RfPB project plan for testing reliability and diagnostic accuracy of the Fast Raman device

Summary

The RfPB Fast Raman study of diagnostic accuracy will start on February 3rd, 2020 and involves the measurement of the first Mohs resection layer from a minimum of 170 patients undergoing Mohs surgery at the NUH Treatment Centre for treatment of BCC on the face – head – and neck area. This is a prospective non-inferiority single-centre study designed to determine the diagnostic accuracy of the Fast Raman instrument for the detection of residual basal cell carcinoma on tissue specimens removed through Mohs micrographic surgery. During this study, the reliability of the Fast Raman instrument, which shows to what extent it is prone to random error and to between-operator variations, will also be investigated. A statistical sample size calculation suggests that a sample size of 170 patients will provide an estimate of sensitivity of 96% with (95%CI 0.89, 0.99). In order to measure 170 patients for the study, an average of 4 patients would need to be investigated per week.

Study design

The proposed study will be split into a measurement of instrument reliability and a measurement of instrument diagnostic accuracy.

Measurement of instrument reliability

Instrument reliability will be tested by investigating the same tissue blocks on replicate times. Twenty specimens will be utilised to determine inter-user variability and 20 separate specimens will be utilised to assess the instrument's proneness to random error. To assess inter-user variability of the Fast Raman instrument, the measurement of each of the 20 samples will be performed twice, once by the instrument operator and once by a research nurse. The two users are not allowed to talk or interfere with each other's measurements. Ideally, the specimens investigated in repeat measurements for the reliability test would be re-measured at a later date as a new measurement. This is not possible during the current study as the tissue specimen is processed and destroyed during frozen section histopathology. Therefore the repeat measurements need to be performed immediately after the initial measurements in order to minimise the delay for the patient's surgical procedure.

The measurement entails placing the tissue samples in the sample cassette, closing the cassette, placing the cassette within the instrument and starting a measurement. Sample positioning within the cassette will be decided by each user so that the familiarity that the Raman operator has developed with the tissue samples does not lead to bias in the positioning of the samples into the Raman cassettes for the second measurement.

Sample pre-processing for Fast Raman requires surgeons to cut the specimens as per the standard Mohs procedure, wash them in RBC lysis buffer, blot them to remove superficial blood and then use coloured marker pens to mark the edges on the resulting tissue samples.

Loading the samples into the cassette will be performed by the instrument user. After Raman measurements, the surgeons will need to ink the tissue specimens according to the markings on the cassette coverslip and to send the samples to be processed for frozen section histopathology.

Specimen blotting for blood removal is performed as follows: The sample is blotted with tissue paper in order to remove the majority of the superficial blood. Blotting consists of applying downwards pressure with a finger on the epidermal side of the sample, with the sample being placed between two layers of tissue paper. The sample is then immersed in 1x RBC lysis solution for 5 seconds and then immersed in saline for 5 seconds, to remove the resulting debris from RBC lysis. The sample is then blotted with tissue paper again by firmly pressing on the middle and around the edges on the sample's epidermal side until the sample does not leave blood marks on the bottom layer of the tissue paper. The blotting paper used for each sample will be photographed and logged to investigate the effect of blotting efficacy on the Fast Raman results.

During the measurement procedure, user-induced variation can be introduced in two ways:

- Via sample pre-processing (which aims to remove superficial blood from the sample); this task is currently performed by the surgeon. As blood removal can be only performed once per specimen, this step will not be included in the intra-user variability assessment.

- Via placing the tissue sample inside the cassette (ensuring that the entire resection surface is in contact with the coverslip), placing the sample cassette within the instrument; performed by the instrument operator and the research nurse. These steps are included in the intra-user variability assessment.

In order to mitigate user-induced variation, a full measurement protocol is provided which guides the user through tissue pre-processing and instrument operation. Accompanying video guides which show a full measurement procedure will also be provided.

To assess the instrument's proneness to random error, a separate set of 20 samples will be measured twice without any changes in sample positioning within the cassette. One sample will be measured in repeat times per week, every Monday, for the first 20 weeks of the study. In order to minimise delay to tissue processing for frozen section histopathology, this specimen will be the last specimen measured during the day, when Mohs wait times are generally longer.

Two separate sets of samples will be selected to investigate the instrument's proneness to random error and intra-user variation to reduce measurement time and delay for frozen section histopathology processing.

A kappa coefficient calculation will be performed to determine agreement between the interpretations of the Raman results by multiple observers as it corrects for chance agreement. An acceptable chance corrected agreement index (k) would be over 0.61 for repeat measurements for this study.

<u>Measurement of instrument diagnostic accuracy</u>

The validity of the instrument entails the comparison between the diagnostic results provided by the Raman instrument and the results following the histopathologic evaluation of the frozen H&E sections. Validity is measured using two primary parameters: sensitivity and specificity. For the current application, sensitivity of detection is the most important performance metric of the instrument. It is of utmost importance to ensure that the entirety of the tumour is removed during the surgical procedure. A lower specificity is permitted, as a false positive detection would only induce the removal of an extra tissue layer. This is preferable to having residual tumour still present in the lesion after the operation. A range of other parameters such as positive and negative predictive values and likelihood ratios are derived from the sensitivity and specificity. In order to ensure that the results of Mohs are not influenced by Fast Raman and vice-versa and that index test bias is kept to a minimum, the Raman operator will perform the index technique first, whilst blinded from the standard of reference. Afterwards, the surgeon will perform the histopathologic diagnosis while not aware of the results of the index test.

The sensitivity of the instrument will be calculated as the probability of a positive measurement for samples which have been determined as BCC positive by frozen section histopathology. The specificity of the instrument will be calculated as the probability of a negative measurement for samples which have been determined as BCC negative by frozen section histopathology. Surgeons are encouraged to be cautious when performing Mohs, in order to maximise probability of complete tumour removal. Therefore there may be instances where the resection surface is clear of BCC but the specimen is deemed BCC positive, such as in instances where there is uncertainty when interpreting the H&E sections. A third category will be therefore created (alongside BCC positive and BCC negative) which will contain specimens which are deemed as BCC positive by the investigation of frozen section histopathology, but for which the surgeon is not certain of whether the tumour is indeed present at the resection surface. The sensitivity and specificity of the instrument will be calculated twice, with the uncertain class being considered as BCC positive and with it being not included, to determine what impact this uncertainty has on the accuracy of the instrument.

What will be regarded as a correct Raman detection?

- In the case of a BCC negative specimen, a Raman result is correct if there are no red segments within the image and incorrect if there is at least one red segment within the image.

- In the case of a BCC positive specimen, a correct detection (red segment) will be located in the region within 5 mm from where the tumour is identified in the frozen section H&E. The 5 mm distance is chosen to be comparable with the size of acceptable resection which would allow investigation via frozen section H&E. An incorrect detection occurs when there is no red segment detected within the 5 mm diameter associated with the BCC occurrence detected via frozen section histopathology.

In order to reduce bias when comparing the fast Raman and the frozen histology results and to ensure that the neither of the two techniques is influenced by the other, an automated software will perform the comparison. The Fast Raman results are displayed as a tissue map where BCC is depicted as red regions. After the Fast Raman measurement, the outline of the investigated specimens will be printed out and handed to the surgeon. This map contains no diagnostic information or features other than the outline of the tissue specimens. The surgeon, following the histopathologic assessment, will annotate the outline map with the location where he suspects cancer cells are present at the resection surface and the approximate size of the

tumours. The annotated map will be digitised and registered to the Fast Raman result. An automated algorithm will compare the two results and will determine whether the Fast Raman BCC detections are within the 5 mm distance from the histopathology detections.

The validity of the instrument can be calculated on a per sample basis, on a per layer basis or on a per patient basis. The statistical analysis will be performed on a per sample basis (which represents the tissue layer split into smaller pieces), as it would best retain the localisation of the tumour within the lesion and would be therefore be truer to a real-life scenario application of the technique.

Reduction of bias

In order to ensure that the study presents a highly accurate assessment of the performance of the Fast Raman instrument in a clinical setting, the study design should attempt to minimise biases. The main biases that could be encountered in a diagnostic tests of accuracy are: selection bias, verification bias, incorporation bias, observer bias and differential reference bias.

Selection bias

Selection bias refers to bias induced by the criteria used in the selection of recruited patients. In the present study, a BCC lesion will be considered for enrolment using predefined inclusion and exclusion criteria:

Patient inclusion criteria and sample size

Minimum 170 consenting patients undergoing Mohs surgery at NUH Treatment Centre for treatment of BCC on face – head – and neck area will be included in the present study.

As part of the study, the entire first resection layer will be measured with the Fast Raman instrument followed by assessment by frozen section histopathology. In instances where the specimen is larger than the 2 cm x 2 cm permitted by the sample cassette, the layer will be spilt into smaller samples (as processed for standard frozen histopathology) and one of the smaller samples will be investigated instead of the entire layer.

In order to ensure that the measurements mimic the spectrum of all patients undergoing Mohs, sample selection will be performed through stratified sampling to simulate the occurrence of specimen types encountered in a common Mohs unit. Surgery records at the NUH Treatment Centre from 29th July to 3rd December show the following sample distribution: nose: 48%, eyelid: 26%, cheek: 10%, lip: 5%, forehead/scalp: 4%, eyebrow: 4%, ear: 3%, neck: 1%. Therefore, the current study aims to investigate 81 nose lesions, 44 eyelid lesions, 17 cheek lesions, 8 lip lesions, 7 forehead/scalp lesion, 7 eyebrow lesions, 5 ear lesions and 2 neck lesions. The first patient during each day of measurement will not be investigated via Fast Raman as the specimens will be sent directly to the histopathology lab, to reduce delays to the patients' standard of care. Specimens from consecutive patients will be investigated via Fast Raman until the required quota is met for a certain tissue type. After a tissue type quota is met, other tissue types will be prioritised. Even if the quota is met for a certain tissue type, samples from that tissue type will still be investigated, provided it does not interfere with meeting the quota for other tissue types.

The surgical excision and histological evaluation will be performed by three surgeons of varying experience: Dr Sandeep Varma, Dr Ashish Sharma and Dr David Vietch.

Patients with multiple BCC lesions will be included for only the lesion most suitable to be investigated with the index test, as determined by the surgeon.

While measuring consecutive patients would theoretically reduce measurement bias and would therefore be more compliant with QUADAS-2 guidelines, stratified sampling would ensure that samples from all anatomical locations are investigated. Though selecting consecutive patients would generally result in a sample set that is diverse, tissue type distribution during the Mohs workflow at the Treatment Centre is surgeon dependent. As not all Mohs surgeons in the centre participate in the study, the sample set would likely be skewed towards disproportionately more eyelid specimens. In order to reduce bias, consecutive specimens will be measured within each tissue type category, provided they are not highlighted by the patient exclusion criteria.

Patient exclusion criteria

After patients have consented to participate in the study, a decision will be made of whether their specimens are suitable to be included. Patients will not be included in the study if their tissue specimens have the following characteristics:

- Layers which are larger than 2 cm in diameter will not be fully analysed, as they do not fit within the sample cassette. Instead, a smaller sample piece, representing ½ - ¼ of resection surface will be investigated instead.

- Samples for which the quality of the histopathology section is not sufficient to allow for a definitive reference standard will also not be included. In order to reduce bias, this decision will be made by the Mohs surgeon by investigating the first 10 consecutive frozen H&E sections (10 μm thickness). This will allow the surgeon to determine if the actual resection surface is represented in the H&E sections. To reduce bias, this decision will be made at the end of each Mohs workday, with the surgeon being blinded towards the Fast Raman results.

- Samples for which the excision surface does not come in contact with the cassette coverslip. The surgeon will be able to use the autofluorescence image to determine whether the resection surface is in full contact with the cassette coverslip. To reduce bias, this decision will be made at the end of each Mohs workday, with the surgeon being blinded towards the Fast Raman results.

Verification bias

Verification bias occurs when only some of the patients who have the index test go on to have the reference standard. This is not the case for this study as all samples which will be measured with the Fast Raman device will be sent to be processed through frozen section histopathology.

Incorporation bias

Incorporation bias occurs when the reference standard is defined in part by a positive index test. This is also not applicable to the study as the reference standard will be performed

regardless of the results of the index test. Moreover, the surgeon that analyses the frozen section histopathology slides as part of the reference standard will be blinded towards the results of the index test.

Observer bias

Observer bias can occur when the person responsible of interpreting the results of the reference standard knows the results of the index test. In order to limit bias, the surgeon will interpret the frozen section histopathology slides whilst being unaware of the Fast Raman results. The surgeon will annotate an outline of the specimen as per the standard Mohs clinical workflow at the NUH Treatment Centre. The diagram will then be compared with the Fast Raman result at the end of the Mohs workday by an automated algorithm.

Differential reference bias

Differential reference bias occurs when patients that are investigated with the index test are then being treated with different reference standards. This is not the case for the present study, as all samples measured with the Fast Raman instrument will be processed for frozen section histopathology.

Imperfect reference standard

Frozen section histopathology as performed for Mohs is an imperfect reference standard for our study as it does not investigate the true resection surface in one image, rather it sections through the sample in 100 μm intervals, likely being slightly biased towards providing positive diagnoses.

Surgeons are encouraged to be cautious when performing Mohs, in order to maximise probability of complete tumour removal. Therefore there may be instances where the resection surface is clear of BCC but the specimen is deemed BCC positive, such as in instances where there is uncertainty when interpreting the H&E sections. A third category will be therefore created (alongside BCC positive and BCC negative) which will contain specimens which are deemed as BCC positive by frozen section histopathology, but for which the surgeon is not certain of whether the tumour is indeed present at the resection surface. The sensitivity and specificity of the instrument will be calculated twice, with the uncertain class being considered as BCC positive and with it being not included, to determine what impact this uncertainty has on the accuracy of the instrument.

One concern is that there could be instances when there is no residual BCC on the resection surface of the investigated sample, but the tumour is brought to the surface by shavings performed by the Mohs technicians when processing the frozen samples. In such an instance, the instrument would correctly identify a lesion as being BCC negative, while histopathology would identify it as being BCC positive. In order to reduce such occurrences, the frozen section H&E technicians will process and stain the first 10 consecutive shavings (10 μm thickness), after which they would process the remaining sections in 100 μm increments. This will allow the surgeon that performs the investigation to assess the agreement between the two techniques within a well-defined error margin and to better determine the quality of the H&E sections.

Update 14/05/2021

Variations to Mohs tissue processing procedure at recruitment centre

From October 2020, the surgery staff at the NUH Nottingham Treatment Centre modified the standard tissue processing procedure for Mohs micrographic surgery. After surgical excision, instead of splitting the tissue layer into smaller specimens, it is now kept as a single specimen. Due to the conical shape of these "full-face" tissue layers, it is more challenging to ensure that the entire resection surface is in contact with the cassette coverslip for Fast Raman measurements.
To increase the likelihood of obtaining the entire resection surface in as few H&E sections as possible (which would suggest a flat resection surface), the layer is frozen flat with the resection surface against a microscope slide during Mohs processing. This procedure is performed after the Fast Raman measurements and cannot be replicated using the Fast Raman cassettes.

Modification to tissue cassette

To account for the decrease in tissue-cassette window contact caused by the change in tissue processing at the NUH Nottingham Treatment Centre, a new cassette lid was developed. The new cassette lid was designed to apply different amounts of pressure to different regions withing the cassette. A set of nine grub screws within the new cassette lid can be tightened or loosened depending on the thickness of a tissue layer at a certain region. This increases the amount of tissue surface area that is in contact with the cassette window, without overly squeezing thicker regions of the layer. This new cassette lid was implemented in the study in April 2021.

Changes to sample size calculation

The original sample size calculation was made prior to the COVID-19 pandemic, which has halted measurements for 6 months and has impacted patient recruitment figures. The sample size calculation was made under the assumption that the prevalence of a BCC-positive first layer during Mohs is 45% (from the literature). During the course of the study, we have observed that the BCC-positive rate at our recruitment centre (NUH Nottingham Treatment Centre) was lower, at ~22%. This may be caused by selection bias, as tissue specimens that are larger than 2 cm, and therefore have a higher likelihood of being BCC-positive, cannot be investigated by the Fast Raman device (limited by the tissue cassette size). In order to predict the sensitivity of the Fast Raman device of 96% with (95%CI 0.81, 0.99) using the binomial exact method, 27 BCC-positive tissue layers need to be investigated. Therefore, tissue layers will be investigated until 27 BCC-positive layers are measured. Based on the observed prevalence, approximately 122 fresh tissue layers will be investigated during this study.