Research for Patient Benefit project plan for testing reliability and diagnostic accuracy of the Fast Raman device

Summary

The Research for Patient Benefit (RfPB) Fast Raman study of diagnostic accuracy will start on March 2nd, 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 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 contain an assessment of instrument reliability and an assessment of instrument diagnostic accuracy.

Measurement of instrument reliability

Instrument reliability will be tested by investigating the same tissue blocks on replicate times. Fast Raman measurements entail 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 surgeon so that the familiarity that the initial 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.

Forty specimens will be utilised to assess inter-user variability. For each Fast Raman measurement preformed during this study, user-induced variation can be introduced in three ways:

- Via sample pre-processing (which aims to remove superficial blood from the sample); this task is 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 specimen inside the sample cassette and ensuring that the entire resection surface is in contact with the coverslip; this task is performed by the surgeon. This step is included in the intra-user variability assessment.

- Via placing the loaded cassette in the Fast Raman instrument and ensuring that the cassette is correctly slotted; performed by the instrument operator and the research nurse. This step is included in the intra-user variability assessment.

As there are two users involved in each measurement, the surgeon and the instrument operator, the variability introduced by each will be assessed independently. Twenty tissue specimens will be used to assess intra-surgeon variability and 20 tissue specimens will be used to assess intra-instrument operator variability.

Twenty Mohs layers will be used to assess inter-surgeon variability in placing the tissue specimens in the cassette. Each layer will be first loaded in the cassette by the surgeon which performed the Mohs procedure and tissue blotting. The specimen will be measured with the Fast Raman device by the instrument operator and the cassette will be returned to the surgeon. The surgeon will then remove the tissue specimens, place them in a petri dish and hand it to a different Mohs surgeon, which will load the tissue specimens in a new cassette. The Fast Raman measurement will then be performed by the same instrument operator.

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, use coloured marker pens to mark the edges on the resulting tissue samples and load the samples into the sample cassettes. After Raman measurements, the surgeons 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.

To assess inter-instrument operator variability, the measurement of 20 separate samples will be performed twice, once by the spectroscopy expert and once by a research nurse. Each of the Mohs layers will be pre-processed and loaded in a cassette by one of the surgeons. The cassette will be loaded in the Fast Raman instrument by the instrument operator and will be measured. Afterwards, the cassette will be handed to the research nurse, which will also measure the same tissue sample, within the same loaded cassette.

As tissue specimens are destroyed during processing for frozen section histopathology, intra-user variability cannot be reliably tested within the Mohs workflow restrictions. An intra-user variability assessment entails the replicate measurement of a specimen by the same user, blindly, as a new specimen. The familiarity of the user withta specimen when performing consecutive measurements leads to bias in cassette positioning and sample placement.

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

A single Mohs layer will be measured per week to assess inter-surgeon and inter-instrument operator variability. These measurements will be performed every Monday, for the first 40 weeks of the study. In order to minimise delay to tissue processing for frozen section histopathology, the last specimen measured during the day will be measured twice, as it is when Mohs wait times are generally longer.

Two separate sets of samples will be selected to investigate inter-surgeon and inter-instrument operator variability to reduce measurement time and delay for frozen section histopathology processing.

Measurement of instrument diagnostic accuracy

The validity of the instrument entails the comparison between the diagnostic results provided by the Fast 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 instrument 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.

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.

What is 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 generated within the image and incorrect if there is at least one red segment generated within the image.

- In the case of a BCC positive specimen, a correct detection (red segment) will be located in the region within 2.5 mm from where the tumour is identified in the frozen section H&E. The 2.5 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 2.5 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 does not contain 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 of the cancer cells on the resection surface and the approximate size of the tumours. The map will be marked in red for certain BCC detections and in green for the presence of other markers which would result in a second layer being removed (e.g. inflammation, missing epidermis, etc.). 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 2.5 mm distance from the histopathology detections. The software will assess whether the Raman measurement was successful and display the results for the user.

Three different spectral classification models will run concurrently during each measurement and will each generate a Fast Raman result map. Each of the three Fast Raman maps will be compared to the H&E assessment independently, with the described automated software. Therefore, three independent validity parameter sets (sensitivity and specificity) will be calculated, each highlighting the performance of one of the classification models. The classification model with the highest BCC detection accuracy will be retained and will represent the diagnostic accuracy of the Fast Raman device.

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 inclusion 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 study.

As part of the study, the entire first Mohs 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 a single, smaller sample 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 delay 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 Veitch.

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 will 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 clinic 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 Fast Raman measurement does not complete successfully due to either software or hardware errors. These measurements will be kept and reported, but they will not be included in the detection accuracy calculations.

- Samples for which any of the following data is missing: consent forms, surgeon BCC annotation maps, H&E sections.

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 maximise the probability of complete tumour removal when performing Mohs. 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.

Statistical analysis plan

The statistical analysis will be performed on a per layer basis (which represents the entire tissue layer removed during a single round of surgery).

Sample size calculation

A sample size of 170 will provide an estimate of sensitivity of 96% with (95%CI 0.89, 0.99), based on a prevalence of tumour clearance for the first Mohs layer of 45%. As Raman assessment is a rule-out test (i.e. we need to ensure any new tissue slice is clear of tumour), the sample size has been based on sensitivity.

Inter-surgeon variability

As Fast Raman results do not require user interpretation, user variance will be determined by calculating the surface area of the tissue specimens which is not in contact with the cassette coverslip. This surface area will denote the efficiency with which the samples are placed within the cassette by the surgeons. As this area is a continuous outcome, interclass correlation coefficient (ICC) will be calculated to assess differences in tissue handling between surgeons. The ICC will be calculated by investigating 20 specimens measured twice, each time being placed within a cassette by different surgeons. The ICC will be interpreted as poor<0.50, fair 0.50 to 0.75, good 0.75-0.9, excellent >0.9.

Inter-instrument operator variability

As Fast Raman results do not require user interpretation, user variance will be determined by calculating the surface area of the tissue specimens which is not in contact with the cassette coverslip. This surface area will denote the efficiency with which the cassettes are placed within the instrument by the instrument operator. As this area is a continuous outcome, interclass correlation coefficient (ICC) will be calculated to assess differences in tissue cassette handling between instrument operators. The ICC will be calculated by investigating 20 specimens measured twice, each time being placed within a cassette by different surgeons. The ICC will be interpreted as poor<0.50, fair 0.50 to 0.75, good 0.75-0.9, excellent >0.9.

Sensitivity and specificity

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.

Sensitivity analysis: 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.

Likelihood ratios will be calculated as:

Positive likelihood ratio = sensitivity / (1-specificity)

Negative likelihood ratio = (1-sensitivity) / specificity

If the prevalence of tumour clearance for the first Mohs layer is the same as in the population, the positive and negative predictive values will be calculated:

Positive predictive value: True positive / (True positive + False positive)

Negative predictive value: True negative (True negative + False negative)