Science & Justice 65 (2025) 139–148

Contents lists available at ScienceDirect

Science & Justice

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

Evaluating the scope of peer review in digital Forensics: Insights from Norway and the U.K.

Rune Kenneth Bauge [a], Elénore Ryser [b,*], Nina Sunde [c], Graeme Horsman [b]

[a] Vest politidistrikt, Allehelgens gate 6, Bergen 5016, Norway
[b] Crandfield Forensic Institute, Cranfield University, 33b College Rd, Wharley End, Bedford MK43 0AL, United Kingdom
[c] Norwegian Police University College, Slemdalsveien 5, Oslo 0369, Norway

ARTICLE INFO

Keywords:
Digital Forensic Science
Peer Review
Digital evidence
Quality assurance
Quality control

ABSTRACT

This paper investigates the implementation and utilisation of peer review practices in digital forensics (DF) within Norway and the U.K. Through a comprehensive survey of 113 DF practitioners and managers, we explore the extent to which peer review is integrated into DF investigations and the variations in practices between these two countries. Our findings reveal that while both Norway and the U.K. recognize the importance of peer review in ensuring the integrity and accuracy of DF work, there is a tendency to limit peer reviews to the examination of reports, rather than extending them to more thorough verification of results and methodologies. Utilising the Peer Review Hierarchy for DF as an analytical framework, our study highlights a significant gap in the depth of peer review practices, with both countries primarily focusing on lower-level reviews that are less likely to detect critical errors. The paper discusses the implications of these findings in the field of DF, emphasising the need for more robust and comprehensive peer review mechanisms to enhance the quality and reliability of digital evidence. Furthermore, we discuss the systemic and resource-related challenges that may hinder the implementation of more extensive peer review practices.

1. Introduction

Digital evidence is obtained in most criminal investigations today [1,2]. It is therefore of paramount importance that such evidence is accurate and derived from robust forensic processes where numerous works have highlighted the risk of erroneous or misleading evidential findings resulting from DF investigations e.g. [3–8]. Therefore it is essential that all work conducted during these stages is subjected to a quality review.

Page et al. [9] compared the quality procedures in the DF domain to those in other forensic science disciplines, such as DNA, fingerprinting and body fluid examination, revealing significant deficiencies in the DF quality processes. They highlighted that existing quality systems were mainly directed towards laboratory environments, which represent only a small segment of the digital evidence production chain. Emphasising the need for verification of key findings, they called for a more nuanced peer review approach in DF, transcending the conventional administrative-technical review dichotomy. In this traditional paradigm, administrative review pertains to management and control of documentation, while the technical review centres on the validation of findings [10,11]. Horsman and Sunde expanded this idea in two papers [12,13], proposing the Phase-Oriented Advice and Review Structure (PARS) as a comprehensive quality assurance framework. PARS aims to prevent errors and optimise decision making throughout the DF process by introducing checkpoints, peer advisor supervision, and quality control through peer review by a qualified peer reviewer. The Peer Review Hierarchy for DF (the Hierarchy) was proposed as a more detailed approach to the scope of peer review.

Our paper examines quality management in DF in both the U.K. and Norway, focusing specifically on peer review practices. Although both countries conduct similar DF examinations, they have adopted fundamentally different approaches to standardisation and organisational requirements for such work in criminal investigations. The U.K. system is characterised by a high level of formal standardisation, often linked to ISO accreditation, and is largely supported by private sector forensic services. In contrast, DF work in Norway is primarily conducted by law enforcement through in-house units, which typically operate with minimal standardisation and use non-accredited laboratories. While several law enforcement DF units in the U.K. are transitioning to full accreditation, some have only achieved partial accreditation to date. It is

* Corresponding author.
E-mail address: elenore.ryser@cranfield.ac.uk (E. Ryser).

https://doi.org/10.1016/j.scijus.2025.01.005
Received 4 October 2024; Received in revised form 6 January 2025; Accepted 28 January 2025
Available online 6 February 2025

worth mentioning that, while there are no specific validation tests required for traces to be admitted as evidence in U.K. common law, they undergo various checks. These include internal quality management system checks, the judge's approval of the evidence to be presented, judicial evaluation by jury members, and cross-examination. However, this paper does not aim to classify the importance or relevance of these processes, nor does it seek to evaluate whether expert testimony should be considered a weak or strong epistemic authority. The focus on peer review as a validation methodology results from the need of DF processes to adhere to DF science principles, where peer review is commonly used in science to validate outputs.

The PARS framework, and particularly the Hierarchy forms the analytical basis of the survey (Fig. 1). The Hierarchy is a flexible model comprising seven levels of peer review, ranging from the least resource-intensive administrative checks to the most comprehensive re-examination of the evidential dataset. The initial levels include administrative checks, proof checks, sense reviews and conceptual reviews, all of which focus solely on the report. More resource-intensive levels, such as sampled and full-verification reviews, extend beyond the report to involve an examination of the evidential dataset. At the top level is a re-examination, requiring a second, full analysis of the dataset by an independent examiner not previously involved in the case. While lower-level reviews are less resource-intensive, they are also less effective at detecting critical errors. It is assumed that a more thorough assessment of the findings and interpretations increase the likelihood of uncovering errors and misleading conclusions, with partial or full verification reviews further enhancing error detection. The Hierarchy was utilised in the survey to determine not only the frequency of peer review activities but also the depth of these quality control measures.

This study addresses two key points, (1) the extent to which peer review is implemented and utilised in DF investigations in Norway and the U.K. and (2) how peer review practices may differ between the two countries, including possible explanations for these differences.

While previous studies on quality management have been conducted, as discussed in the background section, and some papers have touched on peer review, no published research has yet explored its practical implementation in DF units. While peer review in DF may seem well-documented, much of the existing literature focuses on theoretical frameworks and high-level quality systems, rather than on the practical details of how peer review is conducted. This study fills that gap by offering a detailed, practical investigation, making a novel contribution to a field often dominated by theoretical discussions. Moreover, its comparative analysis between Norway and the U.K. adds a new dimension to the peer review conversation, providing insights into how legal frameworks and resource allocation affect peer review implementation.

Our evaluation of these points is based on a survey conducted among DF practitioners and managers in both countries, with a total of 113 participants – 60 from Norway and 53 from the U.K.. The Norwegian part of the survey was conducted as part of Bauge's [14] Master's thesis, and the findings are further explored here for comparative purposes. The survey aimed to gather insights into their attitudes toward quality control and the application of peer review practices. Following the survey analysis, we discuss the findings, explore their implications for the DF field and suggest potential for further research.

2. Background

2.1. DF investigation in the U.K

In the U.K., law enforcement is overseen by 45 regional forces (43 in England and Wales, the Police Service in Northern Ireland and Police Scotland) and 3 specialised forces (Civil Nuclear Constabulary, Ministry of Defence Police and the British Transport Police) [15]. According to the Home Office, there are approximately 227,000 employees in the police (full-time equivalent), thereof 142 000 police officers in England and Wales as of 2022 [16].

Due to the complexity of the model used in England and Wales for providing DF services, it is difficult to determine the number of professionals employed in this industry. At a general level, the Forensic

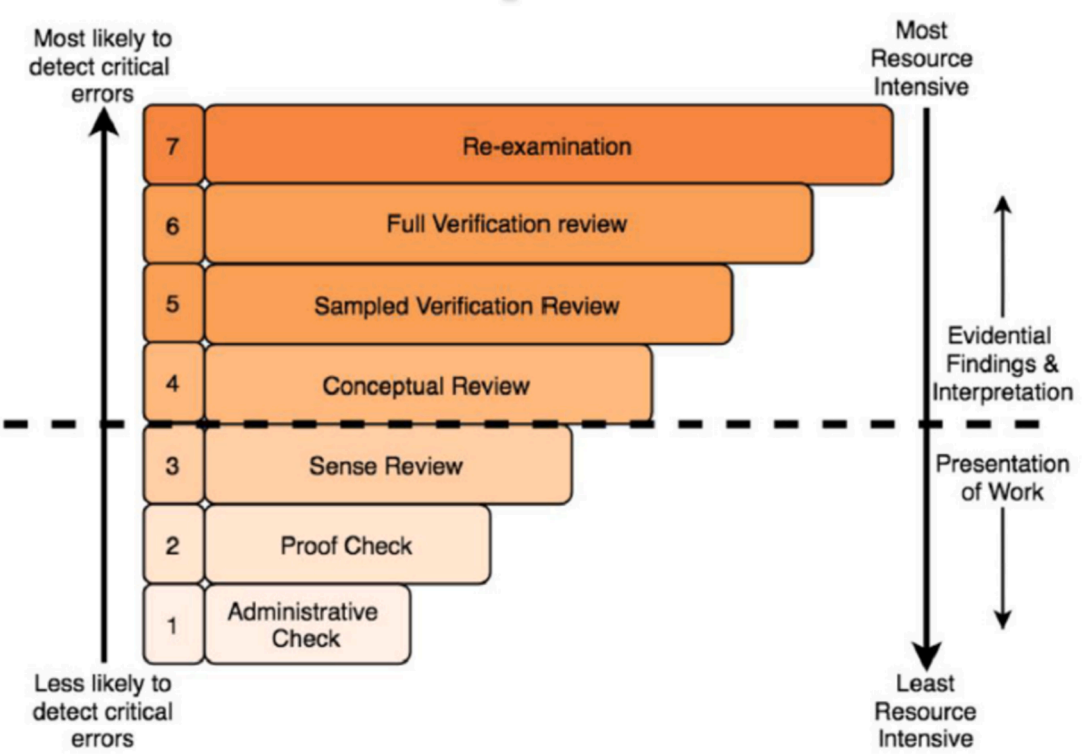

Fig. 1. The Peer Review Hierarchy for DF as presented in [12].

Capability Network [17] suggested around 4,000 individuals were working in forensic roles in England and Wales; however, this includes all branches of forensic science. At a force-level, the Metropolitan Police have acknowledged, in response to a freedom of information request in 2023, that *"103 police officers and 275 police staff undertaking a wide range of DF tasks"* [18], Q1. Additionally, a report by His Majesty's Inspectorate of Constabulary and Fire & Rescue Services [19] into DF provision in the U.K. noted that demand for these services outweighs the resources currently available.

The deployment of DF services within criminal investigations often involves professionals from both government agencies and/or individuals from the private sector. Regional forces may maintain their own dedicated DF unit, which is tasked with providing DF services to its own catchment area. However, forces may also seek to outsource excess or specialist work to private sector DF organisations who may conduct examinations on their behalf. Since 2017, the Codes of Practice and Conduct mandated accreditation to ISO/IEC 17025 for all DF laboratories involved in criminal investigations.

A literature review performed by Bauge [14] revealed that there is little empirical research on peer review practices within the DF domain. Although formalised systems for peer review exist and are part of the requirements of accredited laboratories (e.g., ISO 17025), little is known about how and when peer review is performed in DF units. Tully et al. [20] examined the implementation and compliance with ISO 17025 in England and Wales by analysing UKAS's initial assessments from 2015 to 2019 and quality referrals to the Forensic Science Regulator from 2012 to 2019. They concluded that accreditation to appropriate standards provides external assurance of an organisation's competence to produce reliable results within the accredited scope, but emphasised that accreditation alone does not eliminate errors. Similarly, Wilson-Kovacs and Wyatt [21] examined the quality standards implementation in England and Wales, identifying significant challenges in adopting these standards due to organisational preparedness. They point out that cost, time and resources pose considerable obstacles. While standards are crucial in DF, they argue that they cannot substitute the need for scientific and analytic rigour and reliable methods.

2.2. DF investigation in Norway

While some external laboratories provide services to private customers as well as the police, the majority of DF work for criminal investigations in Norway is conducted by internal DF examiners within local and national police units. Unlike the U.K., no DF unit or police DF laboratory is accredited (in part or full) to ISO 17025, nor is there a statutory requirement for such accreditation.

The Norwegian police, operating as a unified force, is divided into 12 police districts and 10 national or specialised services, employing around 18, 000 people as of November 2023 [22]. Each police district has its own DF unit, while specialised services like The Norwegian Criminal Investigation Service (NCIS – or Kripos in Norwegian) and the National Authority for Investigation and Prosecution of Economic and Environmental Crime (Økokrim in Norwegian) handle advanced DF tasks. The minimum qualification for DF practitioners is completing the NCFI Core course, which provides 15 ECTS credits in digital evidence handling. DF practitioners with civil backgrounds are also required to complete a course on criminal investigation regulations to gain limited police authority.

The DF work performed in criminal investigation is regulated by the Norwegian criminal procedure code, but no formalised guidelines exist on practitioner competency or procedures. Jahren [23] found a lack of systematic quality management across three police districts, with peer review initiated only by practitioners and typically focused on spelling and grammar issues rather than the evidential findings. This lack of quality management is further emphasised by Stoykova et al. [24] who highlight issues with the chain of custody. Their analysis of DF reports from 21 Norwegian criminal cases revealed that digital evidence could not be traced back to its source, stressing the need for better quality assurance.

2.3. The peer review Hierarchy for DF

The survey utilised in this work referenced the Hierarchy proposed in the papers by Horsman and Sunde [12,13] to identify the levels of quality control and frequency of review activities performed by the participants. The Hierarchy was regarded as a suitable method for quality control, as it is specifically tailored to the unique requirements of DF. It provides a comprehensive, checklist-based framework that guides the quality control process across various hierarchy levels. Importantly, this method does not replace other essential quality assurance measures, such as competence requirements, proficiency testing, and the validation of tools and methodologies. Instead, it serves to complement these measures, contributing to a holistic and robust approach to quality assurance in DF practices. A more detailed explanation of the Hierarchy is provided below.

Level 1, the 'Administrative check' assesses if the practitioner fulfilled the investigation requirements, met client specifications, and completed necessary tasks on relevant exhibits. It is a straightforward administrative process, requiring minimal resources and time.

Level 2, the 'Proof check' focuses on reviewing reports for spelling and grammatical errors. This low labour approach ensures that the document is written in an acceptable manner, emphasising a straightforward proofreading process for language-related issues.

Level 3, the 'Sense review' assesses the clarity and coherence of the report, and ensures the report makes sense as a deliverable piece of evidence without delving into the evaluation of the evidence itself. It's a low-labour process, suitable for individuals with limited technical knowledge, aiming to meet acceptable standards of clarity and organisation.

Level 4, the 'Conceptual Review' is a thorough evaluation of the report's content, focusing on the scientific and logical foundation without verifying the findings. It examines the relationship between the evidence and conclusions, assessing the soundness of the report's conceptual aspects. Conducting a conceptual review can be resource-intensive and requires reviewers with expertise and experience equal to or greater than the principal examiner. It involves assessing evidence descriptions, interpretations, and the overall validity of the documented experimental design, methods, results, and conclusions. This level of peer review relies on professional and scientific expertise as it cannot replicate experimental methods or data.

Level 5, 'Sampled Verification Review' entails verification of selected findings using a different tool or methodology than that was used during the original examination. This type of review focuses on evaluating the practitioner's dataset and scrutinising their interpretation of data, and may uncover implementation errors in tools, and helps to identify whether the practitioner may have misinterpreted data. A limitation is, however, that the approach only follows up on the findings and does not assess whether the practitioner may have overlooked relevant information.

Level 6, 'Full Verification Review' involves a similar approach as for level 5; however, here, all reported results are assessed. As with level 6, only findings are reviewed, and the procedure has limited potential to uncover additional relevant findings.

Level 7, 'Re-examination' involves a complete examination of the case by personnel with no prior knowledge or involvement. It is considered the most robust form of peer review in the Hierarchy and has the potential to uncover additional relevant findings. However, it is also the most resource-intensive approach, and the time and resources required for a second practitioner to examine and interpret the data might render this approach feasible only in extraordinary situations.

As indicated in Fig. 1, the lower levels of the Hierarchy are less resource-intensive compared to the upper levels. At the same time, the upper levels are more likely to detect critical errors than the lower

levels. It is also assumed that peer review in levels 1–3 may be performed by personnel without DF expertise. Level 4 would require the peer reviewer to have similar or higher competency than the reviewee, and the level above would require higher expertise since it would involve verification of findings.

3. Method

3.1. Design and distribution

This paper presents the results of a survey which was conducted anonymously with informed consent during September 2022-June 2023, starting in Norway, and then in the U.K. In Norway, approval for handling personal data was obtained from the Norwegian Agency for Shared Services in Education and Research (Sikt). In the U.K., data collection and storage adhered to the guidelines of the U.K. Data Protection Act 2018 following an internal ethical review. The survey was distributed directly by email or through networks such as the Forensic Capability Network and the Chartered Institute of Information Security. In the later stages, social media platforms were also used in an effort to increase participant number.

The survey, comprising multiple-choice, Likert scale, and open-text questions, was divided in three parts. The first part collected demographic data, such as education, experience, role, and police district or national unit affiliation. The second part focused on quality control procedures at the respondent's unit, examining the presence of quality procedures and documentation and the roles involved. The third part, referencing the Hierarchy, examined the frequency and scope of quality control activities. The levels of the Hierarchy were used as a basis for measuring the extent of quality control. An explanation of what each level entails was provided, allowing the survey to be conducted regardless of the participants' prior familiarity with the framework. In the U.K., all respondents answered the same set of questions, while in Norway, questions regarding the number of cases involving each level of quality control (see Fig. 8) were directed to DF managers, and questions about the frequency of quality control activities in the past 12 months (see Fig. 9) were posed to DF practitioners.

While the survey included some country-specific questions, this paper primarily focuses on the comparable matters. However, due to its relevance, the paper reports the U.K. respondents' definition of quality control in DF and key measures for its implementation (Fig. 9), as this question was added after collecting the Norwegian data.

3.1.1. Sample and population

The surveys received a total of 113 responses, with 60 from Norway and 53 from the U.K. In Norway, 50 respondents were DF practitioners, and 10 were managers. Based on the total population of 185 practitioners and managers in Norway as of March 2023 [14], this represents a response rate of approximately 32 %. In contrast, gathering responses in the U.K. proved more challenging, as reflected in the lower participation rate. Due to the lack of reliable data on the total number of DF practitioners and managers in the U.K., it is not possible to assess how representative the U.K. sample is to the wider community. With only 53 U.K. employees responding, the sample likely represents less than 1 % of the estimated DF practitioner population. The implications of these limitations will be addressed in the ensuing discussion.

3.2. Data consolidation

Due to differences in the organisational structure of DF work in Norway and the U.K., some survey questions were slightly varied. As a result, data needed to be reorganised for comparison, primarily by consolidating responses based on roles. The Norwegian survey offered a binary choice between 'Employee' or 'Manager', while the U.K. survey provided three roles: 'Practitioner', 'Manager', 'Technician', along with an 'Other' option.

To analyse the data effectively, employees in the Norwegian survey were considered as practitioners, aligning with the role of practitioners and technicians combined in the U.K. survey. The 'Other' category in the U.K. survey was subdivided into practitioners and managers based on respondent answers. For example, if a respondent indicated being a *"Supervisor but also a practitioner"* this was categorised as a managerial role. This underscores the diversity of DF roles in the U.K., presenting a divergence from Norway, which is further addressed in the discussion.

A discrepancy arose between the number of reported practitioners and managers and the data linked directly to the Hierarchy. While all Norwegian participants completed the survey, 22 U.K. did not. To maximise the use of the U.K. data, the survey was treated as two parts: The first encompassing demographics and peer review, and the second focusing on the Hierarchy. For easier comparison, percentages were used instead of absolute figures.

Further adjustments were made to harmonise frequency groupings, aligning Norway's 6–9 group with the U.K.'s 6–10, and consolidating U.K. data for 11–26 + under 10 + to match the scale used in Norway.

Finally, the levels of the Hierarchy were merged to simplify the interpretation and presentation of results: Levels 1–3 and 5–6 were grouped, and levels 4 and 7 remained separate due to their distinct functions.

4. Results

4.1. Demographics

The survey received responses from 60 participants in Norway, all of whom completed the survey. All Norwegian respondents were employed in law enforcement (Table 1), with an even distribution of 50 percent having a civilian background and 50 percent a police background. In the U.K., 53 participants responded, and of these, 31 respondents completed the survey.

Of the 53 U.K. participants, 46 were employed in law enforcement and 6 in the private sector, with most respondents having a law enforcement background (Table 1). Compared to the Norwegian participants, the U.K. participants have a more diverse educational background. In Norway, 77 percent of participants held an academic degree, with 37 percent having a Master's degree and 40 percent a Bachelor's degree. In contrast, one-fifth of the U.K. participants held a Master's degree, and twice this number had a Bachelor's degree. Concerning

Table 1
Demographic data.

Categories	*Norway*	*U.K.*	*Total*
Background (n = 113)			
Civil	*30*	*42*	*72*
Police	*18*	*11*	*29*
Police/additional background	*12*	*0*	*12*
Highest degree (n = 113)			
BSc	*12*	*17*	*29*
MSc	*11*	*10*	*21*
PhD	*0*	*1*	*1*
No answer/other education	*37*	*25*	*62*
Employment (n = 113)			
Law enforcement	*60*	*46*	*106*
Private sector	*0*	*6*	*6*
Other	*0*	*1*	*1*
Years of experience[a] *(n = 113)*			
0–2 years	*11*	*9*	*20*
3–5 years	*10*	*15*	*25*
6–8 years	*19*	*10*	*29*
9 + years	*20*	*19*	*39*
Role (n = 113)			
Employee	*50*	*33*	*83*
Manager	*10*	*16*	*26*
Other	*0*	*4*	*4*

[a] within the digital forensic discipline.

alternative educational backgrounds (Table 1), 30 Norwegian participants did not specify their education level, while 7 Norwegian and 25 U.K. participants reported education other than university level, such as school or college education. Additionally, in both Norway and the U.K., most participants held employee positions at the time of the survey (Table 1) and had more than six years of experience (Table 1).

4.2. General aspects of quality control

4.2.1. Attitude toward quality control

Overall, participants from both Norway and the U.K. strongly support integrating quality control measures into DF processes.

In Norway, most participants (52, n = 60) fully agreed, while a smaller proportion (8, n = 60) partially agreed that quality control should be integrated into the work process of the DF units (Fig. 2). Respondents justified the need for quality control to safeguard legal rights and prevent investigations from collapsing due to flawed digital evidence. However, one respondent commented that lack of resources made systematic quality control challenging.

Similarly, most U.K. participants (34, n = 40) fully agreed on the importance of quality control in DF casework. They emphasised that it should be integrated within the forensic process, not only before sending the information to court but also before sharing results with the wider investigative team (Fig. 2). Two participants argued that previous quality assurance efforts had revealed gaps that forensic tool providers did not adequately address, and that the adversarial court system does not effectively reduce the risk of insufficient quality control and its consequences.

Three U.K. respondents partially agreed, noting that while quality control enhances output standards, its effectiveness may be limited in cases where evidence presence on devices is uncertain, and triage is required. Another participant, adopting a neutral stance, acknowledges the importance of quality control but argued that it has been taken too far, stifling innovation and hampering practitioners' ability to use their experience to solve investigative problems.

While most agree on the need for quality control, there are notable differences in how they define its implementation (see Fig. 3). Two questions added in the U.K. survey asked respondents to define quality control and identify the most effective method to ensure it. Generally, most comments agree that the purpose of quality control is to ensure the quality of the final product. However, there is diversity in how they interpret the concept of 'quality'. Some respondents highlight accuracy, correctness, and reliability, while others emphasise meeting customer requirements, ensuring consistency, and repeatability. Additionally, some respondents highlighted the significance of clarity in the reports and the importance of finding and sharing sufficient information.

In defining quality control, several respondents mentioned Standard Operating Procedures (SOPs), adherence to standards such as ISO, and the adoption of best practices. Additionally, four participants described quality control as a means to preserve evidence integrity, while two others refer to it as a way to reduce risks and ensure a forensically sound process, though they did not specify what "forensically sound" means.

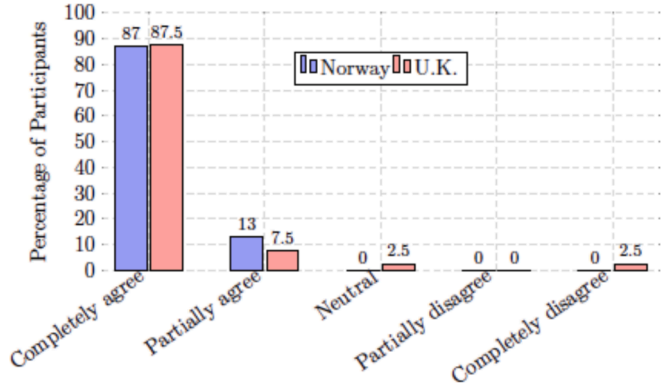

Fig. 2. The need to implement quality control within the work processes (NO-Q12 / UK –Q11) Answers in percentage.

When the U.K. participants were asked about the most effective method to ensure quality control, the majority mentioned peer review as a key component. However, their responses also included other measures, as shown in Fig. 3. The second most frequently cited approach was the implementation of audits. Among the various themes identified, tool validation or verification stood out as the most prominent. Furthermore, training and documentation were also mentioned as essential means to ensure quality control and integral to the overall quality control process.

4.2.2. Application of quality control

The survey reveals two contrasting scenarios regarding the application of systematic quality control. The majority (92.5 percent) of U.K. participants agreed or partly agreed that their unit had implemented a systematic quality control mechanism (Fig. 4). In contrast, only 22 percent of the Norwegian participants shared this view. Approximately three-fourths of the participants (U.K.: 72.5 percent, 29, n = 40, NO: 67 percent, 40, n = 60) indicated that they had initiated this process within the twelve months preceding the survey.

According to the Norwegian participants, the role of those performing quality control was divided between managers (41 percent, 7, n = 17), employees (53 percent, 9, n = 17) and prosecutors (6 percent, 1, n = 17). It is important to note, however, that 72 percent (43, n = 60) of participants did not answer that question (see Fig. 5). In the U.K., most participants (65 percent, 23, n = 40) indicated that employees typically handled quality control in their units (see Fig. 5).

However, eleven respondents described a different situation. One participant noted that Digital Crime Scene Managers were responsible for implementing quality control in their organisation. Eight others explained that roles were not clearly defined, with managers, practitioners, and technicians all involved in the processes. Two U.K. participants emphasised that peer review tasks were delegated based on expertise rather than rank, with specialists, such as those knowledgeable in Apple systems, being assigned relevant reviews regardless of their position.

The survey also highlights contrasting conditions in Norway and the U.K. regarding time and documentation for quality control. In Norway, only 27 percent (16, n = 60) of respondents partly or completely agreed they had enough time, while the majority (59 percent, 35, n = 60) partly or completely disagreed. Their free text responses indicated a divide in attitude towards the issue of time. Some viewed quality control as an additional task, separate from the workflow, while others emphasised that it was already an integral part, as one respondent noted: "*We always have time to do what we have to do depending on the situation*" [14], p. 42. In the U.K., despite the absence of dedicated roles, participants were evenly split between complete and partial agreement on having sufficient time for quality control implementation (Fig. 6). However, examining their free text answers reveals concerns about the volume and scope of peer review, with participants citing a lack of time and staff: *"Demand is so high that we don't have the time or staff to perform all of these reviews."* and *"Level 7 is time intensive and not possible with client deadlines. […]"*.

A similar pattern is observed in both the U.K. and Norway concerning the availability of sufficient documentation to guide the quality control process, such as routine and procedure descriptions (Fig. 7). In the U.K. the majority (82.5 percent) completely or partly agree that documentation is in place. In contrast, only 31 percent of Norwegian participants share this view, with over two thirds (62 percent) partly or completely disagreeing about the availability of such documentation.

4.3. The scope of quality control

As noted earlier, this section of the survey had fewer responses from

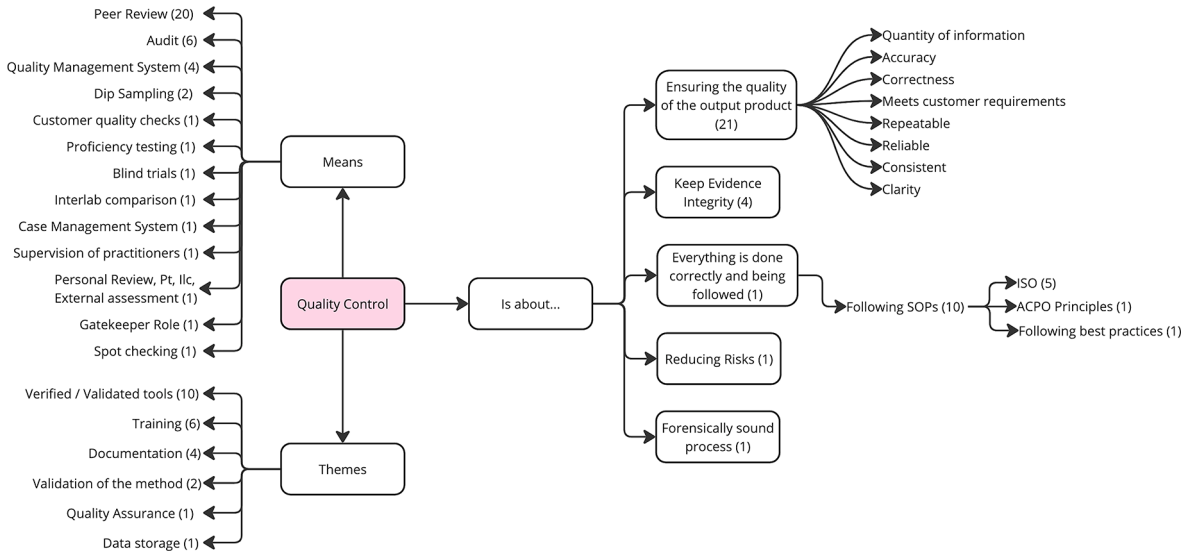

Fig. 3. U.K. Survey: what is Quality Control? The number in parentheses indicates how often a particular notion appeared in the responses, except for the mentioned qualities of the output product, which were not counted.

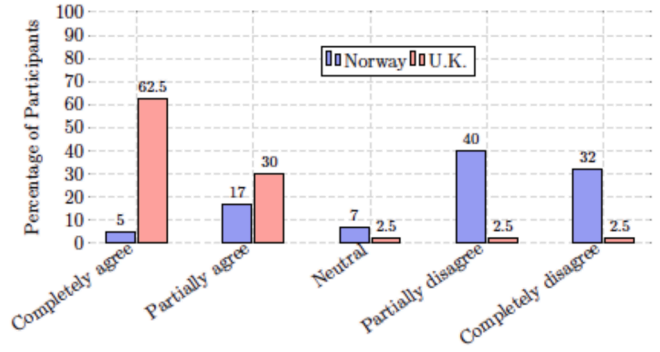

Fig. 4. Systematic control in place (NO-Q9 / UK-Q9) Answers in percentage.

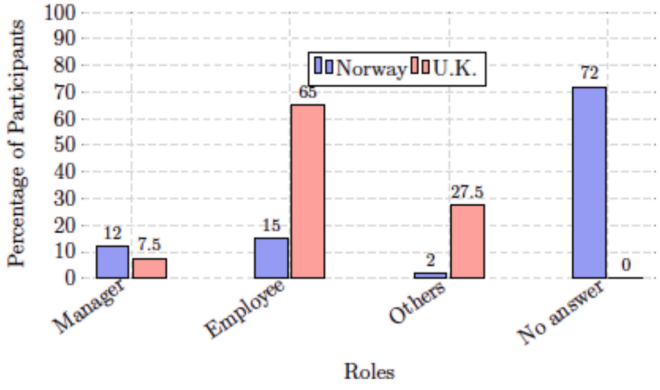

Fig. 5. Roles taken by those implementing quality control (No-Q14 / UK-Q13) per-centage.

the U.K., with only 31 participants completing all the questions. This part of the survey centres on the scope of peer review referring to the seven levels of the Hierarchy. They were asked to indicate the extent to which quality control was implemented at each level, and, if applicable, how many cases were subject to review (All, Most, A few, None). Additionally, participants were asked to report the frequency of quality control checks performed at each level over the past twelve months.

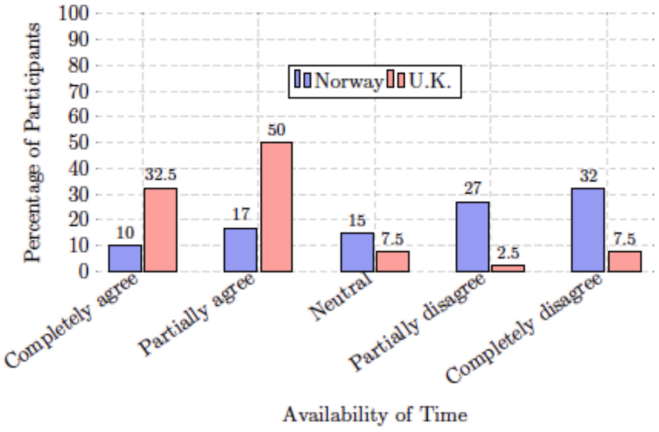

Fig. 6. Availability of Time (NO-Q11 / UK-Q10) Answers in Percentage.

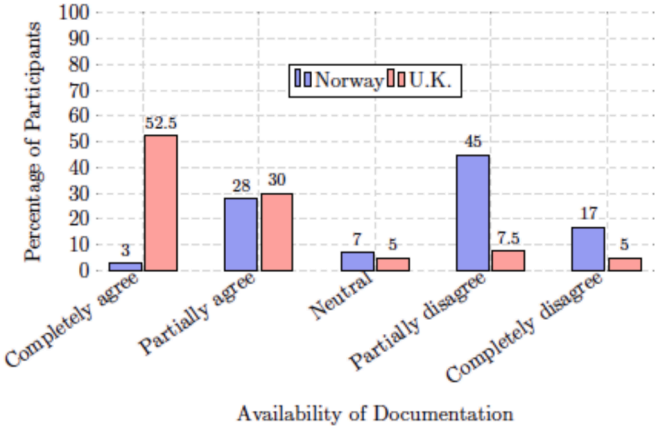

Fig. 7. Availability of Documentation (NO-Q8/ UK-Q8) Answers in Percentage.

4.3.1. *Implementation of quality control*

Fig. 8 summarises the responses regarding the number of cases for which each level of quality control is typically implemented. Half of the U.K. participants reported that all cases undergo quality control at the

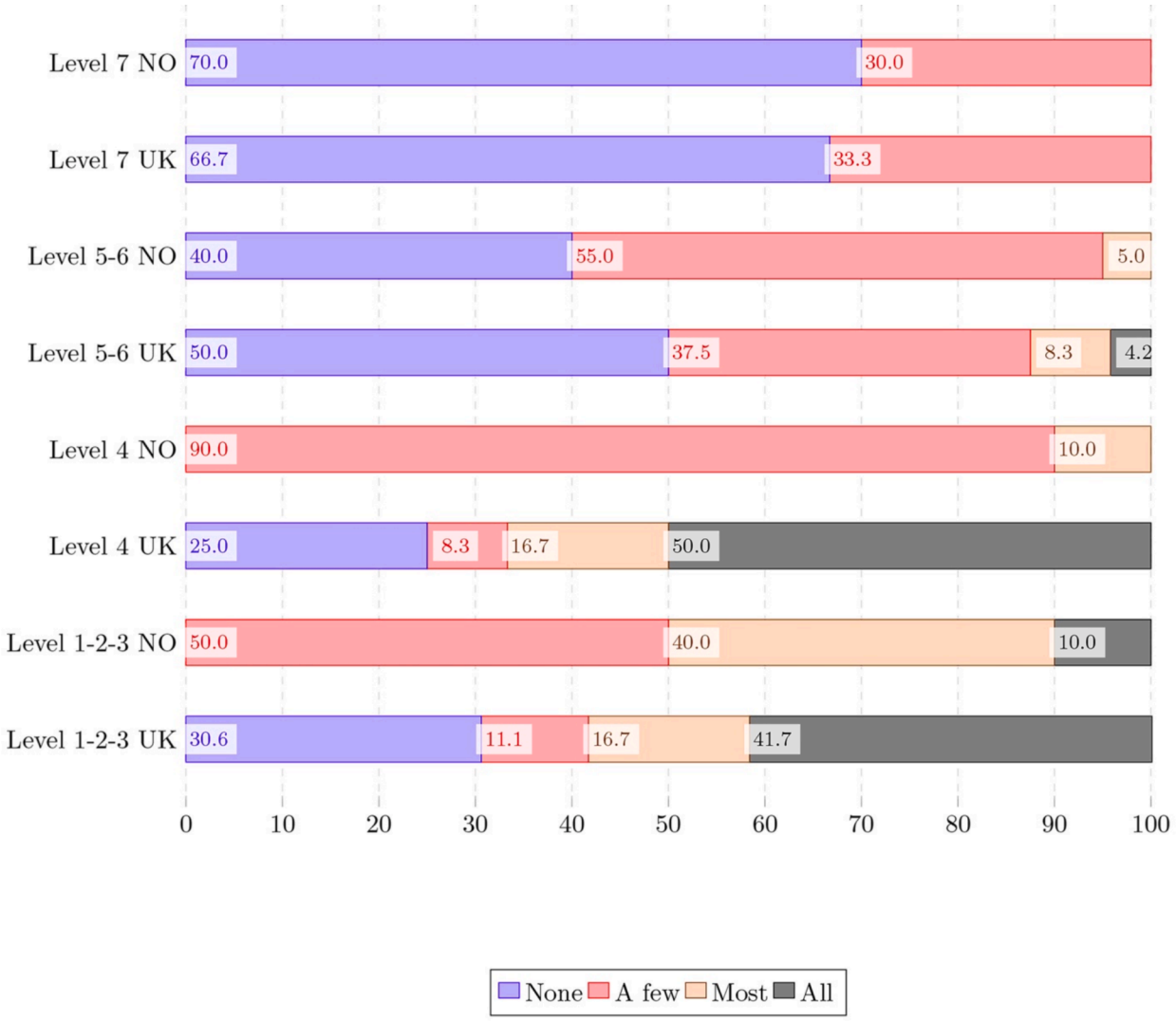

Fig. 8. % Managers (NO (n = 10)/UK (n = 12)): Level of quality control.

first four levels of the Hierarchy. In contrast, none of the Norwegian participants stated the same, with only 10 percent indicating that most cases would be subject to quality control across these initial four levels of the Hierarchy.

For the upper levels of the Hierarchy, both U.K. and Norwegian respondents indicated that verification reviews and re-examinations are less frequently performed. Approximately 10 percent of U.K. respondents and 5 percent of Norwegian respondents stated that most or all cases would undergo full or partial verification reviews. Furthermore, about one third of respondents from both countries stated that only a few cases would be re-examined.

These findings indicate that while the lower levels of the Hierarchy are commonly applied, the more extensive reviews and verifications at the higher levels are less frequently utilised in practice.

4.3.2. *Occurrence of quality control*

A similar issue arises when interpreting responses to the question: *How many times during the last 12 months have you performed quality control of case work at level 1–7?* As shown in Fig. 9, it is notable that approximately one-third (36.8 percent) of U.K. participants reported that no cases were checked at level 4 (Conceptual review). A similar tendency is observed among Norwegian participants, where 36 percent indicated that no cases were reviewed at this level. However, fewer Norwegian respondents (21 percent) compared to their U.K. counterparts (36.8 percent) reported that no cases were checked at level 3 (Sense review).

The remaining results in Fig. 9 are less surprising. Given that most participants indicated that cases rarely or never were reviewed at a level 5, 6, or 7 of the Hierarchy (Fig. 8), it is expected that the vast majority of both U.K. and Norwegian respondents reported that none or only a few (1–5) underwent quality control at these levels in the past twelve months. This observation aligns with the responses in Fig. 8, further highlighting the infrequent application of higher-level quality control checks in practice.

5. Discussion

The study aimed to investigate (1) the extent to which peer review is implemented and utilised in DF investigations in Norway and the U.K. and (2) how peer review practices differ, along with possible explanations for these differences. However, some limitations must be acknowledged. First, there was a significant disparity in response rates between Norway and the U.K., which affects the generalisability of the findings.

In Norway, the survey achieved a high response rate in comparison to the known DF practitioner population. In contrast, the DF practitioners' population in the U.K. is unknown, and the low response rate compared to conservative estimates weakens the generalisability of the

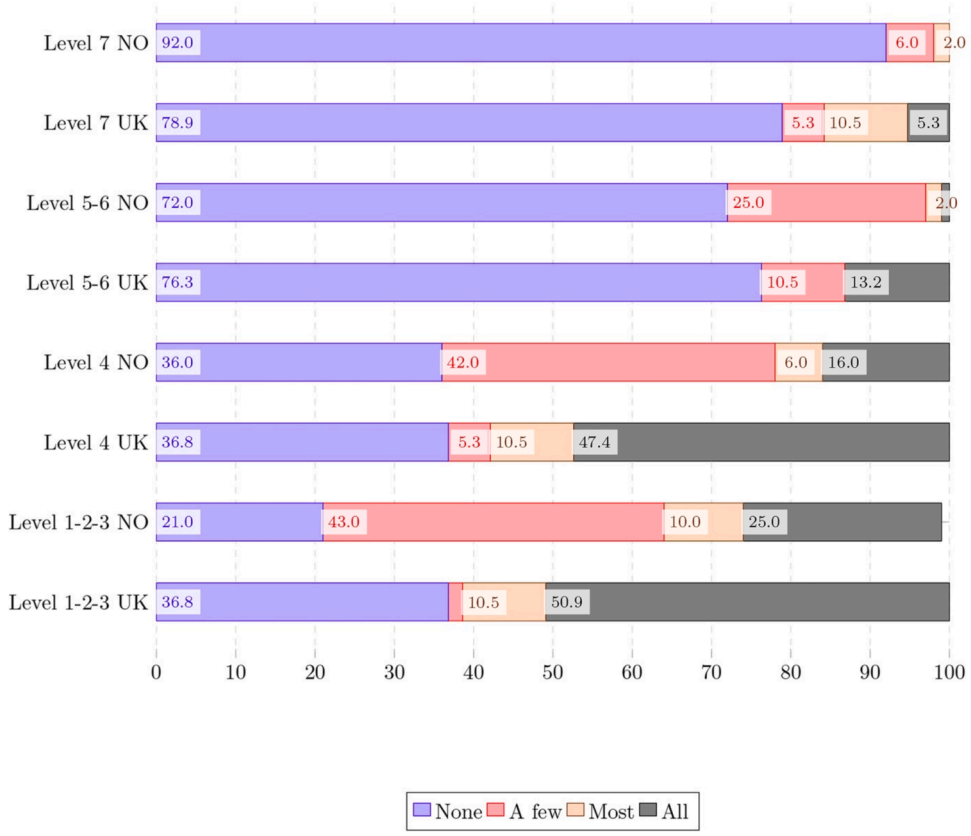

Fig. 9. % Practitioners (NO (n = 50)/UK (n = 19)): Nb of quality control performed in the last 12 months.

U.K. results. Second, data adjustments were required for effective analysis. Norwegian employees were classified as practitioners, aligning with the combined roles of practitioners and technicians in the U.K. survey. In the U.K. sample, respondents in the 'Other' category were classified based on their descriptions; for example, those who identified as both 'Supervisor' and 'Practitioner' were categorised as 'Managers'.

5.1. To what extent is peer review implemented and utilised in DF investigations conducted in both Norway and the U.K.?

Limitations notwithstanding, this study provides valuable insights into whether, how, and to what extent peer review is performed in two distinct systemic contexts. The results indicate a consensus among the respondents from both countries that peer review is important and that it should be implemented in the workflow. While the U.K. appears to have a more formalised implementation with sufficient documentation guiding quality control procedures, the Norwegian survey results indicate a stark contrast, revealing little to no procedural documentation in place. While the U.K. is moving towards standardisation and formalised quality management systems for forensic services providing DF examinations, DF work is done in-house in the Norwegian police with little standardisation or formalised quality management.

5.2. How do practices for peer review diverge, and what are possible explanations for this?

5.2.1. Role distribution: Who is peer-reviewing?

In both countries, quality control responsibilities appear to be shared between managers and employees, with a particular emphasis on employees. In Norway, peer review is largely informal, sporadically initiated and conducted by DF practitioners without a formal structure. In contrast, the U.K. has more formalised quality control systems, where peer review is integrated into the hierarchical structure. However, responses to the survey reveal diverse approaches to peer review in the U.K., suggesting that the task is often delegated based on expertise rather than position. This indicates a shift towards a more dynamic, skill-based approach, where specialised knowledge is prioritised over formal rank when assigning peer review responsibilities.

In both U.K. and Norway, peer review is generally performed in-house. However, some participants raised the possibility of involving external entities in quality control, prompting discussions on whether external scrutiny and oversight could enhance the peer review process.

The participants' responses regarding the implementation of quality control at various levels showed notable variations, raising important questions about how peer review should be structured and who should be responsible for it.

5.2.2. Levels of peer review: How is peer-reviewing implemented?

The survey responses regarding levels of the Hierarchy revealed that most peer review activities in both Norway and the U.K. are concentrated at levels 1–4, focusing primarily on the reported results. Fewer activities involve verification of results or re-examinations of evidence. There are some notable differences between the countries, as the data in Figs. 8 and 9 indicate that a higher proportion of cases would undergo review at level 4 (Conceptual review) in the U.K. compared to Norway. However, in both countries, quality control at higher levels is less frequent. A consistent trend shows that managers tend to hold a more favourable view of the scope and frequency of the peer review conducted in their units than practitioners.

The choice of peer review level significantly affects its effectiveness in detecting critical errors. A level 4 – Conceptual review – might uncover logical inconsistencies in the reported findings and conclusions, but technical errors, such as data misrepresentation due to software 'bugs' or misinterpretations stemming from inadequate technical expertise, may go unnoticed if only the report, and not the findings themselves, is assessed.

The focus on reviewing reports rather than verifying findings likely stems from resource constraints. Verification reviews require more time and expertise than conceptual reviews. In the U.K., this tendency may also be influenced by the marketisation and privatisation of forensic services, where service providers compete to deliver low-cost services that meet the minimum standards set by the Forensic Science Regulator [25]. This competitive environment could result in limiting quality control tasks to the bare minimum required.

In Norway, most DF work is conducted in-house by the police, where standardisation and formal quality management are minimal. Despite the lack of marketisation, a trend similar to that in the U.K. is observed. The combination of tight budgets, limited standardisation, and lack of mandatory quality controls has led to regular detectives taking over more DF tasks, especially in the analysis and reporting stages of the DF process.

As a result, DF practitioners primarily focus on data collection and processing, with fewer technical reports produced for review. This may explain the low frequency of peer review reported by Norwegian participants.

It is worth noting that the legal systems and evidential rules in Norway and the U.K. differ significantly. While peer review is not a mandatory requirement for presenting digital evidence in court in either country, it plays a role in assessing whether expert evidence meets the required evidential standards. Norway's legal framework, closely aligned with the civil law tradition, does not impose strict admissibility rules for expert evidence. A lack of peer review would typically not result in the exclusion of evidence, as the system prioritises an evaluation of relevance and reliability during the trial itself.

In contrast, the U.K.'s common law system applies stricter admissibility criteria for expert evidence, including Daubert-like principles, which require evidence to be grounded in sound scientific methodology and demonstrably relevant to the case in order to be admissible. Given these stricter criteria, it is reasonable to expect that peer reviews are more frequently utilised in the U.K. and that they encompass higher-level reviews (levels 5 and 6) to ensure the evidence withstands scrutiny and avoids exclusion during the admissibility stage. However, in practice, both countries appear to conduct few peer reviews at these higher levels.

Future studies should investigate the types of peer review that courts deem sufficient to establish the reliability of digital evidence and explore how these requirements impact the acceptance and use of such evidence across different legal systems.

6. Conclusion

The study offers novel and valuable insights into the extent and manner peer review is conducted among DF practitioners in Norway and the U.K. It reveals that while peer review is recognised as crucial for maintaining the integrity of DF casework and the resulting digital evidence, its implementation is largely limited to the review of reports. This form of review has limitations in detecting errors, particularly technical issues or misinterpretations of evidence. More comprehensive evaluations, such as the verification of the findings or re-examinations of evidence, remain infrequent in both countries. While limited scope of peer review may be driven by resource constraints, it also reflects systemic factors, including the marketisation and privatisation of forensic services in the U.K., and the decentralised and informal nature of quality control in Norway. In both contexts, the legal and operational frameworks influence the extent to which peer review is prioritised.

The study underscores the need for critical assessment of current quality control practices to ensure they are not merely symbolic, but genuinely effective in detecting critical errors and ensuring accurate presentations of findings. Without rigorous evaluations, there is a risk of fostering a false sense of security about the reliability of digital evidence. Further research should examine peer review mechanisms across the different levels of the Hierarchy and seek to identify the factors that influence whether these controls are effectively implemented.

Ethics statement

The survey was based on informed consent to participate and publish the survey.

In Norway, approval for handling personal data was obtained from the Norwegian Agency for Shared Services in Education and Research (Sikt). In the U.K., data collection and storage adhered to the guidelines of the U.K. Data Protection Act 2018 following an internal ethical review. We submitted the project to the internal ethics committee, CURES, and it was deemed to adhere to Cranfield's ethical policies.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Rune Kenneth Bauge: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Elénore Ryser:** Methodology, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Nina Sunde:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Graeme Horsman:** Methodology, Writing – original draft, Writing – review & editing.

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.

Acknowledgements

Nothing to declare.

Appendix

(For questions in Norwegian, see [14], Appendix C, p. 121–152).

Q1: What type of digital forensic organisation do you work for?

Q2: What is your role?

Q3: What is your professional background?

Q4: What is your highest academic degree?

Q5: How many years of experience do you have within the digital forensics discipline? (if a year is not complete, round down).

Q6: Define what quality control in digital forensics means to you and the most relevant measures to implement it.

Q7: Do you think the quality control of case work in digital forensics is needed? (You can elaborate in the free field if needed).

Q8: In my unit, there is sufficient documentation available (such as procedures, routines, templates etc) that describes and supports appropriate measures for the maintenance of quality control of the digital forensic work performed at the unit.

Q9: In my unit, the examination reports undergo systematic quality control. This question concerns reports resulting from examination/analysis of digital data.

Q10: At my unit, we have enough time to perform quality control of the casework undertaken.

Q11: Quality control should be integrated into the work processes at digital forensics units.

Q12: Quality control of results prior to them being shared with the wider investigative team or officer in charge of a case, should be a natural part of the work flow in a digital forensic unit.

Q13: What is the role of those performing quality control of the case work performed at your unit?

Q14: Is the role of those performing quality control of the case work undertaken at your unit a dedicated 'role' assigned to certain individuals?

Q15: During the last 12 months, have you initiated processes for the quality control of your own case work?

Q16: At our unit, quality control of case work is performed at level 1, Administrative Check.

Q17: How many times during the last 12 months have you performed quality control of case work at level 1 (Administrative Check).

Q18: At our unit, quality control of reports is performed at level 2, Proof Check.

Q19: How many times during the last 12 months have you performed quality control of case work at level 2 (Proof Check)? Q19:At our unit, quality control of reports is performed at level 3, Sense Review.

Q20: How many times during the last 12 months have you performed quality control of case work at level 3 (Sense Review).

Q21: At our unit, quality control of reports is performed at level 4, Conceptual Review.

Q22: How many times during the last 12 months have you performed quality control of case work at level 4 (Conceptual Review).

Q23: At our unit, quality control of reports is performed at level 5, Sampled Verification Review.

Q24: How many times during the last 12 months have you performed quality control of case work at level 5 (Sampled Verification Review).

Q25: At our unit, quality control of reports is performed at level 6, Full Verification Review.

Q27: How many times during the last 12 months have you performed quality control of case work at level 6 (Full Verification Review).

Q28: At our unit, quality control of reports is performed at level 7, Re-examination.

Q29: How many times during the last 12 months have you performed quality control of case work at level 7 (Re-examination).

Q30: Do you have any remarks?

Data availability Statement

Data supporting this study are openly available from CORD at DOI: 10.57996/cran.ceres-2665. This dataset includes only quantitative data; due to ethical restrictions, free-text answers are not available.

References

[1] NPCC, Digital forensic science strategy. National Police Chief's Council (NPCC). https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/publicationslog/2020/national-digital-forensic-science-strategy.pdf, 2020 (accessed 26 September 2024).

[2] C.M. Miller, A survey of prosecutors and investigators using digital evidence: A starting point, Forensic Sci. Int.: Synergy 6 (2023) 100296, https://doi.org/10.1016/j.fsisyn.2022.100296.

[3] E. Casey, Error, uncertainty and loss in digital evidence, International Journal of Digital Evidence 1 (2002) 1–45.

[4] C. Chessman, A" source" of error: Computer code, criminal defendants, and the constitution, Calif. Law Rev. 105 (2017) 179–228, https://doi.org/10.2139/ssrn.2707101.

[5] G. Horsman, "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 (2018) 433–440, https://doi.org/10.1016/j.scijus.2018.04.001.

[6] J. Collie, Digital forensic evidence-flaws in the criminal justice system, Forensic Sci. Int. 289 (2018) 154–155, https://doi.org/10.1016/j.forsciint.2018.05.014.

[7] N. Sunde, I.E. Dror, Cognitive and human factors in digital forensics: Problems, challenges, and the way forward, Digit. Investig. 29 (2019) 101–108, https://doi.org/10.1016/j.diin.2019.03.011.

[8] E.A. Vincze, Challenges in digital forensics, Police Pract. Res. 17 (2016) 183–194, https://doi.org/10.1080/15614263.2015.1128163.

[9] H. Page, G. Horsman, A. Sarna, J. Foster, A review of quality procedures in the U.K. forensic sciences: What can the field of digital forensics learn? Sci. Justice 59 (2019) 83–92, https://doi.org/10.1016/j.scijus.2018.09.006.

[10] K.N. Ballantyne, G. Edmond, B. Found, Peer review in forensic science, Forensic Sci. Int. 277 (2017) 66–76, https://doi.org/10.1016/j.forsciint.2017.05.020.

[11] D.L. Watson, A. Jones, A blueprint for implementing best practice procedures in a digital forensic laboratory, 2nd ed., Academic Press, San Diego, CA, 2023.

[12] G. Horsman, N. Sunde, Part 1: The need for peer review in digital forensics, Forensic Sci. Int.: Digital Invest. 35 (2020) 301062, https://doi.org/10.1016/j.fsidi.2020.301062.

[13] N. Sunde, G. Horsman, Part 2: The phase-oriented advice and review structure (pars) for digital forensic investigations, Forensic Sci. Int.: Digital Invest. 36 (2021) 301074, https://doi.org/10.1016/j.fsidi.2020.301074.

[14] Bauge, R.K., Risks and Quality Control in Norwegian Police's Digital Forensics Process: The Digital Forensics Examiner's Tightrope Walk. Norwegian University of Science and Technology, 2023, Master's thesis. URL: https://hdl.handle.net/11250/3079082.

[15] Police, U.K., Find a police force. URL: https://www.police.uk/pu/find-a-police-force, 2024 (accessed 26 September 2024).

[16] GOV.UK, Police Workforce England and Wales - 30 September 2022. URL: https://www.gov.uk/government/statistics/police-workforce-england-and-wales-30-september-2022/police-workforce-england-and-wales-30-september-2022, 2022 (accessed 26 September 2024).

[17] Forensic Capability Network, "We have to look after them": FCN begins policing's first national forensic workforce strategy. URL: https://www.fcn.police.uk/news/2021-10/we-have-look-after-them-fcn-begins-policings-first-national-forensic-workforce, 2021 (accessed 26 September 2024).

[18] MET, M.P., 2024. Information on digital forensic investigators and work they do. Freedom of information request reference no:01.FOI.23.029031. URL: https://www.met.police.uk/cy-GB/foi-ai/metropolitan-police/disclosure-2023/may-2023/information-digital-forensic-investigators-work-they-do/, 2024 (accessed 26 September 2024).

[19] HMICFRS, Police forces overwhelmed and ineffective when it comes to digital forensics, new report finds. His Majesty's Inspectorate of Constabulary and Fire & Rescue Services. URL: https://hmicfrs.justiceinspectorates.gov.uk/news/news-feed/police-forces-overwhelmed-and-ineffective-when-it-comes-to-digital-forensics/, 2022 (accessed 26 September 2024).

[20] G. Tully, N. Cohen, D. Compton, G. Davies, R. Isbell, T. Watson, Quality standards for digital forensics: Learning from experience in England & Wales, Forensic Sci. Int.: Digital Invest. 32 (2020) 200905, https://doi.org/10.1016/j.fsidi.2020.200905.

[21] D. Wilson-Kovacs, D. Wyatt, The long journey of resistance toward acceptance: Understanding digital forensic accreditation in England and Wales from a social science perspective, Wiley Interdisciplinary Reviews: Forensic Science. 6 (2023) e1501, https://doi.org/10.1002/wfs2.1501.

[22] Politiet, Norway, Employee statistics for the police districts and national units. URL: https://www.politiet.no/om-politiet/tall-og-fakta/bemanning/, 2024 (accessed 26 September 2024).

[23] J.H. Jahren, Is the quality assurance in digital forensic work in the Norwegian police adequate? Norwegian University of Science and Technology, 2020. Master's thesis. URL: https://hdl.handle.net/11250/2781174.

[24] R. Stoykova, S. Andersen, K. Franke, S. Axelsson, Reliability assessment of digital forensic investigations in the Norwegian police, Forensic Sci. Int.: Digital Invest. 40 (2022) 301351, https://doi.org/10.1016/j.fsidi.2022.301351.

[25] K.M. Richmond, Marketisation and Forensic Science Provision in England and Wales, Emerald Publishing Limited (2022), https://doi.org/10.1108/978-1-83909-123-020221009.