Non-Open Access article acknowledgement

This article has been accepted for publication in Journal of Clinical Pathology following peer review. The definitive copyedited, typeset version

Quantification of histochemical stains using whole slide imaging: development of a method and demonstration of its usefulness in laboratory quality control.

J Clin Pathol. 2014 Dec 5. pii: jclinpath-2014-202526. doi: 10.1136/jclinpath-2014-202526. [Epub ahead of print]

is available online at: http://jcp.bmj.com/content/early/2014/12/05/jclinpath-2014-202526.abstract

Title

Quantification of histochemical stains using whole slide imaging: Development of a method and demonstration of its utility in laboratory quality control

Authors

Allan Gray BSc 1

Alex Wright BSc(Computing) 2

Pete Jackson FIBMS CSi MPhil 1

Mike Hale BSc 2

Darren Treanor BSc (Computing) PhD FRCPath 1, 2

Affiliations:

1. Leeds Teaching Hospitals NHS Trust, Leeds, UK

2. University of Leeds, Leeds, UK

Corresponding author:

Dr. Darren Treanor

darrentreanor@nhs.net

St James University Hospital, Beckett St., Leeds LS9 7TF, UK

Tel. +44(0)113 20 67546 Fax. +44(0)113 20 67803

Keywords: DIGITAL PATHOLOGY, IMAGE ANALYSIS, QUALITY ASSURANCE, QUALITY CONTROL, HISTOPATHOLOGY

Word count: 5866 title page, abstract, references, figures and tables

Competing Interest:

None declared.

Abstract

Aims: Histochemical staining of tissue is a fundamental technique in tissue diagnosis and research, but it suffers from significant variability. Efforts to address this include laboratory quality controls and quality assurance schemes, but these rely on subjective interpretation of stain quality, are laborious and have low reproducibility. We aimed (i) to develop a method for histochemical stain quantification using whole slide imaging and image analysis and (ii) to demonstrate its utility in measuring staining variation.

Methods: A method to quantify the individual stain components of histochemical stains on virtual slides was developed. It was evaluated for repeatability and reproducibility, then applied to control sections of appendix to quantify H&E staining (H/E intensities and H:E ratio) between automated staining machines, and to measure differences between six regional diagnostic laboratories.

Results: The method was validated with less than 0.5 % variation in H:E ratio measurement when using the same scanner for a batch of slides (i.e. it was repeatable) but was not highly reproducible between scanners or over time, where variation of 7% was found. Application of the method showed H:E ratios between three staining machines varied from 0.69 to 0.93, H:E ratio variation over time was observed. Inter-laboratory comparison demonstrated differences in H:E ratio between regional laboratories from 0.57 to 0.89.

Conclusions: A simple method using whole slide imaging can be used to quantify and compare histochemical staining. This method could be deployed in routine quality assurance and quality control. Work is needed on whole slide imaging devices to improve reproducibility.

Introduction

Histochemical staining of tissue sections is a long established and widely used diagnostic and research tool. The interpretation of haematoxylin and eosin (H&E) stained tissue is central to the diagnosis of cancer and other diseases, and it is relied on by pathologists and scientists worldwide for the initial assessment of tissue in a wide variety of human and animal conditions.

However histochemical staining is more of an art than a science. A large number of variables can affect the final appearance of histochemically stained tissue. These include pre-analytic factors (such as tissue origin, disease status, duration of warm and cold ischaemia, duration and type of fixation) as well as analytic factors (such as the dyes and other reagents used, the technique applied, and variation in technician performance) [1]. Even in relatively controlled research environments where preanalytic factors are optimised (generating large numbers of identically stained serial sections of a single tissue block for 3D reconstruction) we have observed significant colour variability [2,3].

In clinical practice, H&E staining is known to be rather variable in appearance. Pathologists commonly recut and restain cases referred from other institutions because of unfamiliarity or discomfort with the appearances of the staining. A poorly stained section may prolong the time taken to make a diagnosis (by increasing the interpretative effort of the pathologist or leading to delays as slides are recut and restained) or even lead to misdiagnosis. Variability in histochemical staining will also affect the accuracy and reproducibility of image analysis in clinical practice and research.

Steps to manage staining variability

Given the importance of histochemical staining in diagnosis, efforts to make it more consistent have been developed at multiple levels. In individual laboratories, standardised protocols are employed to increase consistency of staining. Automation of staining is now widely employed to attempt to make the staining procedure as uniform as possible, and laboratory dye production includes quality control steps to ensure consistent chemical composition and concentration of dyes. After staining,

visual quality control of individual stained slides (or a sampled subset) is often performed before they are released from the laboratory.

Nationally, laboratory quality control programs have been established in an effort to increase inter-laboratory consistency. In the UK, the National External Quality Assurance Scheme [4] run inter-laboratory circulations of slides to assess the results of their histochemical staining of both internally and externally procured & sectioned tissue. In the US, the Biological Stain Commission [5]regulates the composition of biological stains, and the College of American Pathologists Quality Assurance Program [6]conducts inter-laboratory comparisons of staining. Such programs have heightened awareness of the problem of staining variability and encouraged efforts to reduce it.

Generally, national/ regional quality assurance programs provide a standard stained section (or description of an ideal stain) as an example of best practice, and provide marking systems comparing the submitted slides against what is thought to be an ideal stain. For example, the UK NEQAS system assesses each submitted slide for a variety of factors including fixation, processing, microtomy and staining from 0 (good quality) to 9 (poor quality). It uses these scores to perform inter-laboratory comparisons, identify areas of poor performance and provide feedback to individual laboratories. Schemes may include an educational element and a developmental element to assist laboratories with quality issues.

However, such quality assurance (QA) programmes usually rely on subjective assessment of histochemical stains in order to compare the performance of laboratories. In the UK, expert assessors (usually histology technicians and pathologists) assess slides submitted for QA visually, looking for characteristic staining expected for each histochemical stain and assigning a semi-quantitative score to the slide. For example, Haematoxylin is expected to selectively stain nuclei a dark purple and eosin is expected to stain the cytoplasm of most cells, with selective enhancement in certain cells such as muscle cells and eosinophils, but should not excessively stain stromal connective tissue.

Although such semi-quantitative scoring is well established for the assessment of stain quality and intensity in all areas of pathology and histotechnology, it suffers from several problems. Firstly it is subjective and prone to intraobserver variation, leading to reduced repeatability. Secondly, due to interobserver variation it lacks reproducibility, so comparisons cannot reliably be made between institutions, across time, or between countries. Finally, it requires manual assessment by experts of slides, so is relatively costly and time consuming and cannot be applied en mass or in routine daily practice.

In this paper a novel method for quantification of histochemical stains is described which combines whole slide imaging with image analysis. It is demonstrated that the method is repeatable within certain constraints and its application to the assessment of stain intensity in several settings is described, including comparisons of staining within and between laboratories.

Materials and Methods

Institutional ethical approval was obtained for the research. The following paragraphs describe the four stages of the work.

Experiment 1: Validation of stain quantification method

To determine the repeatability of the analysis, repeated measurements were made of standardised sections of tissue. A 5mm long cross section of human appendix was obtained from surplus surgically removed human appendix. The piece was fixed in 10% formal saline for 24 hours, processed on a routine process (Sakura Tissue – Tek VIP 150 4891), using the process described in

Step	Solution	Immersion (hrs:mins)
1	Formal Saline (10%)	0:15
2	70% Alcohol	0:30
3 & 4	Absolute Alcohol (Genta medical)	0:30
5	Absolute Alcohol	1:00
6 & 7	Absolute Alcohol	1:30
8	Xylene (Genta Medical)	1:00
9 & 10	Xylene	1:30
Wax 1	Wax (VWR wax 36114)	1:00
Wax 2	Wax	1:30
Wax 3	Wax	1:45

Table 1. The tissue was embedded in paraffin wax using moulds and an embedding station (Leica EG 1160 embedding station 038630352). The paraffin block was removed from the mould and trimmed on a rotary microtome until the full face of the tissue was seen, (defined as when all of the tissue on one face can be seen on the block).

To minimise variation in section thickness, the block was placed on ice for 10 minutes allowing the wax to cool and harden, allowing a thin section to be cut more easily. After each 5 sections were taken the block was placed on ice again to cool and harden the paraffin. Serial sections were cut with a microtome (Leica rotary microtome RM2235) with the section thickness set to 3 microns. Sections were floated out on a water bath (Thermo scientific 3120058) at 45 degrees Celsius and placed on a non coated glass slides (Tissue-Tek yellow frosted slides 9607). After drying on a rack for 5 minutes, the unstained sections were hot plated (Thermo scientific hotplate 3120061) for 10 minutes at 70 degrees Celsius to ensure the section stuck to the slide, then dewaxed in xylene followed by absolute alcohol. They were stained with H&E using our standard departmental protocol (see Table 2) and dehydrated in absolute alcohol and xylene. Slides were cover slipped with an automated coverslipper (Leica coverslipper CV 5030) using glass coverslips (Surgipath Premier cover glass 24x50#1,5 95.8877 Rev D).

Step	Solution	Immersion (hrs:mins)
1	Formal Saline (10%)	0:15
2	70% Alcohol	0:30
3 & 4	Absolute Alcohol (Genta medical)	0:30
5	Absolute Alcohol	1:00
6 & 7	Absolute Alcohol	1:30
8	Xylene (Genta Medical)	1:00
9 & 10	Xylene	1:30
Wax 1	Wax (VWR wax 36114)	1:00
Wax 2	Wax	1:30
Wax 3	Wax	1:45

Table 1 Tissue processing protocol used.

Step	Tub	Solution	Time (mins/secs)
0	Load		
1	Oven (To melt wax)		2:00
2	1 (To de-wax slides)	Xylene	2:00
3	2	Xylene	2:00
4	3	Xylene	1:00
5	4 (To clear xylene)	Absolute Alcohol	1:00
6	5	Absolute Alcohol	0:45
7	6	Absolute Alcohol	0:45
8	Wash 1 (To clear slides of Alcohol)	Water	1:00
9	7	Haematoxylin (Mayers)	2:15
10	Wash 2 (To clear slides of Hamatoxylin)	Water	1:00
11	8 (To turn the nuclear staining blue)	Scott's Tap water substitute	0:30
12	Wash 3 (To clear slides of Scott's)	Water	1:00
13	9	Eosin	1:45
14	Wash 4 (To clear slides of Eosin)	Water	0:45
15	10 (To dehydrate slides)	Absolute Alcohol	0:30
16	11	Absolute Alcohol	0:30
17	12	Absolute Alcohol	0:30
18	13	Absolute Alcohol	0:30

19	14 (To clear slides of Alcohol)	Xylene	0:30
20	15	Xylene	0:30
21	16	Xylene	0:30
22	17	Xylene	0:30
23	18	Xylene	0:30

Table 2 Haematoxylin and Eosin staining protocol used

After staining the slides were scanned with an Aperio AT virtual slide scanner (Aperio, Vista, CA, USA) using an Olympus UPLSAPO 20x objective (Olympus, Southend-on-Sea, UK) and Techniquip Model 21DC Light source (Techniquip, Pleasanton CA, USA) , with a GE quartzline projector lamp model EKE 21V 150W (General Electric, Fairfield, CT, USA) . Images were scanned in one batch to produce a whole slide image ("virtual slide") with a final resolution of 0.46 microns per pixel and a single focal optimal focal plane as determined by the automatic focus detection system on the scanner. The virtual slides were compressed with conventional JPEG compression, quality score 70 and saved using the Aperio SVS file format.

Figure 1 shows the colour quantification process. To select regions of interest in which to quantify the colour, virtual slides were viewed in Aperio Imagescope software version 11 (Aperio, Vista, CA). Rectangular regions of interest (ROIs) were drawn at selected locations by a histology technician trained in tissue recognition. One ROI was drawn on the background of the slide (i.e. a white area of the image free of tissue, representing empty "glass"); a second ROI was drawn as a minimal bounding box around the whole cross section of the appendix.

Custom image analysis software was developed in Matlab version 11 (Mathworks, USA) to estimate the amount of each histochemical stain present in the whole slide image. The algorithm design is as follows.

The hand drawn ROIs were extracted from the virtual slides. Two images representing the intensity of the component stains (Haematoxylin and Eosin) were generated using colour deconvolution [7] and

represented as a Haematoxylin channel and an Eosin channel image. The staining vectors used had been calculated by measurement of staining of parallel sections of control appendix tissue stained with only one stain (i.e. either haematoxylin or eosin).

[FIGURE 1 SHOULD GO HERE]

Non-tissue pixels (i.e. background glass) were subtracted from each image. The pixel values from each staining channel were quantised into ten bins of ten percent, based upon the intensity of the stain identified by colour deconvolution. The mean, median and standard deviation of the overall intensity values were calculated for each slide, expressed as a number from 0 (no stain) to 1 (complete saturation with stain). A H:E ratio was then calculated for each by dividing the median intensity of haematoxylin by the median intensity of eosin. This ratio replicates pathologist's understanding of the images and corresponded reasonably well to how "haematoxylinophilic" or "eosinophilic" the images were.

To determine the repeatability of the measurements on the same instrument over time, 5 slides prepared as described above and chosen to represent a range of staining from eosinophilic to haematoxylinophilic were scanned twice each on three separate days (Day 0, Day 12, and Day 64) on the same scanner. On day 64, all 5 slides were also scanned on 3 other instruments (one XT and two AS scanners) to allow inter-scanner comparison. Slides were kept in a cool dark place between measurements.

Experiment 2: Intra-laboratory quantification of stain intensity

30 serial sections of appendix tissue, handled as described in experiment 1, were produced. Using three automated histochemical stainers (Leica autostainer XL 045638879), one slide was stained every working day for two weeks as part of the routine laboratory workload. Unstained slides were included with other cases in the normal laboratory workload for staining. At the end of

the two week period, all of the stained slides were scanned in one batch on the same scanner and measurement of stain intensity and H:E ratio was performed as described in experiment 1.

Experiment 3: Inter-laboratory comparison of stain intensity

Ten serial sections of appendix tissue, handled as described in experiment 1, were produced. Two slides each were sent by post to five different laboratories within the Yorkshire region of the UK. All laboratories were general hospital histopathology/cytopathology laboratories with a minimum workload of 30,000 specimens per year.

Each laboratory was asked to stain the two slides using their standard H&E protocol, including the slide in a batch of routine diagnostic cases. Slides were returned by post after a delay of up to 4 weeks.

Once all slides had been returned, all of the slides were scanned in one batch on the same scanner and measurement of stain intensity and H:E ratio was performed as described in experiment 1.

Statistical analysis

Statistical analysis was performed with SPSS for Windows 19 (IBM Corp, Armonk, NY, USA). In experiment 1, repeatability was assessed using Pearson's correlation coefficient and Bland Altman plot was used to exclude systematic bias. In other experiments nonparametric tests (Mann-Whitney U test and Kruskal-Wallis test) were used to compare groups (on the assumption that the data were not normally distributed). A P value of less than 0.05 was considered statistically significant.

Results

Experiment 1: Validation of stain quantification method

Five samples underwent repeated scanning. The H:E ratio for each is shown in table 3. The ratios for the samples are all under 1, because the amount of eosin in the tissue was greater than the amount of haematoxylin.

	Variance	Standard deviation	Mean difference as percent of measurement
Same day, same scanner	22.0E-6	2.5E-3	0.47%
Different day, same scanner	114.8E-6	7.6E-3	8.32%
Same day, different scanner	519.0E-6	22.3E-3	7.00%
Different day, different scanner	372.3E-6	19.3E-3	8.30%

Table 3 Variability in H:E ratio within and between scanners. Variance and standard deviation are expressed in scientific notation (e.g. 22.0E-6 means 22.0 x 10^{-6})

The repeatability of the measurement on the same scanner, on the same day, was very high (Pearson correlation coefficient 0.995, P <0.001). The mean magnitude of difference between the two paired measurements was 0.47%, the standard deviation was 2.5E-3and variance was 22.0E-6). A correlation plot and Bland-Altman plot showed no systematic bias in the measurements.

The reproducibility of the H:E ratio measurement on the same scanner, on different days (Days 0, 12 and 64) was lower, with a mean difference of 8.32%, a standard deviation of 7.6E-3 and variance of 114.8E-6. There was no systematic change in H:E ratio across time (i.e. no trend towards increasing or decreasing).

The reproducibility of the measurement on different scanners on the same day was calculated using the second Day 64 scans on all four instruments. This showed a strong correlation between the two

measurements (Pearson correlation coefficient 0.996 to 0.999) but higher variability than the other two comparisons with a mean difference of 7.00%, standard deviation of 22.3E-3 and variance of 519.0E-6.

Experiment 2: Intra-laboratory quantification of stain intensity

[Figure 2 should go here]

[Figure 3 should go here]

Figure 2 shows a boxplot of the H:E ratios for the three automated staining machines used in the experiment.. The median H:E ratio for staining machine 1 was lower than the other two machines (0.69 versus 0.81 and 0.93, $P < 0.01$). Standard deviations for the H:E ratios were 0.11, 0.07, and 0.37 for staining machine 1, 2 and 3 respectively.

The granular differences in H and E intensity between the staining instruments are demonstrated in figure 3. Statistically significant differences in the intensity of individual stains were seen - machines 1 and 2 had a higher eosin intensity than 3, and machine 2 had a higher haematoxylin intensity than 2 and 3 ($P < 0.05$). H and E intensity varied independently of each other in most cases.

[Figure 4 should go here]

Plotting the H:E ratio over time (see figure 4) confirmed that staining machine 3 showed significant variability in staining, with a H:E ratio ranging from 0.7 to over 1.7 . The other two staining machines were more consistent, with H:E ratios varying between 0.58 and 0.75. Two large upward spikes in H:E ratio were observed – one with machine 3 at the start of the experiment and one with machine 1 in the middle of the experiment.

Experiment 3: Inter-laboratory comparison of stain intensity

[Figure 5 should go here]

Figure 5 shows the H:E ratio for the six laboratories in the experiment. . The mean H:E ratio across the 6 sites was 0.64, with a standard deviation of 0.12. Five of the laboratories had similar H:E ratios between 0.57 and 0.63 but one laboratory (laboratory E) had a significantly higher H:E ratio of 0.89 ($P < 0.05$).

Even between laboratories with a similar H:E ratio, there was significant variability in the intensity of the haematoxylin and eosin components of the stains. For example, although laboratories A and D had similar H:E ratios (0.57 and 0.62 respectively) , both median haematoxylin and eosin intensity were slightly higher by 32% and 21% respectively in laboratory D (but these differences were not statistically significant).

Discussion

Experimental results

That there is variability in staining in laboratory practice is well known, but this variability has not been quantified in a way that can be used in routine laboratory quality control or quality assurance. Others have attempted this using simple measurements of grayscale intensities to compare tissue staining [8]. Although this gives an overview of stain intensity it does not allow for a granular understanding of the contribution of each stain component.

In this work we have developed a tool to measure this variability using standard whole slide imaging and relatively simple image analysis. Our method allows for the automated quantification of stain intensity and proportion. It is repeatable when the slides are scanned on the same scanner in the same batch, under which circumstances the measurements vary by less than 0.5% - a precision more than sufficient to detect visually important differences in staining.

The image analysis method is based on the assumption that the amount of stain present in the tissue can be estimated with image analysis. Whole slide images are colour digital images, and the colour of each pixel (recorded as an amount of red, green and blue by the sensor on the whole slide imaging device) represents the overall colour of that point in the image. This overall colour has been created by the interaction of the various histochemical dyes applied to the tissue (as well as their reactions with other reagents and tissue components). It is possible to estimate the overall contribution of each stain to the final colour using a technique called colour deconvolution [7]. This allows the image to be split into separate “channels” representing the intensity of each stain component (in this example, a haematoxylin channel and an eosin channel) which can then be quantified separately. Although the human eye can relatively easily discern different component stains in images, it cannot easily determine when stains overlap or co-exist in the same compartment (for example, eosin stains the protein structures in nuclei but the degree of nuclear

eosin staining is not normally appreciated by eye except in experienced observers). The method therefore provides an objective and accurate way to quantify histochemical stains.

However, measurements taken with different scanners or on different days (a reproducibility measure) showed significant variation, up to 8%, which was of a magnitude similar to visually apparent differences in colour - i.e. the noise was comparable in magnitude to the differences being measured.

Given that the method was repeatable on the same day, and the slides would not have been able to fade significantly in the short duration of this experiment, these differences in measurement (between scanners and on different days) can be ascribed as being due to differences in the virtual slide image – either due to illumination or some difference in performance of the image capture system. These variations cannot be predicted or controlled in the image processing algorithm used to measure the stain intensity, and would almost certainly also apply to other image processing algorithms used to measure intensity with colour deconvolution (and possibly other methods of intensity measurement). This has implications for the use of image analysis in digital pathology – care must be taken to ensure that intensity measurements are repeatable between devices and over time. This problem of variation of colour in virtual slides is being increasingly recognised, with a recent International Colour Consortium Summit on medical imaging [9] and efforts are underway to develop standards for calibration of virtual slide scanners to address it [10].

Once validated, this technique was used to quantify differences in staining between individual automatic staining machines, between laboratories, and to show temporal variability in staining, as well as using the method in an experiment to study the effect of various parameters on staining.

Applying the method in one laboratory showed differences in performance of automated staining machines both between machines and over time. Staining machines showed significant variability in

colour with a H:E ratio ranging between 0.60 and 0.80, resulting in a wide variation in the colour of the section from eosinophilic to haematoxylinophilic.

Staining machine 1 had a significantly lower H:E ratio during the period of observation, resulting in a more eosinophilic image. Further investigation found that this particular machine was used more frequently and for larger tissue samples (on average staining approximately 630 slides in a day, including large format 3”x 2” slides, compared to staining machines 2 and 3 at about 450 slides on average, mostly small biopsies). Large sections of tissue use more of the haematoxylin dye (they have more cells to absorb haematoxylin compared to small biopsies), diluting the concentration of dye remaining in the machine. As more racks of slides were handled by machine 1, the water used to wash the slides is carried on the racks into the stain baths, resulting in the concentration of Haematoxylin in particular being diluted. The overall result is a reduction of the haematoxylin intensity on machine 1, leading to a lower H:E ratio.

Plotting the H:E ratio of the three staining machines over time revealed two peaks in H:E ratio. These peaks corresponded with the last day of the maintenance cycle of each machine, when the dye reservoirs were replenished, returning the H:E ratio to normal ranges.

Comparing the stains obtained between six regional laboratories showed a small amount of variability in the H:E ratio for five labs (H:E ratio 0.57 to 0.63), with one laboratory outlying the others at 0.87. These inter-laboratory differences were well known by local pathologists, but the difference has never been quantified before. The variation between the laboratories may be due to differences in staining protocol as well as the dyes used – laboratory E uses Mayer’s haematoxylin solution, applied progressively, while the other laboratories use Harris haematoxylin regressively, differentiated with 1% acid alcohol. As Mayer’s haematoxylin stains mucin but Harris does not, and the appendix contains significant amounts of mucin, the higher H:E ratio for laboratory E might be explained by staining of appendiceal mucin by the Mayer’s haematoxylin.

Implications and benefits of this method

The variability in histochemical staining within and between laboratories has implications for diagnostic accuracy and quality control in current laboratory practice, but also for the repeatability and reproducibility of image analysis algorithms for diagnostic use in future. Having access to a quantifiable measure of stain intensity which can be readily and repeatedly performed in a mostly automated fashion would be potentially of use in addressing all of these issues. Using whole slide imaging to assess the colour removes the need for specialised colorimetry equipment and allows analysis to be done "offline", at a time remote from image acquisition, possibly as part of a central quality control process.

The method demonstrated here is relatively simple and could be replicated by any reference laboratory with access to appropriate imaging facilities. Sections of appendix were chosen because this tissue is relatively easy to procure in a quantity large enough to permit subdivision into aliquots, has a balance of haematoxylinophilic mucosa and eosinophilic muscle, and has relatively little variation in tissue composition between serial sections over a long distance (hundreds or thousands of microns). Setting up the method requires calibration of the colour deconvolution algorithm. This involves measuring the colour distribution on a slide stained with just one dye, and repeated for each dye composing the final stain (in this case, using one control slide for haematoxylin and one for eosin). The method could be applied to any histochemical stain or combination of stains, up to a combination of 3 or 4 stains depending on how much their colours overlap spectrally.

In our experiments, we made use of the whole slide imaging facility at the University of Leeds [11], which allowed rapid scanning of large batches of slides without significant human input. But in principle the algorithm we describe can be applied to any digital image which has been acquired with a consistent light source and optical system, such as a calibrated microscope-mounted digital camera.

The method described could readily be embedded as a quality control or audit process in routine laboratory workflow. It could also be used to assess the impact (or otherwise) of changes to laboratory workflows, such as changes to fixation or tissue processing.

The image processing used by our method is simple and not processor-intensive, so could easily be embedded on a chip within a whole slide imaging device or even automated staining device to deliver quality control of stains in real time.

At a broader level, the process could be applied in inter-laboratory quality assurance programs. The software can be run remotely so a quality program could run the algorithm on a participating laboratory's images without the need to transport glass slides.

In using this quantification method in quality control, it should be remembered that showing a variation in staining does not necessarily imply that a section is not fit for purpose. The larger variations in staining quantified in this work were detectable by eye, and some slides were noticeably too haematoxylinophilic or eosinophilic, but even the most abnormally stained sections were probably usable for diagnostic purposes to a greater or lesser extent. It is likely that, when used in a quality assurance or quality control setting, this method would be used in conjunction with agreed acceptable limits of variation for each stain, against which quality could be measured.

The method itself can provide very granular information about the intensity of staining. In the demonstrations in this paper, the technique was applied over a whole section, and only three metrics were extracted from the slide (Intensity of haematoxylin, intensity of eosin, and H:E ratio). However the method can be applied with greater spatial precision – for example to measure histochemical staining at different points on the slide, or within tissue compartments (mucosa, submucosa, muscle) cell types (eosinophils, basophils, mast cells) or cellular subcompartments (nucleus, cytoplasm) – by applying regions of interest systematically in those areas. This more detailed analysis has been performed on this dataset, but is not shown here.

Similarly the numerical information extracted from the analysis could be far more detailed, including histograms of staining intensity within regions of interest (showing more subtle shifts in stain intensity). Spatial and numerical information may be combined in visualisations such as heat maps to indicate areas of higher or lower intensity in relation to the original image. This method might suit application to biomarker discovery imaging or tissue microarray stain quality assessment.

Limitations

This method was designed to be applied within normal routine laboratory practice without the need for special colour measurement equipment. It is important to note that the method produces an estimate of the true amount of dye on the slide, not an absolute measurement. The method relies on the red, green and blue colours detected by the scanner image capture device and uses colour deconvolution to separate the likely contribution of each dye. In order to obtain an absolute measurement of the amount of dye present, specialised colorimetry or spectral imaging methods would have to be employed – these are technologies not available in most diagnostic laboratories, and which cannot easily be applied in bulk to pathological sections.

Colour deconvolution is a software method to separate or unmix colours which have been combined in an image, without the need to use specialist hardware acquire multi-spectral images. Deconvolution may produce inaccurate results in two situations. First, when the different histochemical stains applied overlap significantly in colour the method may fail to completely separate them, leading to uncertainty in the intensity and distribution of each component stain. Although haematoxylin and eosin are rather similar in terms of RGB values, they are separable by deconvolution. Second, when the dyes are very concentrated and become saturated, colour deconvolution may fail to distinguish which of the constituent components are causing the saturation. In practice, this second issue would be less likely to happen, as a human operator would recognise that a slide was grossly overstained, and in our experiment neither of the dyes approached saturation in any of the samples. The choice of deconvolution vector may also affect the

output metrics of the algorithm – in this work a vector was derived from two control images stained with only Haematoxylin or Eosin, scanned on a single scanner, and the same vector was applied across all experiments to perform the deconvolution. Of course the inherent variability in colour reproduction of that scanner will also affect the colour estimation for deconvolution vector calculation. But as long as the same vector is used across all comparisons it will not introduce additional variability into the derived stain intensity metrics, allowing comparison of metrics between scanners and staining machines.

We employed a derived metric - the H:E ratio - to compare stain intensity. This is a simple ratio which corresponded reasonably well to the apparent colour of the slide (i.e. haematoxylinophilic or eosinophilic). However further work is needed to investigate the independent differences between H and E staining in laboratory practice.

The colours on a virtual slide image are a representation of the colours in the original slide, but not always an exact reproduction of them. The effect of scanner illumination, scanner and slide optics, on-board image processing (e.g. compression and sharpening) may all interfere with the colours reproduced in the final virtual slide image. As such, the colours measured in the whole slide image may be significantly different from those that might be measured with a colorimeter. However the method is still useful as a relative colour intensity measure, and using the same instrument in one laboratory should produce reasonably consistent results (though we have not carried out a long-term evaluation of the repeatability of our method).

There are almost certainly significant differences between whole slide imaging systems, and possibly between different scanner models from the same vendors. These differences could have a significant effect on measurements taken with our method. Furthermore, the reliability of colour measurements could vary depending on the colour, as the spectral response of different detection systems may be different. Future efforts to develop colour profiles to describe and control the colour variability between devices should address this and the previous problem.

Stain intensity is known to be affected by section thickness [12]. Given that most histochemical stains are light absorbers, they obey Lambert-Beer law and so demonstrate a linear relationship between stain intensity (i.e. optical density of the image) and section thickness. Measuring section thickness accurately requires specialist equipment not readily usable in routine pathology laboratories, such as vertical scanning interferometry [8]. In our experiment, the effect of variation in section thickness was controlled as well as possible by using serial sections of tissue on a microtome and ensuring that tissue blocks were cooled regularly, actions which probably limit the variability in section thickness between samples to 10% as long as the microtome thickness setting is below 10 microns [12]. In real laboratory practice, a large batch of control slides could be cut to the same thickness in this way to provide an appropriate control. If longer term comparison of stain intensity is necessary, then section thickness could be made more consistent using automated sectioning machines or indirectly measured using techniques such as outlined in [12] and [13] . Our work did not evaluate the effect of section thickness on H:E ratio, or whether it affects haematoxylin and eosin quantities differently.

Finally, and perhaps most importantly, while our method allows fast and simple quantification of intensity, it cannot replace the expertise of a well trained experienced histotechnologist. Changes in stain intensity and H:E ratio may be the first indicators of a problem with stain quality, but they are not the only ones. Other factors such as adequacy of fixation and processing, section quality, and more subtle characteristics of staining (like inter-compartment differentiation of eosin) also contribute to the overall quality of a stained section. Our method cannot perform or replace these assessments, but would be suitable as an objective tool to complement them.

Disclosure/ Conflict of interest

The authors have no disclosures or conflicts of interest to declare.

Acknowledgements

The authors would like to acknowledge the assistance of the following histotechnologists who assisted with the inter-laboratory comparisons: Mr. D. Evans (Newcastle Upon Tyne NHS Trust), Mr. D. Bagnall (Mid Yorkshire NHS Trust), Mr. R. Roper (Harrogate and District NHS Trust), Mr. A. Mollitt (Calderdale and Huddersfield NHS Trust), and Mr. D Blythe (Leeds Teaching Hospitals NHS Trust).

Ms. G. Jagger and Mr. M. Brinkworth at the University of Bradford assisted with course supervision for Mr. Gray.

Figure Legends

Figure 1. Illustration of colour estimation method used. A section of appendix stained with H&E was scanned to produce a virtual slide. Regions of interest were drawn around the whole appendix (yellow) and a representative area of slide background (green). Image analysis was then used to separate the H&E image into Haematoxylin (left) and Eosin channels (right). These images were used to quantify the intensity of each stain.

Figure 2. Boxplot of H:E ratio of three different staining machines. Median H:E ratios were 0.69, 0.81 and 0.93 respectively. Two example images are included to illustrate appendix samples with H:E ratios of approximately 0.69 (bottom) and 0.93 (top).

Figure 3. Scatterplot of H and E for three different staining machines. The plot shows the results of experiment 2 where 30 serial sections of appendix were stained on three different machines. Median H and E intensities are plotted on X and Y axes respectively, and the number at each point indicates which staining machine was used. The plot shows that H and E vary between samples and staining machines. Although some samples show co-variation of H and E, many do not.

Figure 4. Line plot of H:E ratio over time for three staining machines. Significant variability in staining is seen, with H:E ratio from 0.7 to 1.7. The large upward spikes seen in the HE ratio correspond to the end of the maintenance cycle for two machines, at which the stains and reagents were replenished, returning the H:E ratio to acceptable values.

Figure 5. Scatterplot of H:E ratio for six regional laboratories. Most samples have H:E ratios between 0.57 and 0.63 but one laboratory (E) had a higher ratio of 0.89, indicating that samples were significantly more haematoyxlinophilic.

Take home messages

- Histochemical staining quality is important, but assessment of staining can be subjective and laborious
- Whole slide imaging (digital pathology) can be used to quantify variation in histochemical staining, producing metrics of stain intensity and H:E ratios
- Applied in our practice, we detected significant differences between automated staining devices which correlated with visually detectable differences in staining
- During the validation process we detected inter-instrument scanner variability which merits further work by digital pathology vendors

References

1. Bancroft JD, Gamble M. *Theory and practice of histological techniques*. 6th ed. Edinburgh: Churchill Livingstone; 2006.

2. Roberts N, Magee D, Song Y, et al. Toward Routine Use of 3D Histopathology as a Research Tool. *Am J Pathol*. 2012. doi:10.1016/j.ajpath.2012.01.033.

3. Magee D, Treanor D, Crellin D, et al. Colour Normalisation in Digital Histopathology Images. In: *Proc. Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI workshop)*. London; 2009.

4. United Kingdom National External Quality Assessment Service (UK NEQAS). Available at: http://www.ukneqas.org.uk. Accessed October 2, 2013.

5. The Biological Stain Commission. Available at: http://www.biologicalstaincommission.org/. Accessed December 31, 2013.

6. College of American Pathologists Accreditation and Laboratory Improvement Program. Available at: http://www.cap.org/apps/cap.portal?_nfpb=true&_pageLabel=accreditation.

7. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol Int Acad Cytol Am Soc Cytol*. 2001;23(4):291–299.

8. Elizabeth Little. What difference does a difference make? In: *Pathology Visions 2010*. San Diego; 2010.

9. International Color Consortium Summit on Color in Medical Imaging. In: Silver Spring MD; 2013.

10. Aldo Badano, Craig Revie, Andrew Casertano, et al. Consistency and standardization of color in medical imaging: a consensus report (in preparation).

11. Virtual pathology at the University of Leeds. 2014. Available at: www.virtualpathology.leeds.ac.uk.

12. Cabrini RL, Folco A, Orrea S, Savino MT, Schwint AM, Itoiz ME. A technique for section thickness evaluation for microphotometry and image analysis of sectioned nuclei. *Anal Cell Pathol J Eur Soc Anal Cell Pathol*. 1998;17(2):125–130.

13. Gschwendtner A, Mairinger T. How thick is your section? The influence of section thickness on DNA-cytometry on histological sections. *Anal Cell Pathol J Eur Soc Anal Cell Pathol*. 1995;9(1):29–37.

Figure 1_compressed.tif

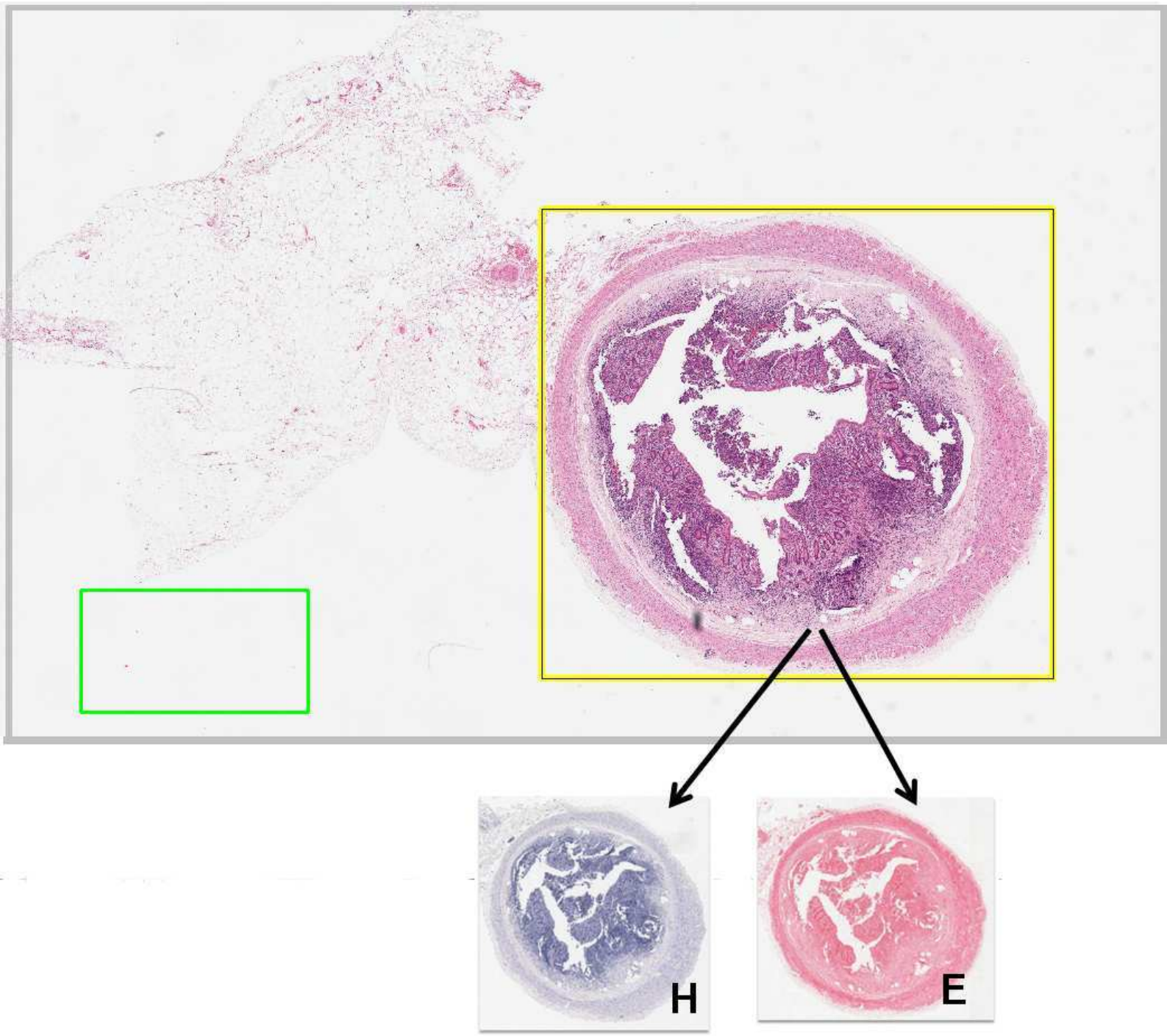

Figure_2_with lines_compressed.tif

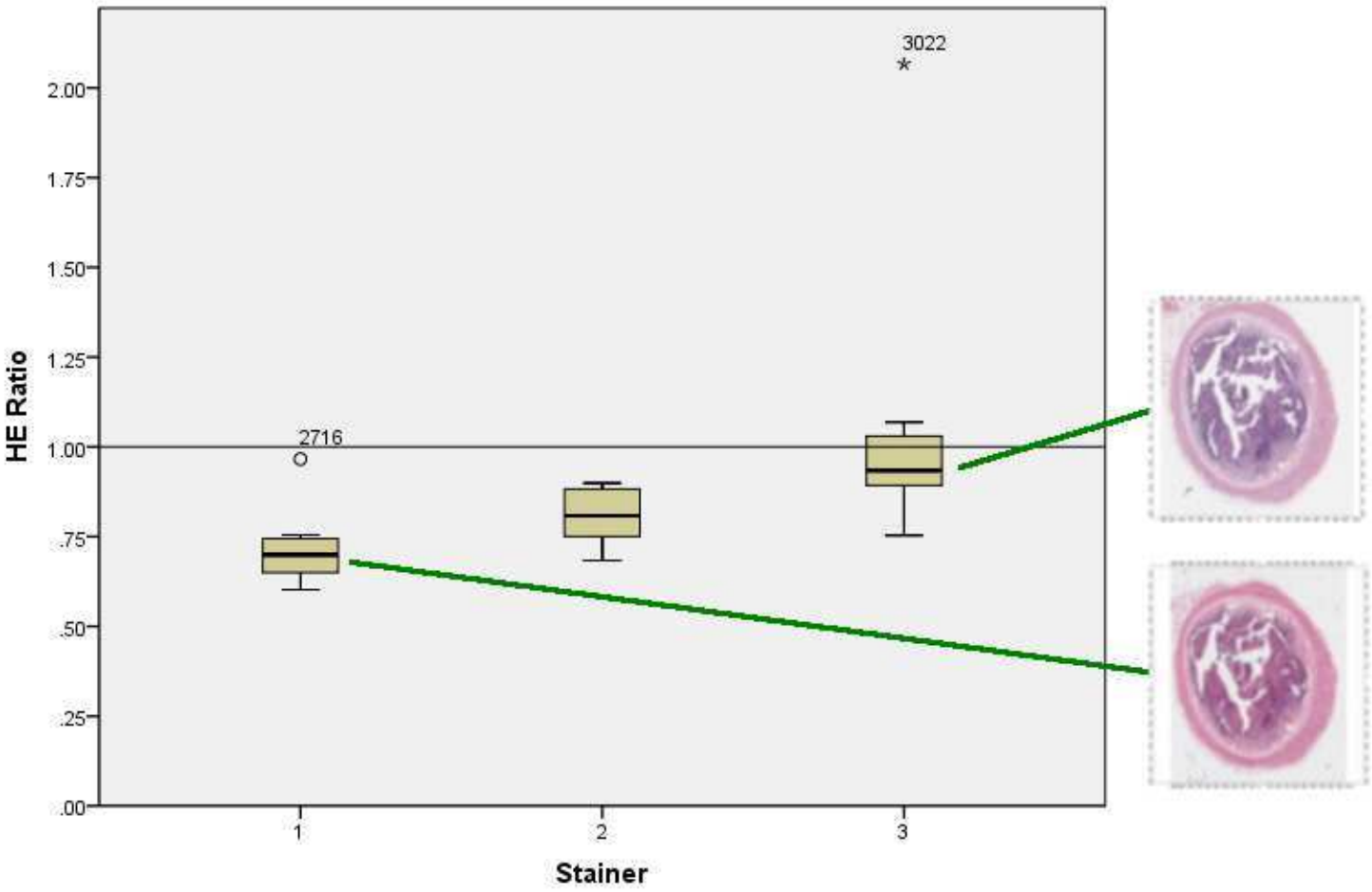

Figure_3_compressed.tif

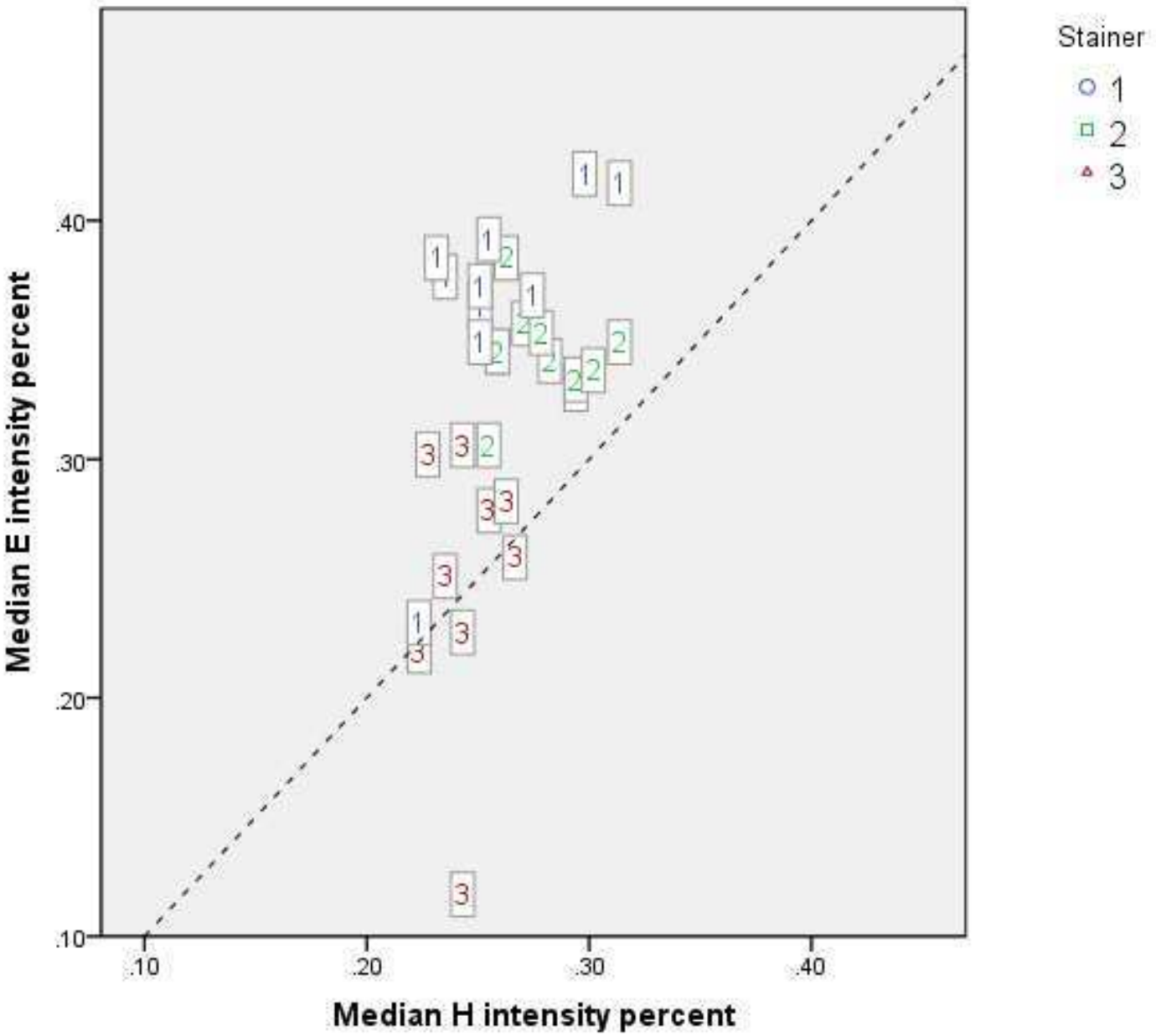

Figure_4_compressed.tif

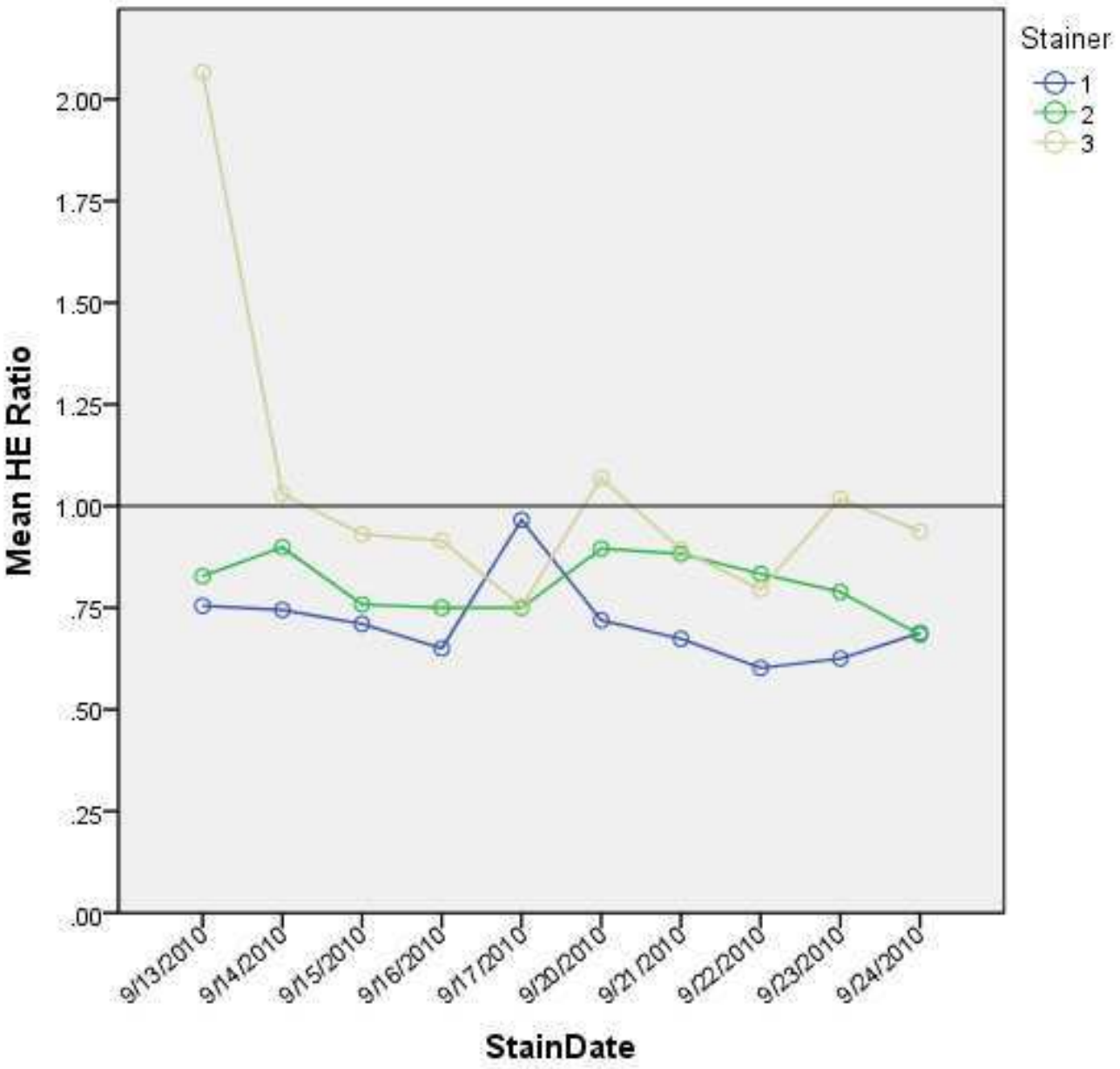

Figure_5_compressed.tif