
Peer reviewed version

License (if available):
CC BY

Link to published version (if available):
10.1680/jinam.23.00004

Link to publication record in Explore Bristol Research
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Ice Publishing at https://doi.org/10.1680/jinam.23.00004. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/
Viability of bridge inspectors determining defect ratings using photographic images

David T. Nepomuceno, MEng PhD (Bristol)
Formerly PhD Student
University of Bristol, Bristol, UK
https://orcid.org/0000-0001-5612-1176

Paul J. Vardanega, BE MEngSc (QldUT) PhD (Cantab) GMICE MASCE MIEAust FHEA
Associate Professor in Civil Engineering
Department of Civil Engineering, University of Bristol, Bristol, UK
https://orcid.org/0000-0001-7177-7851
Email: p.j.vardanega@bristol.ac.uk (corresponding author)

Maria Pregnolato, MSc PhD (Newc) FHEA
Associate Professor
Department of Hydraulic Engineering, TU Delft, the Netherlands
Formerly, Department of Civil Engineering, University of Bristol, Bristol, UK
https://orcid.org/0000-0003-0796-9618

Theo Tryfonas, PhD (Athens) FRSA
Professor of Infrastructure Systems and Urban Innovation
Department of Civil Engineering, University of Bristol, Bristol, UK
https://orcid.org/0000-0003-4024-8003

John Bennetts, MA (Cantab) MSc (Sur) EngD (Bristol) CEng MICE
Associate
WSP, Bristol, UK

Number of words in main text: approx. 6,100
Number of tables: 8 (approx. 550 words)
Number of figures: 12
Viability of bridge inspectors determining defect ratings using photographic images

D. T. Nepomuceno, P. J. Vardanega, M. Pregnolato, T. Tryfonas & J. Bennetts

Abstract

The visual inspection of bridges is a major undertaking for asset owners and operators. In the UK, visual inspections require inspectors to visit bridges on-site and often at night and in unfavourable weather conditions. Therefore, it would be beneficial to move some of the visual inspection process off-site. This paper studies whether the defect classification aspects of the inspection process could be conducted remotely using photographs. This study examines the defect ratings assigned by ten survey participants who were tasked with examining photographs from visual inspections of ten UK bridges. The survey results were compared with the results from the general inspections previously carried out for the bridges in question. From this dataset, the differences in the ratings given and the extent to which defects are missed were examined. The results show that a higher number of defects were identified for a given bridge by the remote inspectors. Statistical analysis shows that aggregated defects rated by off-site inspectors tend to be more severe and of a higher priority rating compared to those from the on-site inspectors. The results also indicate that there is closer agreement between on-site and off-site inspectors for defects of a higher severity rating.

Keywords: Bridges; Monitoring; Service Life

Notation List

AO Asset Owner
BICS Bridge Inspector Certification Scheme
GI General Inspection
ID Identification
N number of data points
p p-value from a statistical test
PI Principal Inspection
μ mean
σ standard deviation
σ² variance

1. Introduction

1.1 Bridge Condition

Bridges are crucial national infrastructure assets. As they age, appropriate routine maintenance is necessary to ensure their continued serviceability and safety. In the UK, of the approximately 8,000...
bridges on the English Strategic Road Network, about 60% are approaching 50 years in service (Mahut & Woodward 2005). Maintaining a bridge stock in the most efficient and cost-effective way is a significant challenge for asset owners.

For infrastructure asset-owning organisations, visual inspection continues to be the most prevalent method of assessing the condition of bridges (e.g., Lea and Middleton 2002, Phares et al. 2004, Bennetts et al. 2016, Bennetts 2019, Bennetts et al. 2020). However, studies have shown that data gathered from visual inspections are highly subjective and that quality can be largely dependent on the inspector’s prior experience, as well as their knowledge of the relevant bridge’s structural behaviour (e.g., Moore et al. 2001, Lea and Middleton 2002; Phares et al. 2004, Bennetts et al. 2016, Bennetts et al. 2018a, Bennetts et al. 2020). Visual inspection procedures can be time-consuming and unsafe for inspectors, especially for large structures located in difficult terrain. Bhreasail et al. (2019) presents a review focussing on how remote sensing can assist with geotechnical asset management for highways. Nepomuceno et al. (2022a) have presented a recent review of future technologies that may be considered for visual inspection of bridge assets.

In the context of bridge management, McRobbie et al. (2007) found that it was possible to visually assess a bridge using only captured images. The findings of this image-based assessment of bridge condition were comparable to those of the on-site inspection. Subsequently, this research was expanded into a series of initiatives investigating the automation of highway infrastructure inspection (McRobbie et al. 2007, McRobbie 2009, McRobbie et al. 2015). The work in this paper follows on directly from a pilot study in which a basic comparison of results derived from an on-site and off-site assessment of a small group of defects was conducted (Nepomuceno et al. 2021). The pilot study (Nepomuceno et al. 2021) also helped shape the study protocol presented in this paper. Some other parts of the collected survey data have been recently published in Nepomuceno et al. (2022b) showing how aspects of the visual inspection process may be incorporated into the digital work environment. This paper addresses an important research need by examining the feasibility and effectiveness of conducting defect classification remotely using photographs, thereby investigating an approach that has the potential to enhance the efficiency, accuracy and prioritisation of bridge maintenance strategies.

1.2 Study Aims & Objectives

This paper focusses on examining how feasible it is for an off-site inspector to assess a bridge for defects using two-dimensional photographs alone. The objective of this paper is to present a proof of
concept showing that some aspects of the visual inspection workflow can be conducted off-site without significant loss of data accuracy. Central to this work is a survey wherein ten participants were tasked with examining photographs taken during on-site inspections and asked to identify and grade the defects observed. The results were then compared with the assigned ratings from the original on-site inspections, allowing for an analysis of the differences between the two approaches. This research considers the human component of defect identification and grading, assessing the differences between results obtained from physical and remote inspections. The results offer an indication of how reliable this method of defect assessment can be and what (if any) implications more widespread implementation would have on the visual inspection process, with a view towards greater automation by e.g., introducing pattern recognition/machine learning algorithms at a later stage. Additionally, the findings would illustrate the proficiency with which human inspectors can assess defects remotely. If it transpired that a human is incapable of performing this task, the promised revolution of machines rating defects will be extremely difficult to achieve. Finally, while the study compares both traditional and remote procedures, it is beyond the scope of this research to make a judgement on which is superior.

2. The Visual Inspection of Bridges

The current industry standard for visual bridge inspection in the UK involves a cycle of routine inspections, comprising of General Inspections (GIs) and Principal Inspections (PIs) (see HE 2021). Bennett et al. (2018a), Bennett et al. (2018b), Bennett (2019), Bennett et al. (2022) present recent work on the use of visual inspection data to study trends at stock/regional level. The work in this paper primarily focuses on implications for the General Inspection procedure.

2.1 Limitations of Visual Inspections

Wallbank (1989, p. 17) noted that: ‘When recording and comparing the visual condition of a wide variety of bridges it is difficult to be precise and consistent’. It can be argued that compared to other aspects of bridge monitoring research, the reliability of visual inspections has received less attention; however, multiple studies (i.e. Moore et al. 2001, Lea and Middleton 2002, Middleton 2004, Phares et al. 2004 and Bennett et al. 2019) detail progress in this research space in part highlighting the significance of the human variables that influence the reliability of visual inspection.

2.2 Study Motivation

A case can be made to enhance the routine inspection process with off-site defect assessment. Viewed from a human resource standpoint, the division of labour can be optimised. A system which allows for
image capture to be conducted by highly competent photographers, thus ensuring that higher-quality images are produced is a potential alternative. These images can then be provided to qualified inspectors in an office setting to identify and rate defects. This modification to the procedure would involve the use of readily available technology such as high-resolution cameras; making this regime more readily adoptable in industry practice and more easily implemented in day-to-day operations. Demonstrating that structures can be adequately inspected remotely in this manner may also instil confidence to enable future innovation to replace or augment the image capture, defect identification or defect grading operations.

3. Study Protocol

The study presented in this paper comprised an online trial designed to compare the defect ratings received from an on-site GI versus a remote inspection for a given bridge. The trial was designed and led by the first author (see Nepomuceno (2022) for further details on the study methodology). As explained in Nepomuceno et al. (2022b), the third author of this paper acted as an academic representative in the survey to assist with benchmarking and whose results have been aggregated with the other results in the paper. This section briefly outlines the study protocol and format of the trial.

3.1 Trial Overview

The trial comprised three parts which were to be completed in succession:

- A stakeholder questionnaire (Questionnaire A),
- Remote inspection survey (Main Survey),
- An evaluation questionnaire (Questionnaire B).

The trial was designed to be completed in approximately one hour. This sub-section will outline the development of each of these three trial components. Participants of the trial were key stakeholders in the bridge management field. Out of the 19 individuals that were initially approached, full results from ten participants were received and analysed in this paper.

The Main Survey was specifically designed and developed to compare the results of an on-site General Inspection and a remote inspection of a specific bridge structure. In 2018-2019, General Inspections were conducted on ten bridges that are part of the highway network in the Southwest of England. For this paper, the inspectors who carried out the original inspections are referred to as ‘Group A’. The bridge ratings obtained by these inspectors are referred to as ‘Set A’. As is standard procedure, during these inspections, inspectors also took photographs of the bridge and subsequently stored them
on their organisational network. These photographs formed the basis of this trial. The remote inspection survey presents the digital photographs taken by Group A to the participants of the study, who were then tasked with observing the photographs for defects, thus ‘remotely’ inspecting a bridge. For brevity, the participants of the study are collectively referred to as ‘Group B’ in this paper. The subsequent defect ratings from these remote inspections are termed ‘Set B’. These key terms are summarised in Table 1. An overall schematic of the trial is shown in Fig. 1.

3.2 Questionnaires

Questionnaire A was developed to obtain information on the profile of participants. This included key questions on their perceived stakeholder role in the bridge management process and their level of bridge inspection experience. Questionnaire B aimed to get the participant’s perception of the Main Survey and additional evaluation of the process. In addition, both questionnaires allowed the participant to give comments for any of the questions. These comments are included in parts of the discussion in Section 5.

3.3 Material Presented to Participants

To complete the trial, each participant was presented with the following:

1. Photographs of the bridge (taken by inspectors in Group A), each labelled with a unique identification number
2. For structures of Asset Owner 1 (see Section 3.5), forms containing background structural information. The relevant forms for structures under Asset Owner 2 were not available for this trial.
3. A spreadsheet to input their observations.
4. A step-by-step guide on how to complete the survey, as well as reference to defect types and severity degrees.
5. Two questionnaires given as Google Forms.

Participants were sent a OneDrive link containing the materials above, which they could access at any point in time. Using the provided spreadsheet, participants were required to record the following defect attributes:

- Photo ID: filename of photo
- Component: the bridge element on which the defect was located
- Type: defect type
Severity: the resulting amount of damage or loss of functionality to a component. This is measured on a scale of 1 to 5 (TSO, 2007).

Extent: a measure of how widespread a defect is on a component. This is measured on a scale of A to E (TSO, 2007).

Priority: estimated priority of defect repair

In another photo?: Indicate whether the defect is seen in another photo (Yes or No)

Relevant Photo ID: if Yes to above, record photo IDs.

Comments: any additional information.

Each cell was pre-populated for ease of input. If a participant had zero confidence in recording an attribute, an option was available to indicate ‘Insufficient Data Available’.

3.4 Participant Profiles

All ten participants at the time of the survey worked in the UK for various organisations including local authorities, independent consultancies, and universities. Respondents identified themselves as holding various professional roles from Graduate Engineer to Inspector to Asset Owner (see Table 2). When asked which bridge management stakeholder roles they could best identify with, half identified themselves as ‘inspectors’ and/or ‘engineers’. Three participants identified as ‘asset owners’ and one participant identified as a ‘researcher’. 70% of participants had at least four years of inspection experience, with 4-6 years being the most indicated experience level (30%). 60% had undertaken an inspection in the 12 months prior to completing the trial. One participant had never undertaken an official general inspection before. Furthermore, only one participant had taken part in the Bridge Inspector Certification Scheme (BICS) (Lantra, 2021), achieving ‘Senior Inspector’ level.

3.5 Structure Profiles

For this study, the on-site defect data was taken from GIs of two groups of structures: (1) highway bridges in the Southwest of England, and (2) bridges in a county in the East of England (see also the pilot study reported in Nepomuceno et al. 2021). In each of these instances two different asset owners are represented, which are referred to as Asset Owner (AO) 1 and AO2 respectively. Each participant was randomly assigned a structure for the trial, with the only condition being that they had not taken part in the GI themselves. The ten structures subsequently included in the study can be found in Table 3. The majority of the structures (80%) are owned by AO1, with the remainder being owned by AO2. The structures range in length from a 10m culvert to an 82m highway overbridge. The structures chosen
for this study were dictated by the GI information made available at the time of the author’s selection. For a future study, it is suggested that structures be selected more deliberately to represent a range of construction types. This selection may involve including a single type of structure or a wider variety of structures.

In the UK, inspectors generally take photographs of a structure and specific defects during an on-site inspection. A subset of these photographs is then used in the final GI report to the asset owner. The quantity of photographs taken from each structure’s most recent GI can also be seen in Table 3. It is worth noting that these inspection photos were taken prior to the inception of this study, and so have not been influenced by the study protocol implemented in this work. Finally, the number in the structure reference is numerically equivalent to the participant that remotely inspected it (e.g., participant ‘P3’ inspected structure ‘S3’)

4. Results

This section presents the data in both Set A and Set B, outlining the main observations found in each dataset. The results were then compared to check for any differences in the ratings given.

4.1 Set A (onsite inspection data)

A total of 196 defects were recorded from the on-site inspections of the ten structures. See Table 4 for definition of the statistical measures used in this paper. Fig. 2a shows the distribution of the severity and extent recorded. It should be noted that score progression from A to E indicates increasing Extent and score progression from 1 to 5 indicates increasing Severity (see e.g., Bennetts (2019) for more detailed information on defect Extent and Severity classification in UK visual inspection process). From this figure, it is observed that 2B is the most common rating accounting for 33% ($N = 65$) of the defects. When considering Severity ratings only, 67% of defects were rated a 2. When analysing Extent, an equivalent numerical rating was assumed where A = 1 and E = 5. Taking an average of the Severity and Extent ratings separately, values of 2.19 and 3.02 are obtained respectively (see Table 4). Fig. 3 shows the frequency of each defect class for both Set A and Set B. The five most common defects in Set A are: (1) corrosion (21%; $N = 39$), (2) crack (19%; $N = 31$), concrete defect (17%, $N = 27$), vegetation/maintenance (16%, $N = 25$), water 10% ($N = 16$).

4.2 Set B (off-site inspection data)

A total of 206 defects were recorded from the remote inspections of the ten structures. Only 186 (out of 206) defects could be confidently classified, where both a severity and extent rating were assigned by
a remote inspector. The Severity and Extent distribution of these 186 defects is shown in Fig. 2b. Similar to Set A, the most common rating recorded was 2B making up 37% ($N = 65$) of the ratings. However, it is notable that there is a higher proportion of defects rated 3 and 4 in Severity. Compared to Set A, a higher average Severity of 2.56 is observed, while a lower average extent of 2.76 is noted. This value seems to indicate that on average Set B inspectors rated defects more severely compared to the Set A inspectors. This result is further examined in Section 4.4.1. Surveying the most frequent defect classes recorded, the same five classes from Set A are also observed, albeit in a different order: (1) vegetation/maintenance (24%; $N = 41$), (2) concrete defect (13%; $N = 22$), (3) water (12%; $N = 20$), (4) crack (10%, $N = 18$) and (5) corrosion (9%; $N = 16$).

4.3 Proportion of Defects Recorded by Both Groups

Table 5 shows the percentage of unique defects recorded by each group. This analysis was conducted by listing the unique defects recorded by both on-site and off-site groups, and calculating the proportion found by each group. Table 5 shows that the on-site inspectors (Group A) found 196/249 (79%) of the total collective unique defects and that the off-site inspectors (Group B) found 206/249 (83%) of the total collective unique defects. It is observed that a slightly higher percentage of unique defects were recorded by the off-site inspectors. This result may be explained by the off-site inspectors erring on the 'safe' side and pointing out a somewhat minor defect 'just in case' it becomes a problem. This outcome is a positive result, as it could be said that by recording a higher volume of defects, the likelihood of discovering a severe defect would increase.

4.4 Attribute Comparisons

This sub-section compares Set A and Set B for the defect attributes of severity, extent, defect class and priority. To clarify, the defects studied in this sub-section relate to the collective unique defects shown in Table 5, to investigate any variation in the data. Therefore, comparisons are made without using data from Set A or Set B as a benchmark for 'true' identifications and ratings. This may be investigated in the future.

4.4.1 Severity

Considering the Severity ratings from both datasets only, it is observed that a rating of 2 was the most frequent for both Set A and Set B (see Fig. 4b). As previously noted, higher frequencies of defects rated 3 and above are seen in Set B, which suggests that Group B inspectors tend to rate defects more severely. To quantify this rate, the Mann-Whitney U-test can be used to compare each dataset.
statistically. This analysis is a rank-based test for comparing the values of two independent groups which does not require normality (Mann and Whitney 1947, The Concise Encyclopedia of Statistics 2008). It is appropriate for both continuous and ordinal data. A significant result indicates that the values of the two groups are distinct. A recent example of the test being used in a civil engineering context can be found in Huang et al. (2021).

A p-value can be calculated to measure this statistical difference which can then be compared to a significance level of 0.05. The p-values in this study were computed using the SciPy package in Python (SciPy, 2021). For the severity ratings of Set A and Set B, this yielded a p-value of 5.07E-8. Given that the p-value is significantly less than the 0.05 level of significance, it is highly likely that the higher values acquired by off-inspectors are statistically significant. This result supports the suggestion that off-site inspectors will often assign a higher severity to defects compared to their on-site counterparts. This is further examined in Section 4.5.3. The authors note that where the Mann-Whitney U-test is employed in this paper, the Unequal Variances t-test was also used.

4.4.2 Extent

Fig. 4b shows that, for both sets, A and B were the least and most frequent extent ratings respectively. In Set B, the count decreases with higher Extent ratings. There were, however, a much larger number of E extent ratings are recorded in Set A compared to Set B. This results in a p-value of 0.06. This indicates that the Extent ratings from both sets are statistically less different when comparing Severity.

4.4.3 Defect Class

A plot of the frequency distribution by defect class is shown in Fig. 5. Both inspector groups had the same five most frequent defect classes. Group B inspectors recorded much more vegetation/maintenance defects compared to Group A inspectors. Similar disproportion is seen in the defect classes ‘Other’ and ‘Paint/Element Surface’. A more in-depth analysis of the data shows that these were generally more superficial defects. Group A inspectors logged higher proportions for the defect classes ‘Concrete Defect’, ‘Corrosion’ and ‘Crack’. This result suggests that Group A inspectors were more confident in classifying these defects. The variability of defect classes between the two groups point to ambiguities in the current list of defects, which could cause inspectors to be unsure which defect to select.
4.4.4 Priority

When observing the Priority ratings given, Fig. 6 indicates that defects in Set B tend to be of higher priority when compared to Set A. This trend reinforces the suggestion that the remote inspectors tend to have a more pessimistic assessment for a given defect than the original on-site inspectors. This tendency may be due to wider situational awareness and a more holistic appreciation of the bridge's issues for the inspectors that are on the ground.

4.5 Defect Indications

This sub-section examines the agreeability between the ratings from both datasets. Many of the defects recorded in Set A consisted of one defect rating description and rating, mapped to various components in the structure. See Fig. 7 and Table 6 as an example. Corrosion of a parapet mesh is recorded with a rating of 1B and is then mapped to six components. This results in six recorded defects, contributing to the overall total of 196 defects. For this comparative analysis, a distinction was made between these recorded defects, and the actual unique ratings given (i.e., solely taking the unique rating as one defect).

From the 196 defects recorded in Set A, this resulted in 100 unique defect ratings. Analysis was conducted on what proportion of these 100 defects in Set A were 'indicated' in Set B. A defect was judged to be an indication if the remote inspector was clearly referring to a defect in Set A. This result was determined using a combination of the comments and Photo-IDs provided by the off-site inspector.

Out of the 100 unique defects in Set A, 67 were indicated by Group B inspectors. The severity and extent ratings of these 67 defects were examined further. Fig. 8a shows the frequency count of Set A defects by severity rating (denoted by grey bar) and what proportion of these defects were indicated by Group B inspectors (denoted by dashed line). From the plot, it is observed that the higher the severity of a defect in Set A, then the number of defects indicated by a Group B inspector also increases. This result suggests that although remote inspectors miss one third of the defects reported by on-site inspectors, of these they were very unlikely to miss high severity defects, but highly likely to miss low severity defects. A similar trend can also be seen for the Extent ratings (Fig. 8b): the higher the Extent rating, the higher the proportion of defects are indicated by an off-site inspector.

4.5.1 Defect Agreement

The percentage agreement for different attributes of the 67 independently recorded defects between Group A and Group B is shown in Table 7. It is observed that 63% of the defects had matched for ‘Class’. This is followed by ‘Severity’ with a 49% match, and then ‘Extent’ with 33% agreement. This
considerable variation in the way defects were recorded by independent inspectors is notable, highlighting the subjectivity of the visual inspection procedure.

These results are similar to the findings in Bennetts et al. (2018a), which compared two independent inspectors’ ratings of defects on 200 bridges on the Highways England network. Each bridge structure was given two inspectors: one from the pertinent service provider, who completed the regularly scheduled PI, and one from WSP Ltd, who detected and rated defects independently but without a thorough inspection (Bennetts et al. 2018a, and the discussion of this study in Nepomuceno et al. 2021). The results showed that individual inspectors documented defects in a highly variable manner (Bennetts et al. 2018a, Nepomuceno et al. 2021). Nearly 30% of the defects reported by WSP inspectors could not be matched to a comparable defect in the service provider’s PI reports (Bennetts et al. 2018a, Nepomuceno et al. 2021). The present study findings revealed that remote inspectors missed 30% of defects; this value is no worse than sending a second independent on-site inspector to conduct the inspection. This result provides some evidence that remote inspection does not degrade inspection quality, although a larger dataset would be required to give further confidence in this observation.

4.5.2 Statistical Difference of Comparable Defects

The results of the Mann-Whitney U-test for these comparable defects for a significance level of 0.05 is shown in Table 8. For Severity, a p-value of 0.001 denotes significant statistical difference between on-site and off-site inspectors. Considering Extent, a larger significant value of 0.390 is observed suggesting better agreement among ratings. This aspect is further explored in the following sub-section.

4.5.3 Rating Difference of Comparable Defects

Fig. 9 shows the distribution of rating difference for these comparable defects. This difference was calculated by subtracting the pertinent Set A value from the corresponding Set B value. Thus, a positive value indicates that a given defect received a more severe and extensive assessment in Set B. A normal distribution can be fitted to the data, with the highest proportion of defects having zero rating difference. Again, the data indicates that more severe ratings are given in B, but a more symmetrical distribution is observed when it comes to Extent. There are slightly more negative values for Extent, meaning that off-site inspectors tend to rate less extensively.
4.6 Structure

This sub-section examines the defects recorded by structure. Fig. 10 shows the defect counts by structure for Sets A and B. Structure S9 had the highest number of recorded defects while S5 had the lowest. The off-site inspectors recorded more defects on structures S1, S3, S4, S5 and S6. It was also observed that S2, S7, S8 and S10 had notably higher defects in Set A. Fig. 11 shows the percentage of defects in Set A indicated by the corresponding Group B inspector. From the plot the low percentages for S7 (38%) and S8 (35%) are notable. Both these structures were in the remit of AO2 and it was observed by the first author that the photos for these structures were of lower quality compared to the photos of the structures in AO1.

4.7 Confidence Levels

Finally, participants were asked to give an indication of their confidence level when rating the ‘Type’, ‘Severity’ and ‘Extent’ of each defect. A choice of ‘Low’, ‘Medium’, and ‘High’ was given. The results are shown in Fig. 12, the results suggest that off-site inspectors were least confident in their ability to rate ‘Extent’. This sentiment is reflected in several of the comments included in Questionnaires A and B:

- “The extent was the hardest due to what seemed like a lack of photos showing the true extent of some of the defects” (P7).
- “…extent [was] practically impossible [to rate] without knowing which other photos related to the same component” (P8).

This outcome is perhaps unsurprising, as one of the key challenges affecting an off-site inspector is a loss of sense of scale.

5. Discussion

The results from this study offer useful information on the feasibility for human inspectors to rate bridge defects using photographic images. Additionally, it adds to the body of data regarding visual inspection subjectivity. The following sub-sections (i.e. sub-sections 5.1 to 5.4) provide some key discussion points that emerged from examination of the results of the study. Sub-section 5.5 gives some comments on how the study participants thought the off-site inspection protocol could be improved. Sub-section 5.6 gives suggestions for future research directions as a result of this study.

5.1 Off-site Inspectors Recorded More Defects Compared to their On-site Counterparts

When considering the number of unique defects listed by inspectors in Group A and Group B, overall, a slightly higher percentage of unique defects were recorded by the off-site inspectors (see Table 5).
This issue may be explained by the off-site inspectors erring on the ‘safe’ side and pointing out a somewhat mild defect ‘just in case’ it is a problem. It could be argued that increasing the number of defects recorded increases the likelihood of discovering a serious defect.

5.2 Off-site Inspectors are Effective at Recording Onerous Defects

The plots in Fig. 8a and Fig. 8b indicate that while off-site inspectors are effective at recording more severe defects, they are less successful at identifying less onerous defects. Out of the 73 defects in Set A rated a 2 or less, 63% were indicated by Set B inspections, whereas the 78% of the Set A defects rated 3 and above were indicated by Set B. This result may be viewed as a promising outcome, as one could argue that it is the higher severity defects that are critical to detect and the most vital to consider when implementing maintenance actions, particularly for General Inspections.

5.3 Off-site Inspectors Tend to Rate Individual Defects More Severely

Of the 67 comparable defects in both datasets, it can be shown that individual defects are typically assigned a higher severity rating by off-site inspectors. This tendency is further supported by the plot shown in Fig. 6, which shows the distribution of Priority ratings assigned; Group B inspectors tended to rate defects of a higher priority. As mentioned, this may be due to a form of task bias within off-site inspectors; because they have the aim of identifying defects (especially more onerous instances), they may naturally be inclined toward assigning more severe ratings to be on the safe side. It is acknowledged that an increased number of defects may make it more difficult for an off-site inspector to appreciate the wider picture and be too focused on individual defects. However, this process would rely on inspectors being appropriately qualified (e.g., BICS (Lantra, 2021)); a qualified inspector who has working knowledge of structural articulation would be expected to understand the criticality of individual aspects of various components with regards to the overall structural integrity of a bridge.

5.4 Loss of Sense of Scale and Orientation is Significant Barrier to Off-Site Inspectors

When participants were surveyed about their degree of confidence, Defect Type was rated as the most confident, whilst Extent was rated as the least confident. There was also particular difficulty when assessing where a defect was on a structure, making location and orientation very difficult to determine. Several participant comments alluded to this:

- “…hard to visualise where some of [the defects] were and how they relate to other defects” (P1).
- “not all photos showed [the] location of defects making it difficult to identify” (P2).
• “The orientation of the pictures was very difficult to identify” (P4).
• “It is difficult to assess extent when the majority of photos are close ups with no sense of scale” (P6).
• “…found it confusing [when trying to work out] which element/direction I was looking” (P9).
• “extent [is] impossible [to rate] when [the] whole element not in photo” (P10).

These findings, along with the comments, point to a major weakness in the method of bridge assessment as explored in this work.

5.5 Evaluating this Remote Inspection Process

5.5.1 Potential Solutions

A number of participant’s comments provided valuable insight on how remote inspection might be improved. Many of these emphasise the need for information relating to defect location: “Anything to allow the person grading the defects to easily identify where on the structure it is would [be] very helpful” (P7).

One bridge inspector noted that “if a sketch was included with some basic comments or notes where defects [were], this would greatly increase confidence [when rating defects]” (P2). Some suggested that supplementing photographs with details such as direction and scale would be of value:

• ”including both close up and wide-angle photos for each defect to aid in locating area of structure. Including a scale in photos” (P6).
• “so long as photo location/direction are recorded, and some scaling tool is available” (P9).

One participant stated that a “walk-through video” (P1) would be helpful. As also noted in Nepomuceno et al. (2022b), another participant mentioned that “a 5-minute chat with the person that took the photos” (P8) would facilitate knowledge transfer pertaining to the bridge under investigation. The use of 360-degree camera imagery as an adjunct to the traditional 2D photos could offer benefits in providing a more comprehensive view of the bridge structure. This approach could assist inspectors in gaining a better understanding of the defect's location and context within the overall structure. By including both close-up and wide-angle photos, along with a scale for reference, inspectors can more easily identify the specific area affected and assess the extent of the defect.

5.5.2 Foreseen Barriers to the Procedure

Valuable comments were provided on the actual procedure itself. Two participants mentioned the time taken to undertake the survey, stating that it “took quite a time to do” (P1) and that it “took more than 1
hour” (P4). One participant gave insight into how the number of photos provided to a remote inspector would have to be carefully considered: “Once structures get above a certain size, ... the number of photos to be reviewed would get very daunting, and the inspector could easily get lost amongst the data. I have experience of uploading others’ inspection notes that this can easily happen” (P8) (this quote is also included in Nepomuceno et al. 2022b). Finally, the quality of photos was stated plainly by one participant, saying “inspectors are not professional photographers, and it shows: defects out of focus, dark and only some on screen” (P10).

5.6 Looking Ahead

The results from this study show promise in utilising remote defect assessment methods to supplement the routine inspection regime, rather than completely replacing it. The findings from this work indicate that ‘off-site’ inspectors are unlikely to miss higher severity defects. By conducting such a process perhaps on an annual basis, the level of granularity needed to track rapidly deteriorating defects on aged structures would increase. This approach would help take a step toward being able to accurately quantify the rate of change of deterioration, which the current assessment system does not sufficiently enable (Bennetts et al. 2021). In turn, this may increase the quality of data taken by visual inspections.

Such a method would also align well with the changing role of a bridge inspector. In the past, the role was typically seen as a long-term career, where one would amass decades of experience and be extremely knowledgeable of the structural behaviour of many bridges in a region of the network. There is now a higher turnover in these posts, where bridge inspectors come into the role for a few years and then move on to a different position. By having a procedure where defect photos are professionally taken and assessed remotely, less experienced inspectors may be able to undertake defect identification and grading with more confidence, facilitating a ‘Digital Stewardship’ of the asset. A recently proposed Schema for the remote inspection of bridges has recently been published by the authors to show how these aims can be achieved (Nepomuceno et al. 2022b).

Regarding the experience level of inspectors, in the future, a similar experimental trial could be devised where one bridge is virtually rated by multiple inspectors of various experience levels. By virtually rating the same bridge, we can compare the scoring ratings and comments provided by inspectors with different experience levels, providing deeper insights into the relationship between experience and the quality of defect assessments.
6. Summary & Conclusions

To investigate the effectiveness of off-site inspections for assessing bridge defect ratings, a targeted trial was designed wherein participants rated bridge defects using photographs taken during the on-site General Inspections of a particular structure. The results were then compared to the ratings assigned to the same structure by the on-site inspector. In all, ten structures were inspected as part of the study.

The on-site inspectors recorded 196 defects across the ten structures, whilst the remote inspectors (i.e., participants of the study) recorded 206 defects (of which 186 could be confidently classed).

Statistical analysis shows that aggregated defects rated by off-site inspectors tend to be perceived as more severe and of a higher priority compared to those from the on-site inspectors. The results also indicate that, for defects of a higher Severity and Extent, there is closer agreement in ratings between on-site and off-site inspectors. The results from this study suggest that there may be promise in standardising remote inspections for the identification and grading of more urgent, higher severity defects. The authors note that any move to a remote inspection process should at first be seen as a complementary approach to traditional on-site inspections, rather than replacing them entirely. A potential future is seen where aspects of GIs are made remote, whilst PIs will remain the same, wherein on-site inspectors can gain practical field experience. Finally, it is noted that a larger data-set would be needed to give further confidence in the trends reported in this paper.

Acknowledgements

The first author would like to acknowledge support from WSP as well as from the Engineering and Physical Sciences Research Council (EPSRC) through the National Productivity Investment Fund (grant number EP/R51245X/1). The authors express thanks to the anonymous individuals who gave their time to participate in this study. The statements made in this paper do not necessarily reflect the opinions of the organisations involved. The authors thank the anonymous reviewers of the paper for their helpful comments which helped improve the paper.

Data Availability Statement

To maintain anonymity of the survey participants the survey data cannot be made available in a data repository. No other experimental data were generated during this study.

References


LIST OF TABLE & FIGURE CAPTIONS

Table 1: Definition of Key Terms
Table 2: Participant Profiles
Table 3: Structure Profiles
Table 4: Descriptive Statistics for Severity for Set A and Set B
Table 5: Percentage of Unique Defects Recorded by Each Group
Table 6: Recorded Attributes for Defect Shown in Fig. 9
Table 7: Percentage Agreement between Defect Attributes
Table 8: Results of Mann-Whitney U-Test for Comparable Defects

Fig. 1. A high level schematic of the trial procedure depicting how photographs taken from the on-site inspections are used by Group B to identify and grade defects. The resulting Set A and Set B are compared in this paper.

Fig. 2. (a) Distribution of Severity and Extent ratings for Set A. Note: Score progression from A to E indicates increasing Extent and score progression from 1 to 5 indicates increasing Severity. Increasing marker size indicates increased number of observations corresponding to that point on the graph, (b) Distribution of Severity and Extent ratings for Set B. Note: Score progression from A to E indicates increasing Extent and score progression from 1 to 5 indicates increasing Severity. Increasing marker size indicates increased number of observations corresponding to that point on the graph.

Fig. 3. Five most frequent defect classes recorded for each dataset.

Fig. 4. (a) Frequency of Severity ratings in Set A and B and (b) Frequency of Extent ratings in Set A and B.

Fig. 5. Frequency of Defect classes recorded in Set A and Set B.

Fig. 6. Frequency of Priority ratings recorded in Set A and Set B.

Fig. 7. Example of format of defect record in Set A. One unique defect rating is assigned to six components. Photograph courtesy of WSP, used with permission. (See Table 6 for recorded attributes)

Fig. 8. (a) Plot showing the number and proportion of defects (dashed line) in Set A that are indicated by inspectors in Group B, by Severity. The dark column bars represent defects indicated in both Set A and Set B, (b) Plot showing the number and proportion of defects (dashed line) in Set A that are indicated by inspectors in Group B, by Extent.

Fig. 9. Plot showing the rating difference between directly comparable defects.

Fig. 10. Number of defects recorded per structure.

Fig. 11. Percentage of Set A defects ratings indicated by off-site inspector per structure.

Fig. 12. Confidence levels indicated by Group B when rating Defect Type, Severity and Extent.
<table>
<thead>
<tr>
<th>Terms</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>On-site bridge inspectors conducting GIs</td>
</tr>
<tr>
<td>Group B</td>
<td>Participants of this study who inspected the bridge using photographs</td>
</tr>
<tr>
<td>Set A</td>
<td>Defect ratings obtained from on-site inspections</td>
</tr>
<tr>
<td>Set B</td>
<td>Defect ratings obtained from remote inspection</td>
</tr>
<tr>
<td>Participant Reference</td>
<td>Stakeholder Role(s)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>P1</td>
<td>Asset owner</td>
</tr>
<tr>
<td>P2</td>
<td>Bridge inspector</td>
</tr>
<tr>
<td>P3</td>
<td>Bridge inspector, bridge engineer, asset owner</td>
</tr>
<tr>
<td>P4</td>
<td>Assistant engineer</td>
</tr>
<tr>
<td>P5</td>
<td>Academic</td>
</tr>
<tr>
<td>P6</td>
<td>Bridge engineer</td>
</tr>
<tr>
<td>P7</td>
<td>Bridge inspector</td>
</tr>
<tr>
<td>P8</td>
<td>Bridge inspector, bridge engineer</td>
</tr>
<tr>
<td>P9</td>
<td>Bridge inspector, bridge engineer</td>
</tr>
<tr>
<td>P10</td>
<td>Asset owner</td>
</tr>
<tr>
<td>Structure Reference</td>
<td>Asset Owner</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>S1</td>
<td>AO1</td>
</tr>
<tr>
<td>S2</td>
<td>AO1</td>
</tr>
<tr>
<td>S3</td>
<td>AO1</td>
</tr>
<tr>
<td>S4</td>
<td>AO1</td>
</tr>
<tr>
<td>S5</td>
<td>AO1</td>
</tr>
<tr>
<td>S6</td>
<td>AO1</td>
</tr>
<tr>
<td>S7</td>
<td>AO2</td>
</tr>
<tr>
<td>S8</td>
<td>AO2</td>
</tr>
<tr>
<td>S9</td>
<td>AO1</td>
</tr>
<tr>
<td>S10</td>
<td>AO1</td>
</tr>
</tbody>
</table>
Table 4: Descriptive Statistics for Severity for Set A and Set B

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of data points (N)</td>
<td>196</td>
<td>199</td>
</tr>
<tr>
<td>Mean ($\mu$)</td>
<td>2.19</td>
<td>2.55</td>
</tr>
<tr>
<td>Mode</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Median</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Standard deviation ($\sigma$)</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>Variance ($\sigma^2$)</td>
<td>0.51</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Table 5: Percentage of Unique Defects Recorded by Each Group

<table>
<thead>
<tr>
<th>Structure</th>
<th>Group A</th>
<th>Group B</th>
<th>Defects in A, indicated by Group B</th>
<th>Collective unique defects</th>
<th>AB Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>16</td>
<td>20</td>
<td>13 (81%)</td>
<td>23</td>
<td>57%</td>
</tr>
<tr>
<td>S2</td>
<td>30</td>
<td>22</td>
<td>20 (67%)</td>
<td>32</td>
<td>63%</td>
</tr>
<tr>
<td>S3</td>
<td>16</td>
<td>28</td>
<td>15 (94%)</td>
<td>29</td>
<td>52%</td>
</tr>
<tr>
<td>S4</td>
<td>24</td>
<td>36</td>
<td>20 (83%)</td>
<td>40</td>
<td>50%</td>
</tr>
<tr>
<td>S5</td>
<td>2</td>
<td>7</td>
<td>2 (100%)</td>
<td>7</td>
<td>29%</td>
</tr>
<tr>
<td>S6</td>
<td>11</td>
<td>14</td>
<td>11 (100%)</td>
<td>14</td>
<td>79%</td>
</tr>
<tr>
<td>S7</td>
<td>13</td>
<td>8</td>
<td>5 (38%)</td>
<td>16</td>
<td>31%</td>
</tr>
<tr>
<td>S8</td>
<td>17</td>
<td>13</td>
<td>6 (35%)</td>
<td>20</td>
<td>30%</td>
</tr>
<tr>
<td>S9</td>
<td>41</td>
<td>40</td>
<td>40 (98%)</td>
<td>41</td>
<td>98%</td>
</tr>
<tr>
<td>S10</td>
<td>26</td>
<td>18</td>
<td>17 (65%)</td>
<td>27</td>
<td>63%</td>
</tr>
</tbody>
</table>
Table 6: Recorded Attributes for Defect Shown in Fig. 9

<table>
<thead>
<tr>
<th>Defect Attribute</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defect Type</td>
<td>RS</td>
<td>RCo</td>
</tr>
<tr>
<td>Defect Class</td>
<td>Corrosion</td>
<td>Corrosion</td>
</tr>
<tr>
<td>Severity</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Extent</td>
<td>B</td>
<td>Insufficient Data Available</td>
</tr>
<tr>
<td>Priority</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Component Type</td>
<td>Multiple (see Figure 9)</td>
<td>Road Vehicle Restraint</td>
</tr>
<tr>
<td>Component Name</td>
<td>Multiple (see Figure 9)</td>
<td>Insufficient Data Available</td>
</tr>
<tr>
<td>Comments</td>
<td>Minor corrosion and distortion to parapet mesh infill panels.</td>
<td>Not clear which parapet mesh infill this is showing. Mesh infill has surface corrosion present.</td>
</tr>
</tbody>
</table>
Table 7: Percentage Agreement between Defect Attributes

<table>
<thead>
<tr>
<th>Defect attribute</th>
<th>Percentage agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>63%</td>
</tr>
<tr>
<td>Severity</td>
<td>49%</td>
</tr>
<tr>
<td>Extent</td>
<td>33%</td>
</tr>
<tr>
<td>Severity and extent</td>
<td>21%</td>
</tr>
<tr>
<td>Class, severity, and extent</td>
<td>16%</td>
</tr>
</tbody>
</table>
Table 8: Results of Mann-Whitney U-Test for Comparable Defects

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Set A</th>
<th></th>
<th></th>
<th>Set B</th>
<th></th>
<th></th>
<th></th>
<th>Significant Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>μ</td>
<td>σ</td>
<td>σ²</td>
<td>N</td>
<td>μ</td>
<td>σ</td>
<td>σ²</td>
</tr>
<tr>
<td>Severity</td>
<td>67</td>
<td>2.34</td>
<td>0.72</td>
<td>0.52</td>
<td>67</td>
<td>3.15</td>
<td>1.16</td>
<td>1.35</td>
</tr>
<tr>
<td>Extent</td>
<td>67</td>
<td>2.75</td>
<td>0.85</td>
<td>0.73</td>
<td>65</td>
<td>2.98</td>
<td>1.99</td>
<td>1.02</td>
</tr>
</tbody>
</table>
Group A (On-site)

1. Conduct on-site inspection
2. Defect identification
3. Defect grading
4. Set A

Group B (Off-site)

1. Defect identification
2. Defect grading
3. Set B

Photographs from on-site inspection

Flow diagram showing the process flow between Group A and Group B.
<table>
<thead>
<tr>
<th>Defect Type</th>
<th>Defect Class</th>
<th>Severity</th>
<th>Extent</th>
<th>Priority</th>
<th>Component Type</th>
<th>Component Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS</td>
<td>Corrosion</td>
<td>1</td>
<td>B</td>
<td>Low</td>
<td>Centre Deck</td>
<td>North Parapet</td>
</tr>
<tr>
<td>RS</td>
<td>Corrosion</td>
<td>1</td>
<td>B</td>
<td>Low</td>
<td>Centre Deck</td>
<td>South Parapet</td>
</tr>
<tr>
<td>RS</td>
<td>Corrosion</td>
<td>1</td>
<td>B</td>
<td>Low</td>
<td>West Deck</td>
<td>North Parapet</td>
</tr>
<tr>
<td>RS</td>
<td>Corrosion</td>
<td>1</td>
<td>B</td>
<td>Low</td>
<td>West Deck</td>
<td>South Parapet</td>
</tr>
<tr>
<td>RS</td>
<td>Corrosion</td>
<td>1</td>
<td>B</td>
<td>Low</td>
<td>East Deck</td>
<td>North Parapet</td>
</tr>
<tr>
<td>RS</td>
<td>Corrosion</td>
<td>1</td>
<td>B</td>
<td>Low</td>
<td>East Deck</td>
<td>South Parapet</td>
</tr>
</tbody>
</table>