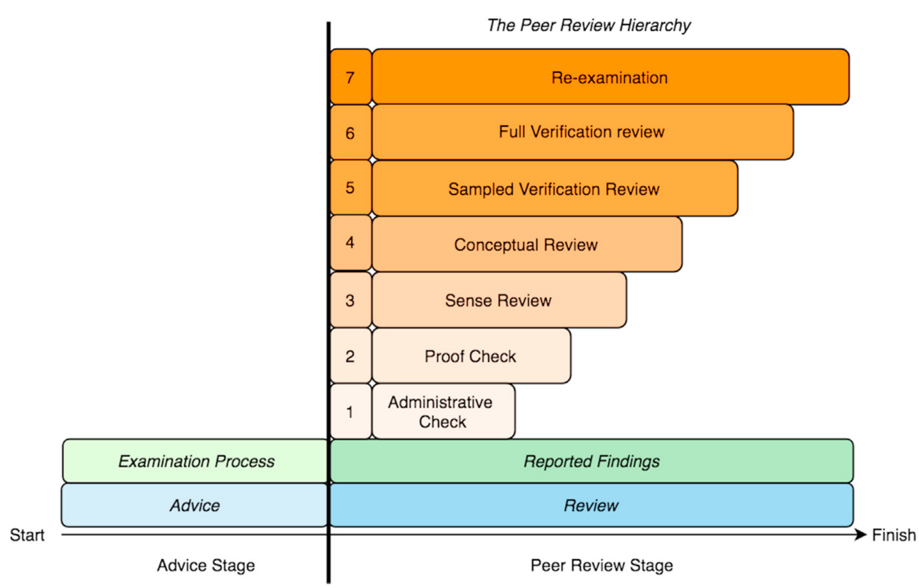

Fig. 2. The advice and peer review stages of the investigative journey.

It is suggested that it is important that practitioners are supported in their investigative decision making at all stages of their cases and that these decisions are exposed to evaluation, preventing any errors early in the investigative process from impacting later work. As a result, peer review should not be confined to the last stage of an examination's journey but form part of a methodology which contains 'advice' stages earlier in the process and throughout each stage of an investigation up until the point of submitting work for review.

In Part 2 (Sunde and Horsman, 2020), the authors offer the Phase-oriented Advice and Review Structure (PARS) for the peer review of digital forensic investigations. PARS is a methodology for peer review which wraps itself around the DF investigatory process combining advice checkpoints and peer review to provide a robust and holistic approach for assessing the quality of practitioner case work and error detection.

5. Conclusion

Peer review mechanisms across all fields and areas have always maintained critics, with the same being true here in DF. What is important to note is that peer review is an important mechanism for DF organisations and it should not be discarded or treated with disdain. We recognise that implementing effective peer review carries a resource burden, but it is argued that this is worth it to ensure quality standards in a given laboratory are maintained. Peer review is arguably the gatekeeper for most organisations, preventing potential erroneous and therefore damaging work from returning to clients and entering legal systems. It is therefore vital that peer review systems are not only in place, but that they are sufficiently rigorous in design and implementation to prevent them from being offered only as a token gesture.

This work has set out the need for peer review in DF, and the benefits that it can bring to the field in terms of quality assurance and error mitigation. The contributions noted here in Part 1 are threefold:

1. An evaluation of the error sources in DF, both technical and non-technical.
2. A critical evaluation of peer review as a concept, in the context of DF, highlighting its potential pitfalls and flaws.
3. The 'Peer Review Hierarchy' is offered, building upon the past work of Page et al. (2019), which offers and defines seven levels of peer review which can be applied to a practitioner's findings.

In this piece, we have discussed the various approaches to peer review available for evaluating presented findings in a practitioner report. The 'Peer Review Hierarchy' outlines seven methods of peer review, which range in depth and practicality, where an organisation can determine which best fits their operating practices. Yet we also emphasise that the actual review of a practitioner's report itself should only form part of an effective peer review process. Proposals are made for the need for a peer review methodology which is structured, documentable, and one which not only evaluates any presented findings, but also offers an advisory role throughout the investigation process and accountability for all those involved in it. It is here that we offer the Phase-oriented Advice and Review Structure (PARS) for DF investigations (Sunde and Horsman, 2020), which is presented and discussed in Part 2.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We are grateful to the anonymous reviewers whose helpful comments, suggestions and constructive criticism have improved the content and sharpened the focus of this article.

This paper was funded by a grant from the Norwegian Police University College awarded to the second author for pursuing a PhD.

References

Appelbaum, P.S., 1992. Forensic psychiatry: the need for self-regulation. J. Am. Acad. Psychiatr. Law 20 (2), 153–162.
Årnes, A. (Ed.), 2018. Digital Forensics. Wiley, Hoboken.
Ask, K., 2013. Bias: fejl og faldgruber i efterforskning. In: Hald, C., Vrist Rønn, K. (Eds.), Om at Opdage: Metodiske Refleksioner over Politiets Undersøgelsespraksis, pp. 149–169. Frederiksberg C: Samfundslitteratur.
Ask, K., 2006. Criminal Investigation: Motivation, Emotion and Cognition in the Processing of Evidence. Doctoral Dissertation. Department of Psychology, Gothenburg University.
Ballantyne, K.N., Edmond, G., Found, B., 2017. Peer review in forensic science. Forensic Sci. Int. 277, 66–76. https://doi.org/10.1016/j.forsciint.2017.05.020.
Budowle, B., Bottrell, M.C., Bunch, S.G., Fram, R., Harrison, D., Meagher, S., Oien, C.T., Peterson, P.E., Seiger, D.P., Smith, M.B., Smrz, M.A., 2009. A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. J. Forensic Sci. 54 (4), 798–809. https://doi.org/10.1111/j.1556-4029.2009.01081.x.
Casey, E., 2011. Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet. Academic press, Amsterdam.
Christensen, A.M., Crowder, C.M., Ousley, S.D., Houck, M.M., 2014. Error and its meaning in forensic science. J. Forensic Sci. 59 (1), 123–126. https://doi.org/10.1111/1556-4029.12275.
Cohen, F., 2010. Toward a science of digital forensic evidence examination. IFIP Adv. Inf. Commun. Technol. 17–35. https://doi.org/10.1007/978-3-642-15506-2_2.
Daniel, H.D., Mittag, S., Bornmann, L., 2007. The potential and Problems of peer evaluation in higher education and research. Quality assessment for higher education in Europe 71–82.
Doyle, S., 2019. Quality Management in Forensic Science. Academic Press, London.
Dror, I.E., 2020. Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. Anal. Chem. 92, 7998–8004. https://doi.org/10.1021/acs.analchem.0c00704.
Dror, I.E., 2017. Human expert performance in forensic decision making: seven different sources of bias. Aust. J. Forensic Sci. 49 (5), 541–547. https://doi.org/10.1080/00450618.2017.1281348.
ENFSI, 2015. Best Practice Manual for the Forensic Examination of Digital Technology, ENFSI-BPM-FOT-01 (November 2015).
Hamer, D.A., Edmond, G., 2019. Forensic Science Evidence, Wrongful Convictions and Adversarial Process. Forensic Science Evidence, Wrongful Convictions and Adversarial Process, vol. 38. University of Queensland Law Journal, p. 2.
Hand, D.J., 2004. Measurement Theory and Practice: the World through Quantification. Arnold, London.
Heilbrun, K., Rogers, R., Otto, R.K., 2004. Forensic assessment. In: Taking Psychology and Law into the Twenty-First Century. Springer, Boston MA, pp. 119–146. https://doi.org/10.1007/0-306-47944-3_4.
Henderson, M., 2010. Problems with peer review. Br. Med. J. 340, c1409. https://doi.org/10.1136/bmj.c1409.
Horsman, G., 2018. "I couldn't find it your honour, it mustn't be there!"–Tool errors, tool limitations and user error in digital forensics. Sci. Justice 58 (6), 433–440. https://doi.org/10.1016/j.scijus.2018.04.001.
Horsman, G., 2019. Tool testing and reliability issues in the field of digital forensics. Digit. Invest. 28, 163–175. https://doi.org/10.1016/j.diin.2019.01.009.
Horsman, G., 2020. Part 1:-quality assurance mechanisms for digital forensic investigations: introducing the Verification of Digital Evidence (VODE) framework. Forensic Sci. Int.: Report 2, 100038.
International Organization for Standardization, 2017. ISO/IEC 17025:2017 (EN) General Requirements for the Competence of Testing and Calibration of Testing and Calibration Laboratories.
International Organization for Standardization, 2015. ISO 9000:2015 (EN) Quality Management Systems - Fundamentals and Vocabulary.
International Organization for Standardization, 2013. ISO/IEC 27001:2013 (EN) Information Technology — Security Techniques — Information Security Management Systems — Requirements.
International Organization of Standardization, 2012. ISO/IEC 27037:2012 (EN) Information Technology — Security Techniques — Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence.
Ivanovic, A.B., 2019. Accreditation and standardization in forensic science - present and future reliability of forensics examination and expertise. JE-Eur. Crim. L. 200.
Jonakait, R.N., 1993. The meaning of Daubert and what that means for forensic science. Cardozo Law Rev. 15, 2103.
Keim, P., 2003. Microbial Forensics: a Scientific Assessment (No. DOEER63403).

American Society for Microbiology (US).
Kukucka, J., Kassin, S.M., Zapf, P.A., Dror, I.E., 2017. Cognitive bias and blindness: a global survey of forensic science examiners. Journal of Applied Research in Memory and Cognition 6 (4), 452–459. https://doi.org/10.1016/j.jarmac.2017.09.001.
Meyers, M., Rogers, M., 2004. Computer forensics: the need for standardization and certification. International Journal of Digital Evidence 3 (2), 1–11.
Meyers, M., Rogers, M., 2005. February. Digital forensics: meeting the challenges of scientific evidence. In: IFIP International Conference on Digital Forensics. Springer, Boston, MA, pp. 43–50. https://doi.org/10.1007/0-387-31163-7_4.
Nakhaeizadeh, S., Dror, I.E., Morgan, R.M., 2014. Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias. Sci. Justice 54 (3), 208–214. https://doi.org/10.1016/j.scijus.2013.11.003.
Obenson, K., Wright, C.M., 2013. The value of 100% retrospective peer review in a forensic pathology practice. Journal of forensic and legal medicine 20 (8), 1066–1068. https://doi.org/10.1016/j.jflm.2013.09.033.
Page, H., Horsman, G., Sarna, A., Foster, J., 2019. A review of quality procedures in the UK forensic sciences: what can the field of digital forensics learn? Sci. Justice 59 (1), 83–92. https://doi.org/10.1016/j.scijus.2018.09.006.
Pollitt, M., 2008. January. Applying traditional forensic taxonomy to digital forensics. In: IFIP International Conference on Digital Forensics. Springer, Boston, MA, pp. 17–26. https://doi.org/10.1007/978-0-387-84927-0_2.
Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., et al., 2006. An intervention to decrease catheter-related bloodstream infections in the ICU. N. Engl. J. Med. 355 (26), 2725–2732. https://doi.org/10.1056/nejmoa061115.
Roux, C., 2019. Forensic Science Isn't 'reliable' or 'unreliable': it Depends on the Questions You're Trying to Answer (Accessed January 21, 2020). https://phys.org/news/2019-09-forensic-science-isnt-reliable-unreliable.html.
Sims, D.N., Langlois, N.E., Byard, R.W., 2013. An approach to peer review in forensic pathology. Journal of Forensic and Legal Medicine 20 (5), 402–403. https://doi.org/10.1016/j.jflm.2013.02.010.
Smit, N.M., Morgan, R.M., Lagnado, D.A., 2018. A systematic analysis of misleading evidence in unsafe rulings in England and Wales. Sci. Justice 58 (2), 128–137. https://doi.org/10.1016/j.scijus.2017.09.005.
Smith, R., 1988. Research Policy: Problems with peer review and alternatives. Br. Med. J. 296 (6624), 774. https://doi.org/10.1136/bmj.296.6624.774.
Sunde, N., 2017. Non-technical Sources of Errors when Handling Digital Evidence within a Criminal Investigation (Master's Thesis). Norwegian University of Science and Technology. Available at: http://hdl.handle.net/11250/2450280.
Sunde, N., Horsman, G., 2020. Part 2: The Phase-oriented Advice and Review Structure (PARS) for Digitial Forensic Investigation. Forensic Science International: Digital Investigation (Accepted 19 september 2020).
SWGDE, 2018. Establishing Confidence in Digital and Multimedia Evidence Forensic Results by Error Mitigation Analysis. Version: 2.0, (November 20, 2018). Available at: https://www.swgde.org/documents/Current%20Documents/SWGDE%20Establishing%20Confidence%20in%20Digital%20Forensic%20Results%20by%20Error%20Mitigation%20Analysis (Accessed March 31, 2020).
Tully, G., 2019. Annual report. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786137/FSRAnnual_Report_2018_v1.0.pdf (Accessed March 22, 2020).
Tully, G., 2020. Annual report. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868052/20200225_FSR_Annual_Report_2019_Final.pdf (Accessed March 22, 2020).
Tully, G., Cohen, N., Compton, D., Davies, G., Isbell, R., Watson, T., 2020. Quality standards for digital forensics: learning from experience in England & Wales. Forensic Sci. Int.: Digit. Invest., 200905 https://doi.org/10.1016/j.fsidi.2020.200905.
Watson, D.L., Jones, A., 2013. Digital Forensics Processing and Procedures: Meeting the Requirements of ISO 17020, ISO 17025, ISO 27001 and Best Practice Requirements. Elsevier, Amsterdam.
Welner, M., Mastellon, T., Stewart, J.J., Weinert, B., Stratton, J.M., 2012. Peer-reviewed forensic consultation: safeguarding expert testimony and protecting the uninformed court. J. Forensic Psychol. Pract. 12 (1), 1–34. https://doi.org/10.1080/15228932.2011.588526.
Welner, M., Davey, E.E., Bernstein, A., 2014. Peer-reviewed forensic consultation in practice: multidisciplinary oversight in common expertise. J. Forensic Sci. 59 (5), 1254–1259. https://doi.org/10.1111/1556-4029.12470.