

# Advanced endoscopic techniques and optical diagnosis in the lower gastrointestinal tract

By

SAMUEL CHARLES LLOYD SMITH

A thesis submitted to the University of Birmingham  
for the degree of  
DOCTOR OF PHILOSOPHY

Institute of Immunology and Immunotherapy  
College of Medical and Dental Sciences  
University of Birmingham  
November 2021

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

# Abstract

Inflammatory bowel disease describes debilitating chronic diseases (Ulcerative colitis and Crohn's Disease) of the gastrointestinal tract that requires many patients to take long term medications, have an increased risk of malignancy and can often lead to major abdominal surgery. With the advent of modern therapies, in particular biological therapies, clinicians and patients can increasingly achieve mucosal healing. Mucosal healing is associated with favourable outcomes such as reduced hospitalisation, colectomy and fewer courses of steroids. However, the definition of mucosal healing was based on endoscopic scoring systems developed in a previous generation of endoscopic technology. Novel endoscopic technology, such as Virtual Electronic Chromoendoscopy, can gain more accurate assessments of inflammatory activity, closing the gap to histological activity. The term optical diagnosis refers to the ability of an endoscopist to accurately predict the histology of an endoscopic finding, albeit in predicting inflammatory activity in inflammatory bowel disease or in predicting the histology of a colorectal polyp. With ever improving endoscopic technology successful optical diagnosis is increasingly possible.

The PICaSSO score is a score that was developed using Virtual Electronic Chromoendoscopy and is the first to define endoscopic features in keeping with mucosal healing. This thesis includes a large multicentre international prospective study which investigates the correlation of the PICaSSO score and other established endoscopic scores (MES and UCEIS) against multiple histological indices (RHI, NHI, ECAP, Geboes, and Villanacci) and prospectively assessed outcomes at 6 and 12 months. There was strong correlation between PICaSSO and histology scores, significantly superior to correlation coefficients of MES and UCEIS with histology scores. A PICaSSO score of  $\leq 3$  detected histologic remission by RHI with AUROC 0.90 (95% CI 0.86-0.94). PICaSSO score  $\leq 3$  predicted better outcomes than PICaSSO  $> 3$ .

The next study moves from optical characterisation to molecular characterisation of inflammatory activity using Raman Spectroscopy. Raman Spectroscopy (RS) describes the scattering of inelastic light giving spectra that are highly specific for individual molecules. This study aimed to establish spectral changes before and after treatment and whether Raman Spectroscopy can accurately differentiate between inflammation and MH. Reductions in intensity at 1003cm<sup>-1</sup> and 1252cm<sup>-1</sup> when a reduction in inflammation was seen post-treatment and when MH was present. MH was associated with an increase in intensity at 1304cm<sup>-1</sup>. A trained neural network differentiated MH from active inflammation with high sensitivity, specificity, PPV, NPV and accuracy in UC and CD. Raman Spectroscopy may represent an additional tool in the assessment of mucosal healing in IBD.

To implement optical diagnosis in practice there needs to be robust training. The next study presents a randomised controlled study comparing the performance of self-training vs. didactic training on the diagnostic accuracy of diminutive/small colonic polyp histological prediction by trainees using established polyp classification tools. The study showed self-learning can achieve results similar to didactic training which, could enable widespread implementation of optical diagnosis in clinical practice. Following this study, I present a meta-analysis and systematic review of optical diagnosis training in small/diminutive colorectal polyps. Optical diagnosis training is effective in improving accuracy of histology prediction in colorectal polyps and didactic and computer-based training show comparable effectiveness in improving diagnostic accuracy. The final study presented is implementing optical diagnosis training and setting quality standards in IBD surveillance endoscopy. A significant improvement in quality was seen in the proportion of procedures using dye-based chromoendoscopy, use of polyp classification tools and an increase lesion detection.

Competing Interests: I declare I have no competing interests.

# Acknowledgements

I would, firstly, like to thank Dr Marietta Iacucci and Prof Subrata Ghosh for granting me the opportunity to conduct this research and for their support and expertise in drafting this thesis.

Thank you to the all the lab team, and in particular Dr Louisa Jeffries and Dr Stuart Hunter for all their help and guidance in supporting my lab work. A special thank you to the chemical engineering team, Dr Pola Goldberg Oppenheimer and Carl Banbury for their help, guidance and opportunity to acquire new skills and knowledge.

I would also like to acknowledge the NIHR Birmingham Biomedical Research Centre team, without who this research would not have been possible. A special mention to the research nurses Christine Bossa and Rosemary Matonhodze for their dedicated work. I would like to acknowledge the research coordinators Hollie Caulfield, Helen Duffy, Gail Heritage and Suzanne Lockyer for their fantastic work and endless patience.

I would like to thank my wife Krupa for her continued support without which none of my achievements would have been possible. Finally, my two beautiful girls, Rani Lily and Indi Lyra, you have provided me endless motivation and are a constant source of joy.

# Abbreviations

|          |                                                    |
|----------|----------------------------------------------------|
| 5-ASA    | Aminosalicylic acids                               |
| ACT      | Acute Ulcerative Colitis Trials                    |
| Anti-TNF | Anti-Tissue Necrosis Factor                        |
| ASGE     | The American Society of Gastrointestinal Endoscopy |
| AUC      | Area under the curve                               |
| BASIC    | BLI Adenoma Serrated International Classification  |
| BCSP     | Bowel Cancer Screening Programme                   |
| BLI      | Blue-light Laser Imaging                           |
| CCD      | Charged Coupled Devices                            |
| CD       | Crohn's Disease                                    |
| CE       | Contrast Enhancement                               |
| CI       | Confidence Interval                                |
| CIR      | Caecal Intubation Rate                             |
| CLE      | Confocal Laser Endomicroscopy                      |
| CRC      | Colorectal Cancer                                  |
| CRF      | Case Report Forms                                  |
| CRP      | C-reactive protein                                 |
| DCE      | Dye-based Chromoendoscopy                          |
| ECAP     | Extent, Chronicity, Activity Plus others           |
| ECCO     | European Crohn's and Colitis Organisation          |
| ER       | Endoscopic Remission                               |
| ESGE     | European Society of Gastrointestinal Endoscopy     |

|          |                                                                        |
|----------|------------------------------------------------------------------------|
| HC       | Healthy Control                                                        |
| HD       | High Definition                                                        |
| HDWL     | High-Definition White Light                                            |
| HR       | Histological Remission                                                 |
| IBD      | Inflammatory Bowel Disease                                             |
| ICC      | Intra-class correlation coefficient                                    |
| IOIBD    | The International Organization for Study of Inflammatory Bowel Disease |
| iScan-OE | iScan Optical Enhancement                                              |
| JNET     | Japanese NBI Expert Team                                               |
| LCI      | Linked Colour Imaging                                                  |
| LDA      | Linear Discriminant Analysis                                           |
| MES      | Mayo Endoscopic Score                                                  |
| MH       | Mucosal Healing                                                        |
| NBI      | Narrow Band Imaging                                                    |
| NICE     | The National Institute of Health and Care Excellence                   |
| NICE     | NBI International Colorectal Endoscopic                                |
| NPV      | Negative Predictive Value                                              |
| NS       | Nil significant                                                        |
| OR       | Odds Ratio                                                             |
| PCA      | Principal Component Analysis                                           |
| PGA      | Physician Global Assessment                                            |
| PICaSSO  | The Paddington International virtual ChromoendoScopy ScOre             |
| PIVI     | Preservation and Incorporation of Valuable endoscopic Innovations      |
| PPV      | Positive Predictive Value                                              |

|        |                                                                                                                  |
|--------|------------------------------------------------------------------------------------------------------------------|
| PSC    | Primary Sclerosing Cholangitis                                                                                   |
| RHI    | Robarts Histological Index                                                                                       |
| ROC    | Receiver Operating Characteristics                                                                               |
| RR     | Relative Risk                                                                                                    |
| RS     | Raman Spectroscopy                                                                                               |
| SCCAI  | Simple Clinical Colitis Activity Index                                                                           |
| SCENIC | Surveillance for Colorectal Endoscopic Neoplasia Detection and Management in Inflammatory Bowel Disease Patients |
| SD     | Standard Deviation                                                                                               |
| SE     | Surface Enhancement                                                                                              |
| SES-CD | Simple Endoscopic Score for Crohn's Disease                                                                      |
| SIMPLE | Simplified Identification Method for Polyp Labelling during Endoscopy                                            |
| SKiNET | Self-optimising Kohonen Index Network                                                                            |
| SOM    | Self-organising Map                                                                                              |
| SOMDI  | Self-organising Map Discriminant Index                                                                           |
| SSA/P  | Sessile Serrated Adenoma/Polyps                                                                                  |
| SSL    | Sessile Serrated Lesion                                                                                          |
| STRIDE | Selecting Therapeutic Targets in Inflammatory Bowel Disease                                                      |
| TE     | Tone Enhancement                                                                                                 |
| UC     | Ulcerative Colitis                                                                                               |
| UCEIS  | Ulcerative Colitis Endoscopic Index of Severity                                                                  |
| VEC    | Virtual Electronic Chromoendoscopy                                                                               |
| WASP   | Workgroup serrated polyps and Polyposis                                                                          |
| WLE    | White Light Endoscopy                                                                                            |

## Publications arising from this thesis

**Smith SCL**, Banbury C, Zardo D, Cannatelli R, Nardone OM, Shivaji UN, Ghosh S, Oppenheimer PG, Iacucci M. Raman Spectroscopy can accurately differentiate mucosal healing from non-healing in inflammatory bowel disease. PLoS ONE 2021 (EPub)

Iacucci M, **Smith SCL**, Bazarova A, Shivaji UN, Bhandari P, Cannatelli R, Daperno M, Ferraz J, Goetz M, Gui X, Hayee B, Hertogh G, Lazarev M, Li J, Nardone OM, Parra-Blanco A, Pastorelli L, Panaccione R, Occhipinti V, Rath T, Tontini GE, Vieth M, Villanacci V, Zardo D, Bisschops R, Kiesslich R, Ghosh S. An International Multicentre Real-Life Prospective Study of Electronic Chromoendoscopy Score PICaSSO in Ulcerative Colitis. Gastroenterology 2021; 160: 1558-1569

**Smith SCL**, Siau K, Cannatelli R, Antonelli G, Shivaji UN, Ghosh S, Saltzman JR, Hassan C, Iacucci M. Training methods in optical diagnosis and characterisation of colorectal polyps: a systematic review and meta-analysis. Endosc Int Open 2021; 9: E716-E726

**Smith SCL**, Cannatelli R, Bazarova A, Sharma N, McCulloch A, Mak J, Shivaji UN, Iqbal T, Kane K, Ghosh S, Cooney R, Iacucci M. Performance measures in IBD surveillance colonoscopy: Implementing changes to practice improves performance. Dig Endoscopy 2020; 32: 592-599

**Smith SCL**, Saltzman J, Shivaji UN, Lethebe BC, Cannatelli R, Birmingham Colonic Polyp Characterisation Group, Ghosh S, Iacucci M. A randomised controlled study of the prediction

of diminutive/small colorectal polyp histology using didactic vs. computer based self-learning module in gastroenterology trainees. Dig Endoscopy 2019; 31: 535-543.

## Publications supported by this thesis

Iacucci M, McQuaid K, Gui SX, Iwao Y, Lethebe BC, Lowerison M, Matsumoto T, Shivaji UN, **Smith SCL**, Subramanian V, Uraoka T, Sanduleanu S, Ghosh S, Kiesslich R. A multimodal (FACILE) classification for optical diagnosis of inflammatory bowel disease associated neoplasia. *Endoscopy* 2019; 51: 1-10

Iacucci M, Furfaro F, Matsumoto T, Uraoka T, **Smith S**, Ghosh S, Kiesslich R. Advanced endoscopic techniques in the assessment of inflammatory bowel disease: new technology, new era. *Gut* 2019; 68: 562-572.

Cannatelli R, Bazarova A, Zardo D, Nardone OM, Shivaji UN, **Smith SCL**, Gkoutos G, Ricci C, Gui XS, Ghosh S, Iacucci M. Fecal calprotectin thresholds to predict endoscopic remission using advanced optical enhancement techniques and histological remission in IBD patients. *Inflamm Bowel Dis* 2020; 27: 647-654

Iacucci M, Jeffrey L, Acharjee A, Nardone OM, Zardo D, **Smith SCL**, Bazarova A, Cannatelli R, Shivaji UN, Williams J, Gkoutos G, Ghosh S. Ultra-high magnification endocytoscopy and molecular makers for defining endoscopic and histological remission in ulcerative colitis- an exploratory study to define deep remission. *Inflamm Bowel Dis* 2021; 27: 1719-1730

Nardone OM, Bazarova A, Bhandari P, Cannatelli R, Daperno M, Ferraz J, Goetz M, Gui XS, Hayee B, De Hertogh G, Lazarev M, Li J, Parra-Blanco A, Pastorelli L, Panaccione R,

Occhipinti V, Rath T, **Smith SCL**, Shivaji UN, Tontini GE, Vieth M, Villanacci V, Zardo D, Bisschops R, Kiesslich R, Ghosh S, Iacucci M. Endoscopic remission assessed with PICA<sup>SSO</sup> electronic chromoendoscopy score is not inferior to combined endoscopic and histological assessment for prediction of clinical outcomes in ulcerative colitis. United European Gastroenterol Journal 2021 (In press)

# Presentations arising from this thesis

## **International meetings:**

### Oral presentation at Digestive Diseases Week 2020, Chicago USA:

- Raman Spectroscopy can differentiate mucosal healing from non-healing in inflammatory bowel disease
- The first real-life multicenter prospective validation study of the electronic chromoendoscopy score (Paddington International Virtual Chromoendoscopy Score) and its outcome in Ulcerative colitis

### Poster presentation at Digestive Diseases Week 2020, Chicago USA:

- Raman Spectroscopy demonstrates biomolecular changes and predicts response to biological therapy in inflammatory bowel disease

### Oral presentation at European Society of Gastrointestinal Endoscopy 2020,

#### Dublin Ireland:

- The first real-life multicenter prospective validation study of the electronic chromoendoscopy score (Paddington International Virtual Chromoendoscopy Score) and its outcome in Ulcerative colitis

E-poster presentation at European Society of Gastrointestinal Endoscopy 2020,

Dublin Ireland:

- Raman Spectroscopy can differentiate mucosal healing from non-healing in inflammatory bowel disease
- Raman Spectroscopy demonstrates biomolecular changes and predicts response to biological therapy in inflammatory bowel disease

Poster presentation at Digestive Diseases Week 2019, San Diego USA:

- Performance measures in IBD surveillance colonoscopy-implementing changes to practice improves performance.

Oral presentation at European Society of Gastrointestinal Endoscopy 2019,

Prague Czech Republic:

- Performance measures in IBD surveillance colonoscopy-implementing changes to practice improves performance.

Poster presentation at United European Gastroenterology Week 2018, Vienna

Austria:

- A randomised controlled study of the prediction of diminutive/small polyp histology using didactic vs. computer-based self-learning module in gastroenterology trainees.

Poster presentation at Digestive Diseases Week 2018, Washington USA:

- A randomised controlled study of the prediction of diminutive/small polyp histology using didactic vs. computer-based self-learning module in gastroenterology trainees.
- Endoscopic characterisation of small/diminutive polyps by gastroenterology trainees: a comparison between the NICE and SIMPLE classification.

## **National meetings:**

E-poster presentation and abstract of distinction at the British Society of Gastroenterology annual meeting 2018.

- A randomised controlled study of the prediction of diminutive/small polyp histology using didactic vs. computer-based self-learning module in gastroenterology trainees.

# Table of contents

|          |                                                                                                      |           |
|----------|------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction.....</b>                                                                             | <b>1</b>  |
| 1.1      | Background of this thesis .....                                                                      | 1         |
| 1.2      | Overview of this thesis .....                                                                        | 4         |
| <b>2</b> | <b>SECTION ONE: Optical characterisation of inflammation by advanced endoscopic techniques .....</b> | <b>6</b>  |
| 2.1      | Overview .....                                                                                       | 6         |
| 2.2      | Currently used scores in IBD.....                                                                    | 6         |
| 2.3      | Relationship between clinical symptoms and endoscopic findings.....                                  | 11        |
| 2.4      | Treat-to-target and problems with current definitions of mucosal healing .....                       | 13        |
| 2.5      | Limitations of current scores .....                                                                  | 16        |
| 2.6      | Novel endoscopic technologies .....                                                                  | 20        |
| 2.6.1    | Dye-based chromoendoscopy .....                                                                      | 21        |
| 2.6.2    | Virtual electronic chromoendoscopy .....                                                             | 22        |
| 2.6.3    | Magnifying colonoscopy .....                                                                         | 27        |
| 2.6.4    | Other advanced endoscopic technologies .....                                                         | 27        |
| <b>3</b> | <b>THE REAL-LIFE PROSPECTIVE VALIDATION STUDY OF THE PICASSO SCORE .....</b>                         | <b>30</b> |
| 3.1      | Overview .....                                                                                       | 30        |
| 3.2      | Background.....                                                                                      | 30        |

|       |                                                                                 |    |
|-------|---------------------------------------------------------------------------------|----|
| 3.3   | Histological assessment of inflammatory activity .....                          | 31 |
| 3.3.1 | Definition of histological healing .....                                        | 31 |
| 3.3.2 | Histological inflammation and endoscopic scoring of inflammatory activity ..... | 32 |
| 3.3.3 | Histological inflammation and risk of relapse .....                             | 34 |
| 3.4   | Aims and objectives .....                                                       | 39 |
| 3.5   | Methods .....                                                                   | 41 |
| 3.5.1 | Study design .....                                                              | 41 |
| 3.5.2 | Endoscopic assessment .....                                                     | 42 |
| 3.5.3 | Histological assessment .....                                                   | 45 |
| 0.0   | No increase .....                                                               | 50 |
| 3.5.4 | Interobserver agreement .....                                                   | 51 |
| 3.5.5 | Follow up and outcomes .....                                                    | 52 |
| 3.5.6 | Statistical analysis .....                                                      | 52 |
| 3.5.7 | Ethical approval .....                                                          | 54 |
| 3.6   | Results .....                                                                   | 55 |
| 3.6.1 | Patient demographics .....                                                      | 55 |
| 3.6.2 | Inter-observer agreement of PICaSSO, UCEIS and MES scores .....                 | 58 |
| 3.6.3 | Correlation of endoscopic scores with histological scores .....                 | 60 |
| 3.6.4 | Relationship between histological remission and endoscopic scores .....         | 65 |
| 3.6.5 | Follow up results and prediction of adverse outcomes .....                      | 73 |
| 3.7   | Discussion .....                                                                | 81 |

## **4 SECTION TWO: RAMAN SPECTROSCOPY IN INFLAMMATORY BOWEL DISEASE .....88**

|       |                                                                                        |     |
|-------|----------------------------------------------------------------------------------------|-----|
| 4.1   | Overview .....                                                                         | 88  |
| 4.2   | Background to Raman Spectroscopy and its application in IBD .....                      | 88  |
| 4.2.1 | What is Raman Spectroscopy? .....                                                      | 88  |
| 4.2.2 | Raman Spectroscopy in differentiating UC from non-IBD.....                             | 90  |
| 4.2.3 | Raman Spectroscopy to differentiate CD from UC .....                                   | 91  |
| 4.2.4 | Raman Spectroscopy to assess inflammatory activity in UC .....                         | 93  |
| 4.3   | Assessment of inflammatory activity using Raman Spectroscopy.....                      | 95  |
| 4.3.1 | Aims and objectives.....                                                               | 97  |
| 4.3.2 | Study design .....                                                                     | 97  |
| 4.4   | Results .....                                                                          | 105 |
| 4.4.1 | Raman Spectroscopy in the assessment of inflammatory activity and mucosal healing: 105 |     |
| 4.4.2 | Biochemical changes following response to biological therapy .....                     | 115 |
| 4.5   | Discussion.....                                                                        | 126 |
| 4.6   | Conclusion .....                                                                       | 132 |

## **5 SECTION THREE: TRAINING IN ADVANCED ENDOSCOPY ...133**

|       |                                                                        |     |
|-------|------------------------------------------------------------------------|-----|
| 5.1   | Overview .....                                                         | 133 |
| 5.2   | Small/diminutive colorectal polyps .....                               | 133 |
| 5.2.1 | Optical diagnosis training of small/diminutive colorectal polyps ..... | 133 |

|          |                                                                                                                                                                                         |            |
|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 5.3      | A randomised controlled study of the prediction of diminutive/small colorectal polyp histology using didactic vs. computer based self-learning module in gastroenterology trainees..... | 139        |
| 5.3.1    | Aims.....                                                                                                                                                                               | 139        |
| 5.3.2    | Methods .....                                                                                                                                                                           | 139        |
| 5.3.3    | Results .....                                                                                                                                                                           | 146        |
| 5.3.4    | Discussion.....                                                                                                                                                                         | 151        |
| 5.4      | Training methods in optical diagnosis and characterisation of colorectal polyps: a systematic review and meta-analysis.....                                                             | 155        |
| 5.4.1    | Introduction: .....                                                                                                                                                                     | 155        |
| 5.4.2    | METHODS:.....                                                                                                                                                                           | 156        |
| 5.4.3    | RESULTS:.....                                                                                                                                                                           | 160        |
| 5.4.4    | Discussion:.....                                                                                                                                                                        | 180        |
| 5.5      | Training in advanced endoscopy in IBD .....                                                                                                                                             | 187        |
| 5.5.1    | Background.....                                                                                                                                                                         | 187        |
| 5.5.2    | Aims.....                                                                                                                                                                               | 188        |
| 5.5.3    | Methods .....                                                                                                                                                                           | 188        |
| 5.5.4    | Results .....                                                                                                                                                                           | 192        |
| 5.5.5    | Discussion.....                                                                                                                                                                         | 199        |
| <b>6</b> | <b>CONCLUSIONS AND FUTURE WORK.....</b>                                                                                                                                                 | <b>204</b> |
| 6.1      | Overview .....                                                                                                                                                                          | 204        |

|          |                                                                       |            |
|----------|-----------------------------------------------------------------------|------------|
| 6.2      | The real-life prospective validation study of the PICaSSO score. .... | 204        |
| 6.3      | Raman Spectroscopy .....                                              | 209        |
| 6.4      | Training in advanced endoscopy .....                                  | 212        |
| 6.5      | Conclusion .....                                                      | 217        |
| <b>7</b> | <b>List of References.....</b>                                        | <b>219</b> |

# List of figures

|                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 2-1 The Mayo Endoscopic Score (MES) .....                                                                                                                | 7  |
| Figure 2-2 The Ulcerative Colitis Endoscopic Index of Severity score .....                                                                                      | 8  |
| Figure 2-3 The Simple Endoscopic Score for Crohn's Disease (SES-CD) .....                                                                                       | 10 |
| Figure 3-1 iScan settings used in the real-life prospective validation study of the PICaSSO score.....                                                          | 43 |
| Figure 3-2 The PICaSSO Score.....                                                                                                                               | 44 |
| Figure 3-3 The Robarts Histological Index .....                                                                                                                 | 46 |
| Figure 3-4 Extent, Chronicity, Activity inflammation and Plus others (ECAP) histological index .....                                                            | 47 |
| Figure 3-5 Nancy histological index .....                                                                                                                       | 48 |
| Figure 3-6 Villanacci histological score .....                                                                                                                  | 49 |
| Figure 3-7 Geboes histological index .....                                                                                                                      | 50 |
| Figure 3-8 Heatmap demonstrating Pearson correlation in the rectum between endoscopic and histological scores in all patients. ....                             | 61 |
| Figure 3-9 Heatmap demonstrating Pearson correlation in the rectum between endoscopic and histological scores in patients with moderate/severe activity.....    | 62 |
| Figure 3-10 Heatmap demonstrating Pearson correlation in the sigmoid between endoscopic and histological scores in all patients. ....                           | 63 |
| Figure 3-11 Heatmap demonstrating Pearson correlation in the sigmoid between endoscopic and histological scores in patients with moderate/severe activity. .... | 64 |
| Figure 3-12 Receiver operator curves for the PICaSSO, UCEIS and MES in predicting histological healing in the rectum.....                                       | 68 |

|                                                                                                                                                            |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 3-13 Receiver operator curves for the PICaSSO, UCEIS and MES in predicting histological healing in the sigmoid colon .....                          | 69  |
| Figure 3-14 Relationship between RHI and Nancy with PICaSSO mucosal and vascular architecture in the rectum .....                                          | 70  |
| Figure 3-15 Cox proportional hazard curves in predicting likelihood of specified clinical outcomes at 6 months .....                                       | 75  |
| Figure 3-16 Cox proportional hazard curves in predicting likelihood of specified clinical outcomes at 12 months .....                                      | 76  |
| Figure 3-17 Cox proportional hazard curves in predicting likelihood of specified clinical outcomes at 6 months .....                                       | 77  |
| Figure 3-18 Cox proportional hazard curves of PICaSSO mucosal score and PICaSSO vascular architectural scores predicting specified clinical outcomes ..... | 78  |
| Figure 3-19 Cox proportional hazard curves of PICaSSO mucosal score and PICaSSO vascular architectural scores predicting clinical outcomes.....            | 79  |
| Figure 4-1 Example of endoscopic assessment of IBD. ....                                                                                                   | 99  |
| Figure 4-2 Overview diagram demonstrating the process of sample acquisition and preparation for Raman Spectral measurement and data analysis .....         | 102 |
| Figure 4-3 Self-organising map (SOM) of UC mucosal healing (Blue) vs. active inflammation (Red) .....                                                      | 108 |
| Figure 4-4 Self-organising map discriminant index (SOMDI) of UC MH (blue) vs active inflammation (red).....                                                | 109 |
| Figure 4-5 SOM of CD MH (Blue) vs. active inflammation (red).....                                                                                          | 112 |
| Figure 4-6 SOMDI of CD MH (Blue) vs active inflammation (red) highlighting the spectral differences from the SOM .....                                     | 113 |

|                                                                                                                                                          |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4-7 SOM of the UC pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black) .....                                                   | 118 |
| Figure 4-8 SOMDI highlighting the spectral differences from the SOM of UC pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black) .....  | 119 |
| Figure 4-9 SOM demonstrating the groupings of CD pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black) .....                           | 122 |
| Figure 4-10 SOMDI highlighting the spectral differences from the SOM of CD pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black) ..... | 123 |
| Figure 5-1 NICE (NBI International Colorectal Endoscopic) Classification .....                                                                           | 142 |
| Figure 5-2 SIMPLE (Simplified Identification Method for Polyp Labelling during the Endoscopy) classification .....                                       | 143 |
| Figure 5-3 Example of training material used in the training module illustrating polyp features included in the NICE and SIMPLE classifications .....    | 144 |
| Figure 5-4 PRISMA Flow Chart .....                                                                                                                       | 165 |
| Figure 5-5 Forest plots for all studies assessing the effect of training on accuracy of optical diagnosis of colorectal polyps .....                     | 169 |
| Figure 5-6 Funnel plots for assessing the impact of training on optical diagnosis accuracy of small/diminutive colorectal polyps in all studies .....    | 169 |
| Figure 5-7 Forest plots for studies assessing the effect of computer-based training on accuracy of optical diagnosis of colorectal polyps .....          | 170 |
| Figure 5-8 Forest plots for studies assessing the effect didactic training on accuracy of optical diagnosis of colorectal polyps .....                   | 170 |

|                                                                                                                                                                                                    |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5-9 Funnel plots for assessing the impact of training on optical diagnosis accuracy of small/diminutive colorectal polyps in computer-based and didactic training in all endoscopists ..... | 171 |
| Figure 5-10 Forest plots for studies assessing the impact of training on the accuracy optical diagnosis of colorectal polyps in experienced endoscopists .....                                     | 173 |
| Figure 5-11 Forest plots for studies assessing the impact of training on the accuracy optical diagnosis of colorectal polyps in trainees .....                                                     | 175 |
| Figure 5-12 Forest plots for all studies assessing the effect of training on accuracy of optical diagnosis of small/diminutive colorectal polyps only .....                                        | 177 |
| Figure 5-13 Risk of bias of studies included using Cochrane Collaboration's risk of bias tool .....                                                                                                | 179 |
| Figure 5-14 Recommended roadmap for high quality IBD surveillance .....                                                                                                                            | 191 |

## List of tables

|                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 3-1 Patient characteristics of the real-life prospective validation study of the PICaSSO score.....                                           | 56  |
| Table 3-2 Endoscopic grading of inflammatory activity of patients included in the real-life prospective validation study of the PICaSSO score ..... | 57  |
| Table 3-3 Inter-rater correlation of the endoscopic scores .....                                                                                    | 59  |
| Table 3-5 Sensitivity, specificity and accuracy of endoscopic scores in predicting histological healing .....                                       | 71  |
| Table 3-6 Sensitivity, specificity and accuracy of endoscopic scores in predicting histological healing in the sigmoid colon.....                   | 72  |
| Table 3-7 Multivariate regression analysis of PICaSSO components in predicting adverse outcome at 12 months.....                                    | 80  |
| Table 3-8 Comparison of commonly used endoscopic scoring systems in UC .....                                                                        | 87  |
| Table 4-1 Patient characteristics of the Raman Spectroscopy in the differentiation of MH from active inflammation in IBD study .....                | 106 |
| Table 4-2 Relative intensity of Raman shifts between UC MH vs active inflammation taken from SOMDI.....                                             | 110 |
| Table 4-3 Example confusion matrix of the trained predictive model at differentiating MH from active UC .....                                       | 110 |
| Table 4-4 Relative intensity of Raman shifts between CD healing vs non-healing taken from SOMDI .....                                               | 114 |
| Table 4-5 Example confusion matrix of the trained predictive model at differentiating MH from active CD .....                                       | 114 |

|                                                                                                                                                   |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 4-6 Patient characteristics of the pre- vs. post-biological therapy study .....                                                             | 116 |
| Table 4-7 Relative intensity of Raman shifts between UC pre biologic and post biologic and HC taken from SOMDI. ....                              | 120 |
| Table 4-8 Relative intensity of Raman shifts between CD pre-biologic vs. CD post-biologic vs. healthy control taken from SOMDI. ....              | 124 |
| Table 5-1 Baseline characteristics of participants in the randomisation-controlled study of optical diagnosis training.....                       | 146 |
| Table 5-2 Pre-training vs. Post-training optical diagnosis performance using the NICE and SIMPLE classification.....                              | 147 |
| Table 5-3 Comparison of the performance at predicting small/diminutive polyp histology between didactic and computer-based training .....         | 148 |
| Table 5-4 Performances using NICE and SIMPLE classification in the didactic training and computer-based training groups .....                     | 149 |
| Table 5-5 Inter-observer agreement comparison between SIMPLE and NICE pre- vs. post-training.....                                                 | 150 |
| Table 5-6 PRISM checklist.....                                                                                                                    | 162 |
| Table 5-7 MOOSE Checklist.....                                                                                                                    | 164 |
| Table 5-8 Characteristics of studies included in the systematic review and meta-analysis ...                                                      | 166 |
| Table 5-9 Patient demographics of the IBD surveillance study .....                                                                                | 193 |
| Table 5-10 Summary of performance in IBD quality indicators pre- and post-August 2016 in all endoscopists and the 4 DCE trained endoscopists..... | 196 |
| Table 5-11 A summary of the reasons DCE could not be used for procedures between January 2014 and October 2018.....                               | 197 |
| Table 5-12 Breakdown of detected lesions.....                                                                                                     | 197 |

Table 5-13 Performance of DCE trained endoscopists vs. all other endoscopists post-August  
2016 ..... 198

# 1 Introduction

## 1.1 Background of this thesis

Inflammatory bowel disease (IBD) describes a group of chronic idiopathic disorders that cause inflammation within the gastrointestinal tract, namely Crohn's disease (CD) and Ulcerative Colitis (UC). CD is a condition that describes inflammation that can involve any part of the gastrointestinal tract, which typically occurs with skip-lesions whereby there is normal mucosa between areas of inflammation. It is a trans-mural disease and therefore complications such as abscesses and fistulae can occur. UC is a condition describing inflammation within the large intestine. It is a mucosal disease and therefore complications that are seen in CD are not typically seen. Both diseases commonly follow a relapsing and remitting course.

The most common presentation is one of a change in bowel habit, abdominal pain, rectal bleeding, passing mucus per rectum and weight loss, however the symptoms will depend on the distribution of the disease. Once seen in mainly western countries where the incidence may be stabilising, it is increasingly diagnosed in developing countries and is now considered a global disease, with a prevalence of 0.3% (Ng, Shi et al. 2018).

IBD represents a debilitating disease that leads to significant morbidity in patients of a working age, which in turn has socioeconomic costs. An Australian population-based cohort study aiming to establish the severity and complications of IBD was published by *Niewiadomski et al (Niewiadomski, Studd et al. 2015)*. The authors used coding to search for patients between 2007-2013, and categorised patients as having disability if there was a

requirement for >2 steroid courses, hospitalisation or surgical intervention (Beaugerie, Seksik et al. 2006). They included 96 patients with UC with 25% defined as having disabling disease (Niewiadomski, Studd et al. 2015). A population-based cohort study in Norway, examined the rate of disability amongst patients with IBD up to 10 years after diagnosis (Hoivik, Moum et al. 2013). Through interviews, examinations, colonoscopy and blood tests at 1 and 10 year follow up periods (Hoivik, Moum et al. 2013) the relative risk of requiring disability pension was higher in patients with UC compared with normal population RR1.80 (95% CI 1.41-2.27), with the risk higher amongst younger patients (30-39 years of age RR 2.78 95% CI 1.44-5.38) (Hoivik, Moum et al. 2013).

These studies confirm the level of morbidity associated with a diagnosis of IBD. This is particularly the case in active disease, not least in terms of symptoms but also in terms of fatigue (OR 4.2 95% CI 2.2-7.8), poor sleep (OR 4.0 95% CI 1.9-8.6) and stress (OR 4.2 95% CI 2.2-8.1) (Graff, Vincent et al. 2011). Complications of poorly controlled disease include steroid use, hospitalisation, increased risk of colorectal cancer (CRC) and surgical intervention in the form of colectomy. The risk of patients being admitted to hospital due to UC over a period of 10 years is up to 66% (Fumery, Singh et al. 2018). Prior to immunosuppressant treatment, the risk of a UC patient of undergoing colectomy over 25 years was as high as 45% (Fumery, Singh et al. 2018), which has reduced to a 14.6% risk after 10 years of follow up (Fumery, Singh et al. 2018), which is likely due to an increase in therapies that are able to modify the disease course.

When relapse occur, patients are at risk of requiring corticosteroids, with associated risks/side effects, thromboembolic events, increased likelihood of hospital admission and surgical intervention. Accurate assessment of inflammatory activity through colonoscopy is vital in the

optimum management of patients with IBD. Advanced endoscopic technologies are increasingly allowing for accurate predictions of histological inflammatory activity and neoplastic activity (Iacucci, Furfaro et al. 2018, Iacucci, McQuaid et al. 2019) which will in turn will allow better decision-making when deciding on escalating/de-escalating therapy and in the management of IBD lesions.

Patients with IBD are at a higher risk of developing CRC compared with the general population (Ekbom, Helmick et al. 1990, Ekbom, Helmick et al. 1990, Dyson and Rutter 2012, Lutgens, van Oijen et al. 2013). Colonoscopy plays a vital role in the management of patients with IBD, not least in making a formal diagnosis but also in monitoring response to treatments and long-term surveillance for neoplastic lesions. Patients with long-standing IBD are invited to surveillance programmes whereby colonoscopy is offered every 1-5 years depending on risk factors and findings (Magro, Gionchetti et al. 2017). The aim of this is to identify lesions (polyps) that have the potential to become malignant. These polyps/lesions if detected early, may be able to be removed endoscopically, reducing the risk of cancer and prevent the need for major surgical intervention such as colectomy.

Colonoscopy also plays a vital role in the non-IBD population as it is the gold standard for screening for bowel cancers and detection of pre-cursors to colorectal cancer (polyps). Much like IBD lesions, early detection of polyps, allows endoscopic removal and therefore reduction in CRC. Following improvements in technology endoscopists are detecting more lesions within the bowel with the majority small/diminutive <5mm (80%), however the clinical relevance of these lesions is minimal as the risk of advanced histology or cancer is <1% (Lieberman, Moravec et al. 2008, Rex, Overhiser et al. 2009, Hewett, Kaltenbach et al. 2012). Novel imaging techniques can help endoscopists characterise these small lesions

between being neoplastic and non-neoplastic (hyperplastic), termed Optical Diagnosis (Chiu, Chang et al. 2007, Sakamoto, Matsuda et al. 2012, McGill, Evangelou et al. 2013, Wu, Li et al. 2013). In the hands of experts using advanced endoscopic technologies, the histology of polyps can be predicted with a high degree of accuracy (Rex 2009). If small polyps can be predicted accurately, it is proposed that they could be removed and discarded without sending to pathology for formal histological analysis, the “Resect and Discard” strategy (Rex, Kahi et al. 2011). In addition, small polyps that are deemed to be hyperplastic in the rectum and sigmoid colon, and therefore do not have malignant potential, can be left in situ. This would have significant cost benefits for healthcare services and reduce the risk of harm to patients due to fewer unnecessary polypectomies. However, the performance of experts has not been replicated in routine clinical practice (Rees, Rajasekhar et al. 2017). For this reason, training will be essential to up-skill endoscopists in optical diagnosis. Endoscopy training has traditionally relied upon ad hoc training intra-procedure; however, this is reliant upon many variables, not least on the type of pathology found. Alternative methods of training might include traditional face-to-face lecture style teaching or more modern computer-based learning; yet the optimum method remains to be seen.

## 1.2 Overview of this thesis

This thesis has three themes: the assessment of inflammatory activity and mucosal healing in ulcerative colitis using advanced endoscopic technology, the application of Raman Spectroscopy in IBD and finally training in advanced endoscopic techniques. Section one presents the literature on endoscopic and histological assessment of UC inflammatory activity and the concept of mucosal healing. I then present the PICaSSO study, which is a novel virtual electronic chromoendoscopy (VEC) score which, I aim to validate prospectively in a

multicentre study and assess its ability to predict mucosal healing and its correlation with histological indices of inflammatory activity.

In section two I discuss Raman Spectroscopy and present the literature of its application in IBD. I then examine Raman Spectroscopy in IBD and aim to assess its ability to predict mucosal healing using a novel machine learning technique, assessing the changes in the biomolecular make-up of colonic tissue following successful treatment with biological therapy.

Finally, in section three I discuss current training methods in advanced endoscopic techniques and present the findings of a randomised controlled study comparing didactic training and computer-based self-training in the optical diagnosis of small/diminutive colorectal polyps using advanced technologies. I present a systematic review and meta-analysis of training methods in optical diagnosis of colorectal polyps. I also present the findings of a quality improvement project in quality measures in IBD surveillance procedures following service changes and training in advanced endoscopic techniques. The objective was to assess the improvement in key performance indicators following these training and service delivery interventions.

## 2 SECTION ONE: Optical characterisation of inflammation by advanced endoscopic techniques

### 2.1 Overview

In recent years there has been rapid developments in endoscopic technology, allowing enhanced imaging and more precise assessments of disease activity in IBD. In this chapter I give an overview of the existing literature on endoscopic assessments of inflammatory activity including their limitations. I will present the current management paradigm in IBD and how novel endoscopic technologies can help achieve key goals in the assessment of inflammatory activity. The novel endoscopic technologies I will discuss in detail will be those that are commonly available to endoscopists, which is Virtual Electronic Chromoendoscopy (VEC), and remains the focus of this thesis.

### 2.2 Currently used scores in IBD

Historically patients were monitored and treatments adjusted using patient reported scores and non-specific biochemical markers such as CRP (Truelove and Witts 1959). As endoscopic equipment and technology evolved and became more accessible, it has played a vital role in the formal assessment of inflammatory activity. Endoscopic scores were developed to assist endoscopists in grading severity of inflammatory activity, such as the Mayo endoscopic score (MES) (Schroeder, Tremaine et al. 1987) (Figure 2-1).

| Score | Disease activity | Endoscopic features                                            |
|-------|------------------|----------------------------------------------------------------|
| 0     | Normal/inactive  | None                                                           |
| 1     | Mild             | Erythema, decreased vascular pattern, mild friability          |
| 2     | Moderate         | Marked erythema, absent vascular pattern, friability, erosions |
| 3     | Severe           | Spontaneous bleeding, ulceration                               |

**Figure 2-1 The Mayo Endoscopic Score (MES)**

Endoscopic features of activity such as presence of bleeding, ulceration, erythema vascular pattern and erosions are observed and graded as normal, mild, moderate and severe. The MES is commonly used in everyday clinical practice and is often used in clinical trials. Whilst it is simple to use, there have been many criticisms of the score, which will be discussed later in this chapter. One of the main limitations is that it was created in an era before high-definition white light endoscopy, and therefore does not include endoscopic features that are increasingly visible. An example is the descriptor “absent vascular pattern”, however even in severe disease modern endoscopic technology can visualise vessel patterns (Iacucci, Fort Gasia et al. 2015).

There have been subsequent attempts at defining endoscopic activity in UC, including the Ulcerative Colitis Endoscopic Index of Severity (UCEIS) (Travis, Schnell et al. 2012), which was developed with more modern endoscopy equipment (Figure 2-2).

| <b>Descriptor</b>  | <b>Scale Anchor points</b>     | <b>Definition</b>                                                                                             |
|--------------------|--------------------------------|---------------------------------------------------------------------------------------------------------------|
| Vascular pattern   | Normal (0)                     | Normal vascular pattern                                                                                       |
|                    | Patchy obliteration (1)        | Patchy obliteration of vascular pattern                                                                       |
|                    | Obliterated (2)                | Complete obliteration of vascular pattern                                                                     |
| Bleeding           | None (0)                       | No visible blood                                                                                              |
|                    | Mucosal (1)                    | Spots or streaks of coagulated blood on the mucosal surface which can be washed away                          |
|                    | Luminal mild (1)               | Some free liquid blood in the lumen                                                                           |
|                    | Luminal moderate or severe (3) | Visible oozing of blood from haemorrhagic mucosa                                                              |
| Erosions or ulcers | None (0)                       | Normal mucosa, no visible erosions or ulcers                                                                  |
|                    | Erosions (1)                   | Tiny ( $\leq 5$ mm) defects in the mucosa, of a white or yellow colour with a flat edge                       |
|                    | Superficial ulcer (2)          | Larger ( $\geq 5$ mm) defects in the mucosa, which are discrete fibrin covered ulcers, but remain superficial |
|                    | Deep ulcer (3)                 | Deeper excavated ulcers in the mucosa, with a slightly raised edge                                            |

**Figure 2-2 The Ulcerative Colitis Endoscopic Index of Severity score**

The UCEIS represents a relatively simple scoring system that is being increasingly incorporated in routine clinical practice and clinical trials (Iacucci, Furfaro et al. 2018). It defines endoscopic inflammatory activity in three categories: vascular pattern, bleeding and erosions/ulceration. Within these three categories there are sub-definitions according to severity of disease activity, the total of which, ranging from 0-8, gives the UCEIS score. However, much like the MES, there are notable limitations, which will also be discussed later in this chapter. In the field of endoscopic imaging in UC, these scores do not reflect the subtle features that are now visible with the latest generation of endoscope, an example is a regenerative pattern suggestive of healing (Iacucci, Daperno et al. 2017, Iacucci, Furfaro et al. 2018). Therefore, there is a demand for a new score that reflects the quality of imaging that is possible in the current era, which will be discussed later in this chapter.

The most commonly used endoscopic score in CD is the Simple Endoscopic Score for Crohn's Disease (SES-CD) (Daperno, D'Haens et al. 2004). This score consists of four variables for each segment of colon that is examined endoscopically: ulcers, proportion of surface that is ulcerated, proportion of surface that is affected and presence or absence of narrowing (Daperno, D'Haens et al. 2004) (Figure 2-3).

|                      | SES-CD values         |                                         |                                        |                                      |
|----------------------|-----------------------|-----------------------------------------|----------------------------------------|--------------------------------------|
| Variable             | 0                     | 1                                       | 2                                      | 3                                    |
| Ulcers               | None                  | Aphthous ulcers<br>(Diameter<br><0.5cm) | Large ulcers<br>(Diameter 0.5-<br>2cm) | Very large ulcers<br>(Diameter >2cm) |
| Ulcerated<br>surface | None                  | <10%                                    | 10-30%                                 | >30%                                 |
| Affected<br>surface  | Unaffected<br>segment | <50%                                    | 50-75%                                 | >75%                                 |
| Narrowing            | None                  | Single, can be<br>passed                | Multiple, can be<br>passed             | Cannot be passed                     |

**Figure 2-3 The Simple Endoscopic Score for Crohn's Disease (SES-CD)**

The SES-CD score was developed to simplify the assessment of inflammatory activity in CD and has been shown to respond more favourably to treatment than other scores in CD (Khanna, Zou et al. 2017). However, it has limitations such as relying on subjective measurements of size of ulcers, proportion of surface involved and lacks a true definition of healing. In comparison to UC scores, the calculation is complex, with four variables assessed in each segment of the colon explored. It is also not clear how reliable the score is in predicting histological correlation (Iacucci, Furfaro et al. 2018). The score was also developed in an era prior to high-definition endoscopy and novel technology such as virtual electronic chromoendoscopy (VEC).

## 2.3 Relationship between clinical symptoms and endoscopic findings

There is a disconnect between clinical symptoms in IBD and endoscopic findings, with symptom resolution not necessarily equating with healing at endoscopy (Colombel, Keir et al. 2017). *Kim et al*, retrospectively reviewed records of UC patients who had colonoscopic examination whilst in clinical remission, defined as partial mayo 0-1 (D'Haens, Sandborn et al. 2007) and endoscopic remission defined as MES 0 (Kim, Cheon et al. 2016). They included 215 patients in clinical remission, and only 52.6% had endoscopic remission and 27.4% of all patients had a relapse within a median follow up of 80 months (Kim, Cheon et al. 2016). Patients with endoscopic remission had a lower risk of relapse (19.5% vs. 36.3%  $p=0.006$ ) (Kim, Cheon et al. 2016). This retrospective, single centre study showed how symptoms do not necessarily reflect mucosal inflammation. Although the study was limited in its retrospective design and the use of the MES.

*Jharap et al* evaluated the relationship between clinical symptoms of UC, namely rectal bleeding and stool frequency, with endoscopic findings. The authors included patients who were recruited as part of clinical trials, and of those who achieved endoscopic remission following treatment, only 29% reported normal stool frequency further highlighting the disconnect between symptoms and endoscopic findings (Jharap, Sandborn et al. 2015). *Falvey et al* prospectively recruited patients with IBD whom were attending a colonoscopy, and recorded their simple clinical colitis activity index (SCCAI) (Walmsley, Ayres et al. 1998), and endoscopic remission was defined as a Baron score  $<1$  (Baron, Connell et al. 1964), and included 65 UC patients (Falvey, Hoskin et al. 2015). They found that endoscopic severity was not distinguishable by the SSCAI with an accuracy of 63% for endoscopic remission

(Falvey, Hoskin et al. 2015). Whilst this study also confirms the disconnect it is limited by the modest patient numbers and an out-dated endoscopic score of inflammatory activity.

A retrospective chart review of 154 patients with UC, from a single centre was completed by *Brahmania et al*, who underwent colonoscopic examination within one week of a clinic visit (Brahmania and Bernstein 2014). In this study a scoring system based around the UCEIS (Travis, Schnell et al. 2012) which categorised patients into: Remission, mild, moderate and severe disease with remission being described as complete MH that could include pseudopolyps or scars (Brahmania and Bernstein 2014). Clinicians also categorised patients into remission, mild and active groups based on a physician global assessment (PGA) (Brahmania and Bernstein 2014). They found the PGA predicted moderate-severe endoscopic disease with a sensitivity, specificity, PPV and NPV of 30%, 97%, 83% and 74% respectively (Brahmania and Bernstein 2014). Whilst a single centre, retrospective study, and using an invalidated endoscopic scoring system, this shows that physicians can predict active inflammation in UC with high specificity but lacks the sensitivity and NPV to be relied upon in clinical practice.

The mismatch between symptoms and endoscopic findings was further highlighted by *Baars et al*. The authors recruited IBD patients from 2 centres in the Netherlands undergoing surveillance colonoscopy who were in clinical remission-defined as a lack of abdominal pain, diarrhoea, rectal bleeding, fever and weight loss (Baars, Nuij et al. 2012). A remarkable 69% of UC patients had evidence of endoscopic activity, histological activity or both present (Baars, Nuij et al. 2012). This study further confirms that symptoms alone cannot be relied upon, however the results need to be viewed with caution as the follow up results suggest no

association between active UC at baseline and an increase likelihood of relapse during follow up ( $p=0.18$ ), which is contrary to other studies (Baars, Nuij et al. 2012).

## 2.4 Treat-to-target and problems with current definitions of mucosal healing

For many years the focus of treatment in IBD was in symptom resolution/improvement with the aim to achieve clinical remission. This was largely in the pre-immunosuppressant/biologic era whereby medical therapy did not modify the course of the disease and often failed to achieve healing (Olaison, Sjodahl et al. 1990). There are multiple immunomodulatory/biologic therapies now available, and clinicians have a realistic ability to modify disease course and it is increasingly recognised that mucosal healing is possible (D'Haens, Van Deventer et al. 1999, Rutgeerts, Sandborn et al. 2005). Recent ECCO (European Crohn's and Colitis Organisation) guidelines suggest remission should be defined as:

“a combination of clinical parameters (stool frequency  $\leq 3$ /day with no bleeding) and no mucosal lesions at endoscopy” (Magro, Gionchetti et al. 2017)

By aiming for symptom resolution and endoscopic remission (Peyrin-Biroulet, Sandborn et al. 2015), there can be a positive impact on disease course and frequency of adverse outcomes (Colombel, D'Haens et al. 2019). This definition whilst utilising both endoscopic and clinical aspects, the absence of mucosal lesions is vague and stops short of defining endoscopic features that are positively associated with mucosal healing and does not consider histological features. *Rutter et al* demonstrated the impact of active inflammation can have on the risk of

colorectal malignancy, whereby a one unit increase in histological inflammation increased the likelihood of neoplasia by OR 4.69 (95% CI 2.10-10.48  $p < 0.001$ ) (Rutter, Saunders et al. 2004). This was supported by a recent meta-analysis also highlighting the role of inflammation (endoscopic and histologic) in the risk of colorectal cancer with OR 3.53 (95% CI 2.59-4.80) (Flores, O'Connor et al. 2017).

The latest treatment paradigm in IBD is the “Treat to Target” strategy, whereby treatments are adjusted to achieve the goal of mucosal healing (Peyrin-Biroulet, Sandborn et al. 2015, Turner, Ricciuto et al. 2021). Mucosal healing is associated with improved outcomes for patients, such as reduced hospital admissions, lower colectomy rates and fewer steroid courses (Froslic, Jahnsen et al. 2007, Colombel, Rutgeerts et al. 2011, Annese, Daperno et al. 2013). A retrospective study of patients included in a study investigating Golimumab in UC, included 91 patients with moderate-severe disease (Mayo  $\geq 2$ ) prior to induction and maintenance therapy (Bossuyt, Baert et al. 2019). Following a multivariate analysis, mucosal healing (Mayo 0/1) at week 14 predicted steroid free remission at week 52 (OR 9.38 95% CI 2.68-32.84  $p < 0.001$ ) (Bossuyt, Baert et al. 2019). A meta-analysis by *Shah et al* confirmed the importance of MH in UC by finding it was associated with long-term clinical remission ( $> 52$  weeks) (OR 4.5 95% CI 2.12-9.52), avoidance of colectomy (OR 4.15 95% CI 2.53-6.81), steroid-free remission (OR 9.70 95% CI 0.94-99.67) and long term MH (OR 8.40 95% CI 3.13-22.53) (Shah, Colombel et al. 2016).

*Meucci et al* completed a prospective multicentre study, whereby UC patients with active disease were recruited to receive treatment doses of 5-ASA, and those patients who achieved clinical remission (normal frequency of bowel movements, no abdominal pain and no rectal

bleeding) were reviewed with a second colonoscopy after 6 weeks (Meucci, Fasoli et al. 2012). Thereafter patients were followed for 12 months (n=61), 91.5% of which achieved endoscopic remission after 6 weeks of treatment with 5-ASA (Mayo 0/1) (Meucci, Fasoli et al. 2012). The clinical relapse rate of patients with both clinical and endoscopic remission was 23% at 12 months, whereas patients with only clinical remission the rate was 80% ( $p<0.0001$ ) and the only variable found to be associated with lower risk of clinical relapse was endoscopic remission ( $p=0.001$ ) (Meucci, Fasoli et al. 2012). This study further supports the strategy of achieving endoscopic mucosal healing, although it is limited by the use of the MES, modest patient numbers and the clinical measures of relapse as opposed to endoscopic measures.

*Ferrante et al* reported the long-term follow up data from UC patients receiving infliximab as part of the ACT-1 trial (Ferrante, Vermeire et al. 2008). Overall 17% of the 121 patients that were originally included in the trial underwent colectomy after a median follow up of 33.0 months (17.0-49.8) (Ferrante, Vermeire et al. 2008). They found that the absence of mucosal healing (Mayo 2-3) in the short-term (4 or 10 weeks after first infusion) was associated with an increased risk of colectomy (LogRank  $p=0.027$ ) (Ferrante, Vermeire et al. 2008). A further study of long-term outcomes following infliximab therapy in UC was conducted by *Laharie et al*, whereby they retrospectively compared outcomes of patients achieving mucosal healing defined as MED 0/1 (Laharie, Filippi et al. 2013). A total of 63 patients were included, 48% of which were judged to have achieved mucosal healing following infliximab therapy (Laharie, Filippi et al. 2013). Only 3% of mucosal healing patients underwent colectomy during follow up (median 27 (3-79) months), vs. 39% who had not achieved mucosal healing ( $p=0.004$ ) (Laharie, Filippi et al. 2013). In a multivariate analysis, the only feature predictive

of the requirement for colectomy was a MES of 2/3 vs. 0/1 (OR 18.01 95% CI 1.58-204.92 p0.02) (Laharie, Filippi et al. 2013). The authors also found a higher rate of infliximab failure, defined as drug withdrawal due to lack of response, with active inflammation vs. mucosal healing (76% vs. 37% p0.003) OR 3.23 95% CI 1.48-7.03 p0.003 (Laharie, Filippi et al. 2013). This further adds to the evidence of treating to a target of mucosal healing, albeit this study is limited by patient numbers and its retrospective data collection.

The relevance of achieving mucosal healing and the favourable long-term outcomes was shown by *Colombel et al* who reported the outcomes from the ACT-1 and ACT-2 studies (Acute Ulcerative Colitis Trials 1 and 2) (Colombel, Rutgeerts et al. 2011). In these clinical trials, 728 patients with UC received infliximab or placebo and there were favourable outcomes for those patients receiving infliximab. Patients achieving mucosal healing (MES 0/1) had a lower incidence of colectomy, and the Kaplan-Meier curve demonstrates deviation between MH vs. non-MH from week 8 (Colombel, Rutgeerts et al. 2011). Those patients achieving MH at week 8 were more likely to be in clinical remission (73% at week 54 for MES 0, 47% for MES 1, 24% for MES 2 and 10% for MES 3 p<0.0001) as well as being more likely to be steroid-free (Colombel, Rutgeerts et al. 2011). This was further confirmed by higher rates of endoscopic mucosal healing at weeks 30 and 54 (77% for MES 0, 54% for MES 1, 21% for MES 2 and 6.7% for MES 3 p<0.0001), which further confirms the importance of achieving mucosal healing (Colombel, Rutgeerts et al. 2011).

## 2.5 Limitations of current scores

MH is poorly defined, with resolution of friability and/or ulceration commonly used. The definition of MH continues to evolve, however existing scores have limitations and fail to

include features that define MH; rather it is assumed the absence of inflammation is the same as MH. The MES is the most widely used score in UC, and a score of 0 or 1 is often considered as MH in clinical trials. It is not surprising to see a difference in relapse rates between MES 0 and MES 1, with fewer relapses associated with 0 than 1 since MES 1 allows for some endoscopic activity of colitis, with erythema, mild friability and decreased vascular pattern present (Schroeder, Tremaine et al. 1987).

*Boal Carvalho et al* confirmed this by retrospectively reviewing consecutive UC patients in remission (endoscopic score of 0/1 and partial mayo 0-1) and collected data for 12 months post-colonoscopy (Boal Carvalho, Dias de Castro et al. 2016). A total of 138 patients were reviewed, 20.3% of which suffered clinical relapse within 12 months, and through multivariate analysis MES 1 was the only factor associated with relapse (OR 2.89 95% CI 1.137-7.385  $p=0.026$ ) (Boal Carvalho, Dias de Castro et al. 2016). However, this study is limited by its single centre, retrospective design. A prospective study by *Barreiro-de Acosta et al* also further confirmed the relapse rate with Mayo 1 is higher than Mayo 0 (Barreiro-de Acosta, Vallejo et al. 2016). 187 UC patients who were in clinical remission, for at least 3 months (Partial Mayo  $\leq 2$ ), underwent colonoscopy and then observed for 12 months for relapse (Barreiro-de Acosta, Vallejo et al. 2016). Despite all patients having MES 0/1, 19.2% had a clinical relapse, however more were from the MES 1 group (36.1%) than MES 0 (11.1%  $p<0.01$ ) (Barreiro-de Acosta, Vallejo et al. 2016). Similar to *Boal Carvalho et al*, the only factor associated with relapse in multivariate analysis was MES 1 (OR 6.27 95% CI 2.73-14.40  $p<0.001$ ) (Barreiro-de Acosta, Vallejo et al. 2016).

The limitations of the MES were further demonstrated by *Naganuma et al* whereby they followed UC patients who achieved clinical and endoscopic remission following topical steroid therapy (Naganuma, Hirai et al. 2019). Despite achieving clinical remission and having MES 0 disease, up to 70% of patients suffered a relapse, defined clinically, within 2 years (Naganuma, Hirai et al. 2019). Although a retrospective study and including patients with pan-colitis, whereby topical therapy is unlikely to induce meaningful remission, this study adds to the evidence that the MES has limitations in adequately defining MH. *Lemmens et al* aimed to explore the correlation between endoscopic scoring (MES) and histological scoring (Riley score (Riley, Mani et al. 1991) and Geboes (Geboes, Riddell et al. 2000)) in terms of inflammatory activity in UC (Lemmens, Arijis et al. 2013). The authors found that the endoscopic score that correlated the best with histology was MES 0 with 15.7% misclassified, but there were frequent misclassifications at MES 1 disease (79.1%) (Lemmens, Arijis et al. 2013). This suggests that when UC has clear evidence of endoscopic quiescence (MES 0) the likelihood is that the histology will corroborate, however where the issue lies in the mild forms of activity. This likely reflects that the definitions in the MES do not include enough bandwidth for more subtle findings that may reflect healing as opposed to active inflammation.

The UCEIS was developed to better define inflammatory activity in UC. A score of  $\leq 1$  was used by STRIDE as healing, however much like MES there is a difference in relapse rates between a score of 0 and 1. To define healing using UCEIS, it is assumed the absence of variables such as bleeding, erosions etc equates with healing. The limitations of MES have been widely reported, in particular at the remission end of the spectrum with MES 0 patients often having an unacceptably high relapse rate. *Arai et al* aimed to establish if the UCEIS

could predict prognosis in UC patients who had achieved clinical remission (Arai, Naganuma et al. 2016). They retrospectively reviewed the outcomes of patients who had a colonoscopy completed by a single endoscopist, and they found the UCEIS was strongly correlated with MES ( $r=0.937$ ), but a moderate correlation with clinical scores of severity, partial Mayo score ( $r=0.637$ ), and patients in clinical remission had a range of UCEIS scores (0-5) (Arai, Naganuma et al. 2016). They found that a UCEIS score of  $\leq 1$  had a sensitivity of 68%, specificity 57% and AUC of 0.69 for predicting clinical remission (partial Mayo score  $\leq 2$ ) (Arai, Naganuma et al. 2016). It seemed that the bleeding component was a predictor of clinical relapse, since patients with a higher bleeding score such as mucosal bleeding had a remission duration of 19.9 months (95% CI 14.3-36.9) whereas patients with no bleeding present had a remission duration of 34.8 months (95% CI 32.7-36.9) (Arai, Naganuma et al. 2016). Through multivariate analysis, the absence of bleeding ( $p<0.001$ ) and absence of mucosal damage ( $p<0.001$ ) were predictors for a longer duration in clinical remission (Arai, Naganuma et al. 2016). This study was however limited by a single endoscopist in a single centre, in addition to clinical data being collected retrospectively.

*Irani et al* examined the correlation between UCEIS and histological activity in UC in a large single centre study (Irani, Wang et al. 2018). Biopsies were taken at the site of most inflammation, together with a UCEIS score (Travis, Schnell et al. 2012) and were correlated with the Nancy index (Marchal-Bressenot, Salleron et al. 2017) and RHI (Mosli, Feagan et al. 2017), from 125 patients. The correlation coefficient was very good between UCEIS and RHI/Nancy at  $r=0.86$  (95% CI 0.80-0.90  $p<0.001$ )/ $0.84$  (95% CI 0.76-0.89  $p<0.001$ ) respectively (Irani, Wang et al. 2018). When considering histological remission, a Nancy score of  $\leq 1$  and RHI  $\leq 3$  was used, a UCEIS score of 0 was most consistent with remission

(OR 11.1 95% CI 1.3-95.6  $p=0.0143$ ) (Irani, Wang et al. 2018). This is the first study to correlate an endoscopic score and validated histological scores, however the limitation of being a single centre, single endoscopist, who developed the UCEIS score, and single pathologist does limit the results. Although a UCEIS score of 0 may correlate with histological remission, it does not define endoscopic features of healing, rather reflecting the absence of inflammatory activity.

The UCEIS is a useful tool in clinical practice and is starting to be adopted routinely in the central readout for clinical trials. However, it has limitations, not least it does not define MH but also does not define the thresholds for mild to moderate and severe disease. Its advantages over MES remains to be seen (Levesque, Sandborn et al. 2015). Since endoscopic MH is considered a key target of therapy in UC, endoscopic scoring of inflammatory activity is vital. A systematic review by *Samaan et al* included the endoscopic scores available and commented that the MES (Schroeder, Tremaine et al. 1987) and UCEIS (Travis, Schnell et al. 2012) show the most promise with regards to evaluating inflammatory activity but were not validated (Samaan, Mosli et al. 2014). But these scores are limited by lacking a clear definition of MH.

## 2.6 Novel endoscopic technologies

Historically, the gold standard of evaluating inflammatory activity and remission is histological assessment of tissue biopsies taken at the time of the endoscopy. This allows formal assessment of inflammatory activity; however, it relies upon accurate colonic sampling of an area that represents the overall inflammatory status. If subtle endoscopic features of inflammation are present, but biopsies are taken of a quiescent area, then clinicians and patients may be falsely reassured. Biopsies also only provide a snapshot of inflammatory

activity and are unable to provide a dynamic perspective of inflammation. Since the development of MES and UCEIS there has been advancements in endoscopic technology, which enables endoscopists to identify subtle findings that can approximate histology (Iacucci, Fort Gasia et al. 2015). One example of this, is that even in severe disease whereby a loss in vascular pattern is noted in the MES (Schroeder, Tremaine et al. 1987), newer high definition endoscopes are able to give images in greater detail and resolution, up to 2 million pixels, whereas standard definition endoscopes offer a resolution up to 400000 pixels (Iacucci, Furfaro et al. 2018). Therefore, it is possible to visualise a vascular pattern even in the most severe inflammatory activity. High-definition endoscopes are able to enhance the ability to define subtle mucosal and vascular characteristics. There have been attempts at exploiting these innovations in endoscopic techniques to bridge the gap between endoscopic and histological assessment, to allow accurate predictions of histology optically, so called “optical diagnosis”.

### **2.6.1 Dye-based chromoendoscopy**

Dye-based chromoendoscopy (DCE) refers to the technique of applying dye, namely methylene blue or indigo carmine, to the bowel mucosa, allowing detailed assessment of mucosal characteristics. DCE received early interest as a means of characterising colorectal neoplasia (Rembacken, Fujii et al. 2000), and the application of dye allows the pit pattern to be visualised and used to predict neoplasia (Kudo, Rubio et al. 2001). Since DCE enables accurate predictions of non-IBD neoplasia its application in IBD was first examined in a pioneering study by *Kiesslich et al* (Kiesslich, Fritsch et al. 2003). The authors recruited UC patients with inactive and long-standing disease prospectively and randomised to either conventional white light colonoscopy (standard white light endoscopy) or chromoendoscopy,

in this case methylene blue, and a higher detection rate of neoplasia was detected in the chromoendoscopy group ( $p=0.003$ ) and a higher rate of, more difficult to detect, subtle neoplasia ( $p=0.0007$ ) (Kiesslich, Fritsch et al. 2003). As a secondary objective, the authors examined the ability of chromoendoscopy to predict inflammatory activity, compared with histological assessment, graded as no inflammation, mild-moderate or severe. Interestingly there was no statistically significant difference in agreement between endoscopic prediction and histological confirmation between the conventional group and chromoendoscopy group (Kiesslich, Fritsch et al. 2003). However, when mild inflammation was present there was a higher agreement in the chromoendoscopy group (87% vs. 54%  $p=0.0002$ ) (Kiesslich, Fritsch et al. 2003), which raises the possibility of its usefulness in predicting mucosal healing. One of the main limitations of DCE is that it is a time-consuming technique, requires training and expertise, and has limited use when bowel preparation is sub-optimal (Kaltenbach, Soetikno et al. 2017).

### **2.6.2 Virtual electronic chromoendoscopy**

DCE using methylene blue or indigo carmine is effective particularly in the setting of delineating neoplastic lesions in IBD (Rutter, Saunders et al. 2004). However, the cost, time and expertise required limits its application in everyday clinical practice. For this reason, there has been increasing attention to novel endoscopic technologies such as Virtual Electronic Chromoendoscopy (VEC), which includes Narrow Band Imaging (NBI), iScan and Blue-light laser imaging (BLI). These refer to techniques that enhance the mucosal and vascular pattern, at the push of a button on the endoscope, in a similar way to dye-based chromoendoscopy, but without the need to apply dye to bowel mucosa.

### 2.6.2.1 Narrow Band Imaging (NBI)

Narrow Band Imaging (NBI) (Olympus, Tokyo, Japan) was first reported by *Gono et al* and involves the narrowing of bandwidths of light, using a light filter, to 400-700nm (Gono, Yamazaki et al. 2003) . The light at this part of the spectrum is absorbed by haemoglobin, therefore making blood vessels and blood itself darker and more pronounced. During the inflammatory process blood vessels become dilated and when moderate and severe inflammatory activity is present bleeding can occur within the mucosa or lumen, which is reflected by the MES and UCEIS (Schroeder, Tremaine et al. 1987, Travis, Schnell et al. 2012). NBI can allows these changes to be more visible. Its use was investigated by *Kudo et al* who examined the mucosal vascular pattern using NBI and commented as to whether it was normal or considered irregular or tortuous (Kudo, Matsumoto et al. 2009). They found that histological features of acute inflammation, such as acute inflammatory cell infiltrates, were found when an irregular mucosal vascular pattern was present vs. normal mucosal vascular pattern (26% vs. 0%  $p=0.0001$ ) (Kudo, Matsumoto et al. 2009). *Danese et al* noted the vascular pattern in IBD when viewed using NBI, as darker capillary vascular pattern or milder/regular capillary vascular pattern (Danese, Fiorino et al. 2010). A total of 14 patients with IBD were recruited (8 UC 6 CD), whereby biopsies were taken at the same site of the endoscopic assessment, whereby CD31 staining took place (immunohistochemistry) to determine the microvascular density (Danese, Fiorino et al. 2010). They found that in those patients with normal appearances under white light endoscopy (WLE), patients with a darker capillary vascular pattern under NBI showed an increased microvessel density ( $p<0.05$ ) (Danese, Fiorino et al. 2010). Although a small pilot study, this was one of the first to demonstrate the value of novel technology in the ability to show subtle features that can help predict microscopic changes. *Sasanuma et al* recruited patients with UC who were in clinical

and endoscopic remission and completed magnifying NBI colonoscopy (Sasanuma, Ohtsuka et al. 2018). The findings under NBI magnification were categorised into 3 groups describing blood vessel pattern: 1. Honeycomb-like appearance 2. Bare branches 3. Shaped like vines (Sasanuma, Ohtsuka et al. 2018). Mucosa defined as group 3 were more likely to have inflammation (81%) compared with the first 2 groups (4%/3%) (Sasanuma, Ohtsuka et al. 2018). Considering the small sample size (n=52) and single centre, this study showed that NBI can demonstrate vascular features in keeping with inflammation, however using magnification may not be the optimum method of utilising VEC in widespread clinical practice, particularly in the West.

#### *2.6.2.2 iScan imaging*

iScan imaging (Pentax, Tokyo, Japan) is an alternative imaging technique which enhances the surface and vascular pattern of bowel mucosa, by manipulating dark-light borders and red, blue and green components of light. iScan imaging is a post-processing image enhancement method and consists of three modes, namely surface enhancement (SE), contrast enhancement (CE) and tone enhancement (TE) (Iacucci, Fort Gasia et al. 2015). SE describes the process whereby luminance intensity of pixels between an area of interest and the surrounding mucosa is analysed, thereby allowing edges to become more prominent (Kodashima and Fujishiro 2010), whilst not altering colour, giving a sharp image. CE is the setting whereby areas of mucosa are assessed for luminance intensity, and areas that are low compared with surrounding areas, are stained in a bluish colour, giving enhanced imaging of subtle mucosal features (Kodashima and Fujishiro 2010). During TE the Red, Green and Blue component of light are disintegrated and re-synthesised into an image, increasing minor differences in colour tone of mucosa, thereby potentially increasing detection and characterisation of mucosal abnormalities (Kodashima and Fujishiro 2010).

The first study using iScan imaging at assessing mucosal inflammation in IBD was completed by *Iacucci et al* in a prospective single centre study (Iacucci, Kiesslich et al. 2017). In this study the iScan settings (iScan 1: surface enhancement, iScan 2: tone enhancement, iScan 3: contrast enhancement) were used to enhance mucosal (graded as normal, mosaic pattern, tubular-gyrus or nodular roundish rosette) and vascular (graded as normal, spiral isolated vessels, crowded tortuous or irregular vessels) abnormalities. The ability of VEC to demonstrate subtle findings was highlighted in this study, whereby 30.4% of Mayo 0 patients had abnormal mucosal patterns and 73.9% had abnormal vascular patterns when the mucosa was viewed using iScan imaging (Iacucci, Kiesslich et al. 2017). This study provides the evidence for the value of VEC over standard white light endoscopy in detecting subtle inflammatory findings.

*Neumann et al* aimed to assess whether iScan can improve assessment of mucosal features of inflammation in IBD compared with white light endoscopy (Neumann, Vieth et al. 2013). They recruited prospectively IBD patients attending for colonoscopy, including patients who had mild/quiescent disease activity (Mayo 0/1). Patients were randomised to either high-definition white light endoscopy or iScan, and scored the UC patients using the MES (Neumann, Vieth et al. 2013), whereby biopsies were taken from the most distal part of the inflammation and the samples were analysed histologically using the Riley score (Riley, Mani et al. 1988). The agreement between endoscopic prediction of inflammatory activity and the histological grading of activity was 50% in the WL group and 88.9% in the iScan group ( $p=0.046$ ) (Neumann, Vieth et al. 2013). This study, although limited by being a single centre and single endoscopists study, as well as using the MES to grade endoscopic activity, this study does give promise that VEC is not only an accessible technology in everyday clinical practice but can also allow accurate prediction of inflammatory activity.

### 2.6.2.3 Blue-light laser imaging (BLI) and Link-colour imaging (LCI)

BLI and LCI (Fujifilm, Tokyo, Japan) is another mode of VEC. BLI uses direct unfiltered blue light which has a short wavelength of 410nm (Bisschops, Hassan et al. 2018), which is absorbed by vascular microarchitecture, similar to narrow band imaging. LCI uses a narrow-band shortened wavelength and white light simultaneously, which allows differentiation and expansion of colours meaning red colours appear redder (Fukuda, Miura et al. 2015). As previously mentioned, during the process of inflammation there is an increase in blood vessels compared with normal tissue, which can become more apparent using BLI/LCI.

LCI was investigated for the first time by *Uchiyama et al*, following recruitment of 52 patients in clinical remission and a total of 193 areas of mucosa examined (Uchiyama, Takagi et al. 2017). They developed the following LCI criteria which was designed to reflect the ability of LCI to detect redness of colonic mucosa: LCI A: No redness LCI B: Redness with visible vessels LCI C: Redness without visible vessels (Uchiyama, Takagi et al. 2017). Three endoscopists reviewed images taken from the colonoscopy and graded inflammatory activity using the Mayo Endoscopic Score and the LCI criteria and correlated with the Matts Histological Index (Matts 1961). Relapse was defined as clinical relapse or a higher endoscopic score over a 30-month study period. The interobserver agreement between the endoscopists when using the Mayo score and LCI score was excellent. In patients who were scored as LCI-A, 94.0% had Matts 2 index (some infiltration of the mucosa or lamina propria and submucosa) whereas LCI-C patients 18.1% had Matts 2, and there was a lower relapse rate with LCI-A compared with LCI-B and LCI-C ( $p=0.0055$ ) showing that LCI is able to accurately predict inflammatory activity (Uchiyama, Takagi et al. 2017).

### **2.6.3 Magnifying colonoscopy**

To bridge the gap between endoscopic and histological assessments of inflammatory activity, magnifying colonoscopy has been investigated, particularly since its value has been widely reported in the setting of colorectal neoplasia (Sano, Tanaka et al. 2016). In a pioneering study, *Matsumoto et al* aimed to establish the features of UC under magnification and whether it can accurately predict histological activity (Matsumoto, Kuroki et al. 1997). They examined a total of 41 patients with UC, firstly by examining the rectum using standard colonoscopy, and grading inflammatory activity as quiescent or active. The endoscopists then applied dye to allow chromoendoscopy, and then obtained magnified views (100x) and they graded inflammatory activity based on the appearance of the cryptal openings (with or without) and the network pattern (visible/invisible) (Matsumoto, Kuroki et al. 1997). The authors found that when a visible network pattern and cryptal opening was present there was an associated with a lower histological activity of disease (compared with one not present  $p=0.0172$ ). This study gave an insight into how endoscopy can approximate histological findings, however the main limitation of this technology in everyday practice includes the application of the dye, the expertise required to obtain magnified images and the fact that the technology is used in few academic centres.

### **2.6.4 Other advanced endoscopic technologies**

Confocal laser endomicroscopy (CLE, Cellvizio, Mauna Kea Technologies, Paris, France) is a novel technology that provides microscopic images of cellular morphology in the gastrointestinal tract *in vivo*. Light strikes a fluorescent contrast agent and light is then emitted back and refocused onto a detection system (Committee 2014, Rasmussen, Karstensen et al. 2015). Optics are arranged so that elements in focus contribute to the image, light from above

and below are excluded giving a sharp clear image at approximately 1000x magnification (Rasmussen, Karstensen et al. 2015, Tontini, Rath et al. 2016, Watson 2016).

CLE gives histological images below the surface of epithelium in real time during endoscopy (Rasmussen et al 2015), allowing visualisation of architectural, cellular patterns enabling the visualisation of dynamic real-time processes (Neumann, Vieth et al. 2012, Rasmussen, Karstensen et al. 2015). The assessment of inflammatory activity and extent of mucosal inflammation is enhanced (Neumann, Vieth et al. 2012). CLE can offer a particular role in detecting barrier loss in vivo, and dynamic changes in vascular structure that occur during inflammation which may not be apparent on standard histology (Buda, Hatem et al. 2014).

Endocytoscopy (Olympus, Tokyo, Japan) utilises a fixed-focus high power objective lens that allows magnification up to 1400x giving images at the cellular level (Rath, Tontini et al. 2015), which is safe and can predict histology with high accuracy (Neumann, Fuchs et al. 2011). Therefore, imaging that is offered by endocytoscopy could potentially enhance and better assess the diagnosis of mucosal healing and detect subtle inflammation. An early scoring system was developed when using endocytoscopy in UC, identifying features in keeping with inflammation such as shape of crypts, distance between neighbouring crypts and visibility of superficial microvessels (Bessho, Kanai et al. 2011). This score showed good correlation with histopathology ( $\rho=0.713$   $p<0.001$ ) with good interobserver agreement ( $\kappa=0.79$  95% CI 0.71-0.87) (Bessho, Kanai et al. 2011). *Neumann et al* confirmed that endocytoscopy can detect subtle inflammatory mucosal changes in UC by recruiting patients with quiescent IBD and performed endocytoscopy alongside standard white light endoscopy. Despite endoscopic remission, all patients had evidence of inflammatory cells, which meant the level of activity was upgraded to slight activity. The concordance with histopathology was

100% (Neumann, Vieth et al. 2013).

Advancements in technology such as confocal laser endomicroscopy and endocytoscopy, allow imaging at the cellular level in real-time, however as these technologies are only available in few academic units, are expensive, require expertise and training they are not in routine clinical use.

## 3 THE REAL-LIFE PROSPECTIVE VALIDATION STUDY OF THE PICASSO SCORE

### 3.1 Overview

In the following chapter I discuss mucosal healing (MH) as a target for therapy in UC, which was covered more extensively in Chapter two. Following on from the discussion on novel endoscopic technologies, I present the PICaSSO score which is the first endoscopic score in UC to utilise VEC and includes features of MH. I then present the importance of histological assessment of inflammation and healing and the role it plays in predicting future disease course. Finally, I present the first real-life validation study of the PICaSSO score, demonstrating its correlation with histology and accuracy in predicting histological remission.

### 3.2 Background

As previously mentioned in Chapter two, the proposed treatment paradigm in ulcerative colitis (UC) is the “Treat to target” strategy, whereby clinicians aim to adjust therapy to achieve the target of MH (Peyrin-Biroulet, Sandborn et al. 2015). MH is associated with clear benefits including maintaining clinical remission, avoidance of colectomy, reduced steroid use and hospitalisation (Ferrante, Vermeire et al. 2008, Colombel, Rutgeerts et al. 2011, Shah, Colombel et al. 2016). There is increasing evidence that the definition of MH is not fit for purpose with guidelines suggesting an absence of mucosal lesions at endoscopy (Magro, Gionchetti et al. 2017) and clinical trials often using a MES (Schroeder, Tremaine et al. 1987) score of 0 or 1 as MH. Using MES 0/1 as a definition of MH is limited, since relapse rates are higher for MES 1 than MES 0 (Barreiro-de Acosta, Vallejo et al. 2016, Boal Carvalho, Dias

de Castro et al. 2016). Furthermore, in a large retrospective study, only histological normalisation was associated with relapse free survival (Christensen, Hanauer et al. 2017). The PICaSSO (Paddington International Virtual ChromoendoScopy ScOre) is the first endoscopic score to define mucosal and vascular characteristics of inflammation in ulcerative colitis by utilising enhanced imaging that is possible when using VEC (Iacucci, Daperno et al. 2017). It represents the first endoscopic score that defines MH in a positive manner as opposed to merely an absence of inflammation as is the case with the MES and UCEIS. It consists of two components: the mucosal architecture and vascular architecture. During the development of the score and following a 60-minute training session, endoscopic videos of UC were scored by experts in IBD and the PICaSSO score had a very good inter-observer agreement ( $k=0.917$  95% CI 0.87-0.96), which was superior to the MES and UCEIS (Iacucci, Daperno et al. 2017). The PICaSSO score was validated in an *ex vivo* setting amongst endoscopists previously not experienced with iScan (Trivedi, Kiesslich et al. 2018). When comparing pre- vs. post-training performances, there was improved discrimination of quiescent vs. mild disease (AUROC PICaSSO mucosal architecture 0.781 vascular 0.715 Mayo 0.708 UCEIS 0.705) (Trivedi, Kiesslich et al. 2018). The PICaSSO score clearly has several strengths, not least it does not require magnification, uses technology that is widely available and is accessible at the touch of a button but importantly actively defines mucosal healing. However, to date the studies completed in its development are limited by being completed in an *ex vivo* setting using videos/images (Iacucci, Daperno et al. 2017, Trivedi, Kiesslich et al. 2018).

### 3.3 Histological assessment of inflammatory activity

#### 3.3.1 Definition of histological healing

Histological assessment of biopsies has long had a vital role in clinical practice, however, is not currently considered as a definition of healing nor is considered as an end point in the majority of clinical trials (Marchal Bressenot, Riddell et al. 2015). Even though there are 18 invalidated scoring systems in use in UC, *Marchal Bressenot et al* comment that the components required to define healing are present amongst these scores but require “fine tuning” (Marchal Bressenot, Riddell et al. 2015). Indeed, Villanacci and colleagues commented that for a histological score to be widely used in everyday practice, not just in a research setting, the score should be simple, taking into account epithelial cell damage, increase in lymphocytes and plasma cells, and presence of neutrophils indicating active inflammation (Villanacci, Antonelli et al. 2015). Indeed, whilst there is no accepted definition of MH in histological terms, termed histological remission (HR), it is widely agreed that the absence of neutrophils in the crypts and lamina propria is indicative of healing or rather the absence of inflammation (Bryant, Winer et al. 2014, Villanacci, Bassotti et al. 2015). Some authors suggest the absence of basal plasma cells is indicative of HR (Bryant, Winer et al. 2014), however others argue that the presence of basal plasma cells is present in all forms of activity of IBD (Villanacci, Antonelli et al. 2014). HR is an important goal for the treatment of UC although is not currently defined and is not considered in recent recommendations (Peyrin-Biroulet, Sandborn et al. 2015). An Italian pathology group suggested a definition of healing is the “complete resolution of the inflammatory process” and mucosa that have no signs of active inflammation, such as a lack of neutrophils in the lamina propria (Villanacci, Reggiani-Bonetti et al. 2019). The lack of agreement on how HR should be defined represents one of the major barriers for HR being considered a key target.

### **3.3.2 Histological inflammation and endoscopic scoring of inflammatory activity**

In an early study, *Riley et al* recruited 82 UC patients who were in clinical (normal stool frequency and no blood in stools) and endoscopic remission (normal Baron score) (Riley, Mani et al. 1991). 32% of patients had histological evidence of acute inflammation, 33% of patients suffered relapse, which was confirmed at endoscopy as mucosa that had become haemorrhagic (Riley, Mani et al. 1991). In patients with acute inflammatory cell infiltrate on biopsies, the relapse rate was 52% within 12 months, whereas the absence was associated with a rate of 25% ( $p=0.02$ ) and the presence of crypt abscess was associated with higher relapse (78% vs. 27%  $p<0.005$ ) (Riley, Mani et al. 1991). This early study indicates the role histological inflammation can play, albeit being limited by out-dated endoscopic assessment of MH.

Patients with apparent endoscopic MH can suffer early relapse within 12 months, demonstrating the importance of HR as a marker of remission. Microscopic inflammation in patients with otherwise endoscopic quiescent colitis is remarkably common (Bryant, Winer et al. 2014), which further highlights the importance of accurate endoscopic assessments using novel endoscopic technology.

*Kanazawa et al* identified the importance of HR in terms of remission in UC (Kanazawa, Takahashi et al. 2019). The authors retrospectively identified patients with ulcerative colitis who were deemed to have mucosal healing (MES 0/1) at colonoscopy and remained in clinical remission at 12 months without significant changes in medications. Of the 166 patients included, 26.5% had evidence of histological inflammation (Kanazawa, Takahashi et al. 2019). Of the patients with Mayo 0 and Mayo 1 disease, 8.8% and 48.0% had evidence of histological inflammation respectively (Kanazawa, Takahashi et al. 2019). This adds further weight to the argument that Mayo 1 should not be considered as healing.

*Bryant et al* completed a systematic review of the evidence for HR as a target for therapy in IBD (Bryant, Winer et al. 2014). It is clear that when using existing endoscopic scores, despite endoscopic findings of quiescent disease histological inflammation can be identified (Bryant, Winer et al. 2014). However, using HR as a key target has limitations, namely the lack of validation, lack of clear definition of remission and is reliant on mucosal sampling, with the inherent risk and cost (Bryant, Winer et al. 2014).

### **3.3.3 Histological inflammation and risk of relapse**

*Kevans et al* completed a retrospective study assessing the link between histological activity and relapse in UC (Kevans, Kirsch et al. 2019). They found that active histological inflammation in patients with endoscopic quiescent disease was associated with higher clinical relapse at 18 months (60% vs. 16%  $p < 0.05$ ) and was associated with a shorter time to relapse (Kevans, Kirsch et al. 2019). This study demonstrates the importance histological inflammation plays in predicting adverse outcomes in patients, however it is limited by being retrospective, single centre and used the MES 0 and 1 as quiescent disease. As previously mentioned, this is not the most accurate means of assessing quiescence and may infer bias into the study.

A prospective study by *Bryant et al*, recruited 91 UC patients who underwent sigmoidoscopy and graded inflammatory activity using the Baron index (Baron, Connell et al. 1964), using a score of  $\leq 1$  defining remission (Bryant, Burger et al. 2016) and the Truelove and Richard's index to define histological remission ('no significant inflammation') (Truelove and Richards 1956). The agreement between endoscopic and histological activity was only moderate ( $k = 0.56$  95% CI 0.36-0.77) (Bryant, Burger et al. 2016). The authors completed follow up of 6 years following baseline endoscopy, and HR was associated with fewer steroids courses (HR

0.42 95% CI 0.2-0.9 p0.02) and hospitalisation (HR 0.21 95%CI 0.1-0.7 p0.02) (Bryant, Burger et al. 2016). This study does infer that histological remission can predict a favourable prognosis, however the single centre design, the modest number of patients and the use of out-dated endoscopic and histological indices means it may lack relevance in the setting of advanced endoscopic technology.

*Narang et al* completed a prospective observational study whereby UC patients achieving clinical remission (Simple Clinical Colitis Activity Index  $\leq 2$ ) and endoscopic remission (Mayo  $\leq 1$ ) underwent colonoscopy and rectosigmoid biopsies (Narang, Kaur et al. 2018). Histological assessment took place in the form of the Geboes index (Geboes, Riddell et al. 2000) and 67.3% of patients were deemed to have histological remission (Geboes 0/1). At 12 months follow up, patients with endoscopic and histological remission were more likely to avoid relapse (87.1% vs. 46.6% p=0.006) (Narang, Kaur et al. 2018). This single centre, single pathologist study adds further to the argument for histological remission to be considered as a crucial target. It also confirms the limitation of the MES since a third of patients with Mayo 0 had evidence of histological activity (Narang, Kaur et al. 2018).

A Portuguese based study included 60 UC patients who were in clinical remission as defined by the partial Mayo score, and in endoscopic remission, defined as MES 0/1 (Ponte, Pinho et al. 2017). The authors demonstrated that histological inflammation based on a Nancy index score of  $\geq 2$  was an independent predictor of relapse (OR 3.7 95% CI 1.1-12.3 p0.034) (Ponte, Pinho et al. 2017). They also highlighted the increase in relapse with Mayo 1 vs. Mayo 0 (p=0.02) (Ponte, Pinho et al. 2017). This study is limited by its retrospective, single centre design and a small sample size (n=60).

Calafat and colleagues completed a retrospective observational study, whereby UC patients with endoscopic remission (Mayo  $\leq 1$ ) were observed for up to 12 months, including 113 patients, most of which had quiescent disease (Mayo 0=86%) (Calafat, Lobaton et al. 2017). In this study, biopsies were analysed for histological activity by allocating into one of three groups: Normal or quiescent histological activity (NQHA), chronic histological activity (CHA) and acute histological activity (AHA). AHA was associated with relapse within 12 months in univariate analysis ( $p=0.048$ ), which together with previous flare in last 12 months were the only predictors of relapse (Calafat, Lobaton et al. 2017). Although retrospective and single centre, this further highlights the importance of histological assessment.

*Christensen et al* assessed the impact of complete histological normalisation, which was defined as normalisation of mucosa without chronicity, had on clinical outcome (Christensen, Hanauer et al. 2017). Patients who achieved histological normalisation suffered fewer relapses compared with quiescence and activity ( $p<0.05$ ), and, through multivariate analysis, complete histological normalisation was associated with clinical relapse free survival (Christensen, Hanauer et al. 2017). However, this study is limited by its retrospective single centre design, a lack of detailed information on medication use and the use of the MES.

*Zenlea et al* enrolled 179 patients with UC prospectively who were in clinical remission who underwent colonoscopy and were subsequently followed up for 12 months (Zenlea, Yee et al. 2016). The Geboes grade  $\geq 3.1$  was the only variable associated with clinical relapse in univariate and multivariate analysis ( $p=0.006$ ) with a RR of 3.5 (95% CI 1.9-6.4  $p<0.0001$ ) (Zenlea, Yee et al. 2016). Only 7% of patients who were in endoscopic, clinical and histological remission experienced relapse (Zenlea, Yee et al. 2016).

*Park et al* completed a meta-analysis on the effect HR has on predicting outcome (Park, Abdi et al. 2016). Overall, HR was associated with fewer relapses vs. histological inflammation (RR 0.48 95% CI 0.39-0.60), and vs. endoscopic/clinical remission (RR 0.81 95% CI 0.70-0.94) (Park, Abdi et al. 2016). When assessing specific histological features, it is clear that an absence of neutrophils in epithelium (RR 0.32 95% CI 0.23-0.45) and lamina propria (RR 0.43 95% CI 0.32-0.59) are important in defining HR and therefore a lower risk of relapse (Park, Abdi et al. 2016). The authors also noted nearly 30% of patients with endoscopic and clinical evidence of remission had histological evidence of active inflammation (Park, Abdi et al. 2016).

*Bessissow et al* aimed to assess histological variables and the impact on relapse and the correlation with endoscopic activity (Bessissow, Lemmens et al. 2012). The authors retrospectively reviewed all UC patients who underwent surveillance colonoscopy, including patients with inactive disease at endoscopy (MES 0) who had documented follow up for 12 months. A single pathologist scored the biopsies taken, using the Geboes score (Geboes, Riddell et al. 2000), in addition to noting the degree of basal plasmacytosis. Interestingly the presence of basal plasmacytosis was the only predictor of clinical relapse within 12 months in multivariate analysis OR 5.13 (95% CI 1.32-19.99 p0.019) (Bessissow, Lemmens et al. 2012). Of the 75 patients with endoscopic quiescent disease included, histological activity was present in 40% noted as a Geboes score  $\geq 3.1$ , further indicating that endoscopic assessments of inflammatory activity are often inaccurate (Bessissow, Lemmens et al. 2012). The finding of the presence of basal plasmacytosis as a predictor of relapse in UC supports the early finding by *Bitton et al* who prospectively recruited patients with UC whom were in clinical remission (baseline bowel habit and no rectal bleeding) and endoscopic remission (0- normal-vascular pattern clearly visible/1-erythema with loss in vascular pattern) (Bitton, Peppercorn

et al. 2001). Patients underwent sigmoidoscopy at 6 and 12 months after baseline or if clinically indicated for suspicion of relapse. Biopsies were taken and graded as active if neutrophils were noted in the lamina propria, cryptitis, crypt abscesses and erosions/ulceration and basal plasmacytosis was noted as a feature of chronic inflammation (Bitton, Peppercorn et al. 2001). In addition to a younger age and number of previous relapses in females, the presence of basal plasmacytosis was the strongest predictor of relapse within 12 months OR 4.5 95% CI 1.7-11.9 p0.003, and Kaplan-Meier curve analysis shows a favourable survival benefit in the absence of basal plasmacytosis (Bitton, Peppercorn et al. 2001).

Histological analysis is vital in the management of patients with UC and helps guide management. It is increasingly clear that HR is a strong predictor of survival free from relapse and that histological inflammation is frequently present in current endoscopic definitions of MH. One limitation is the lack of consensus as to how to define histological healing but given VEC can allow assessments of inflammation that can approximate histology there is a need for an endoscopic scoring system that reflects this (Iacucci, Fort Gasia et al. 2015).

A major current need in IBD endoscopy is an endoscopic score that describes MH in endoscopic terms which correlates with HR. Currently available advanced images using high definition (HD) endoscopy, with virtual electronic chromoendoscopy, (VEC) can enhance endoscopic assessment and can detect subtle mucosal and vascular changes associated with histological changes (Iacucci, Fort Gasia et al. 2015, Iacucci, Cannatelli et al. 2020). As previously mentioned, the **P**addington **I**nternational virtual **C**hromoendo**S**cOPY **S**cOre (PICaSSO) is the first to use VEC images to assess vascular and mucosal details of healing.

A number of histological indices are currently available that include active or chronic changes or both such as Robarts Histopathology Index (RHI) (Mosli, Feagan et al. 2017), Nancy Histological Index (NHI) (Marchal-Bressenot, Salleron et al. 2017), Villanacci Simple Score (VSS) (Villanacci, Antonelli et al. 2017), Geboes Score (Geboes, Riddell et al. 2000) and ECAP (Iacucci, Fort Gasia et al. 2015) (Extent, Chronicity, Activity and Plus Score). It has been shown that the PICaSSO score using VEC images correlated well with histological scores and had a very good interobserver agreement amongst experts and non-experts (Iacucci, Daperno et al. 2017, Trivedi, Kiesslich et al. 2018).

### 3.4 Aims and objectives

The aims of this study are, for the first time, to prospectively establish the performance and relationship of the PICaSSO score and other endoscopic scores with several histological scores in real-life UC patients and their associated with HR. A further aim was to evaluate whether PICaSSO was predictor of specified clinical outcomes.

The primary objectives for this study were to (a) establish the correlation between endoscopic scoring systems for activity defined by PICaSSO (Iacucci, Daperno et al. 2017) (ranging from 0-15), MES (Schroeder, Tremaine et al. 1987), UCEIS (Travis, Schnell et al. 2012) and histological activity defined by multiple histological scores including RHI (Mosli, Feagan et al. 2017), NHI (Marchal-Bressenot, Salleron et al. 2017), GS (Geboes, Riddell et al. 2000), VSS (Villanacci, Antonelli et al. 2017) and ECAP (Iacucci, Fort Gasia et al. 2015).

The secondary objectives were to (a) to assess ability of PICaSSO score to predict clinical outcomes indicative of relapse in UC patients at 6 and 12 months, (b) to assess the inter-observer agreement of PICaSSO between all raters and compare with other endoscopic

scoring systems and (c) to determine from Receiver Operator Characteristics [ROC] the best thresholds of PICaSSO, MES and UCEIS associated with HR.

## 3.5 Methods

### 3.5.1 Study design

This was a multi-centre study performed in eleven international centres between September 2016 and November 2019. Adult patients with an established diagnosis of UC were prospectively enrolled meeting inclusion criteria:

- Age  $\geq 18$  and  $\leq 80$  years
- A clinical indication for assessment of activity of UC
- An established diagnosis of UC for  $\geq 1$  year.

Patients were excluded if any of the following were present:

- An inability to provide consent
- Presence of serious co-morbidities
- Toxic megacolon
- Pregnancy or breast feeding
- Contraindication of biopsies
- Boston bowel preparation score  $< 2$  (Calderwood and Jacobson 2010) in the rectum or sigmoid colon.
- Diagnosis of unclassified colitis, Crohn's colitis, ischaemic colitis or infectious colitis

In order to study the relationship between PICaSSO and histology, for the first 20 patients recruited, each participating site included quiescent, mild, moderate and severe inflammatory activity based on the clinical partial Mayo score (0-1=quiescent, 2-4=mild, 5-6=moderate, 7-9=severe) at the time of recruitment (Schroeder, Tremaine et al. 1987). Subsequently, to

establish the ability of PICaSSO to predict HR, sites were asked to recruit (n=20) mild/quiescent disease (clinical partial Mayo score 0-4), such as patients attending for surveillance purposes or assessment after therapy. This was in recognition of the challenges of MES-based MH criteria to relate to HR. All patients who had colonoscopy after consent were included in the study unless they had unsatisfactory bowel preparation.

### **3.5.2 Endoscopic assessment**

Endoscopic examinations were performed by designated endoscopists who were experienced in IBD and i-Scan imaging. Prior to the start of the study each centre received refresher training materials on iScan digital image enhanced technology and the PICaSSO score (Iacucci, Daperno et al. 2017). Training lasted 1 hour and consisted of examples of annotated images (n=56) and annotated videos (n=4) representative of each mucosal and vascular endoscopic finding.

A polyethylene glycol-based bowel preparation was administered to patients prior to colonoscopy in accordance with each centres protocol. Data was collected at the time of the colonoscopy including baseline demographics, duration of disease, extent of colitis and current/previous medication history. Data was recorded on case report forms (CRF) before being transferred to REDcap (The Vanderbilt University, Nashville, Tennessee, USA).

All examinations were performed using HD Pentax (Tokyo, Japan) 7010 processor and colonoscopes of the HiLine series. The mucosa was washed thoroughly prior to the sigmoid colon and rectum being viewed for at least 60 seconds in each of the following modes: HD, iScan1, iScan2 and iScan3, which can be switched at the push of a button on the endoscope

(Iacucci, Fort Gasia et al. 2015). Standardised settings were used in each centre (Figure 3-1). Endoscopists first graded the inflammatory activity by using the MES (Schroeder, Tremaine et al. 1987), and UCEIS (Travis, Schnell et al. 2012) using HD white light. This was followed by VEC and the PICaSSO score (Figure 3-2) (Iacucci, Daperno et al. 2017) in the sigmoid and rectum. Two targeted biopsies were taken from the most inflamed area or showing the most representative features of endoscopic remissions determined by PICaSSO. The assessment and biopsy location were captured by video recording for all patients.

|              | iScan 1 | iScan 2 | iScan 3 |
|--------------|---------|---------|---------|
| Brightness   | 0       | +1      | +1      |
| Average/Peak | Ave     | Ave     | Ave     |
| Blue         | 0       | 0       | 0       |
| Red          | 0       | 0       | 0       |
| Enhancement  | Low/+2  | Low/+2  | Low/+2  |
| SE           | +5      | +4      | +4      |
| TE           | Off     | Colon   | Gastric |
| OE           | Off     | Off     | Off     |

**Figure 3-1 iScan settings used in the real-life prospective validation study of the PICaSSO score**

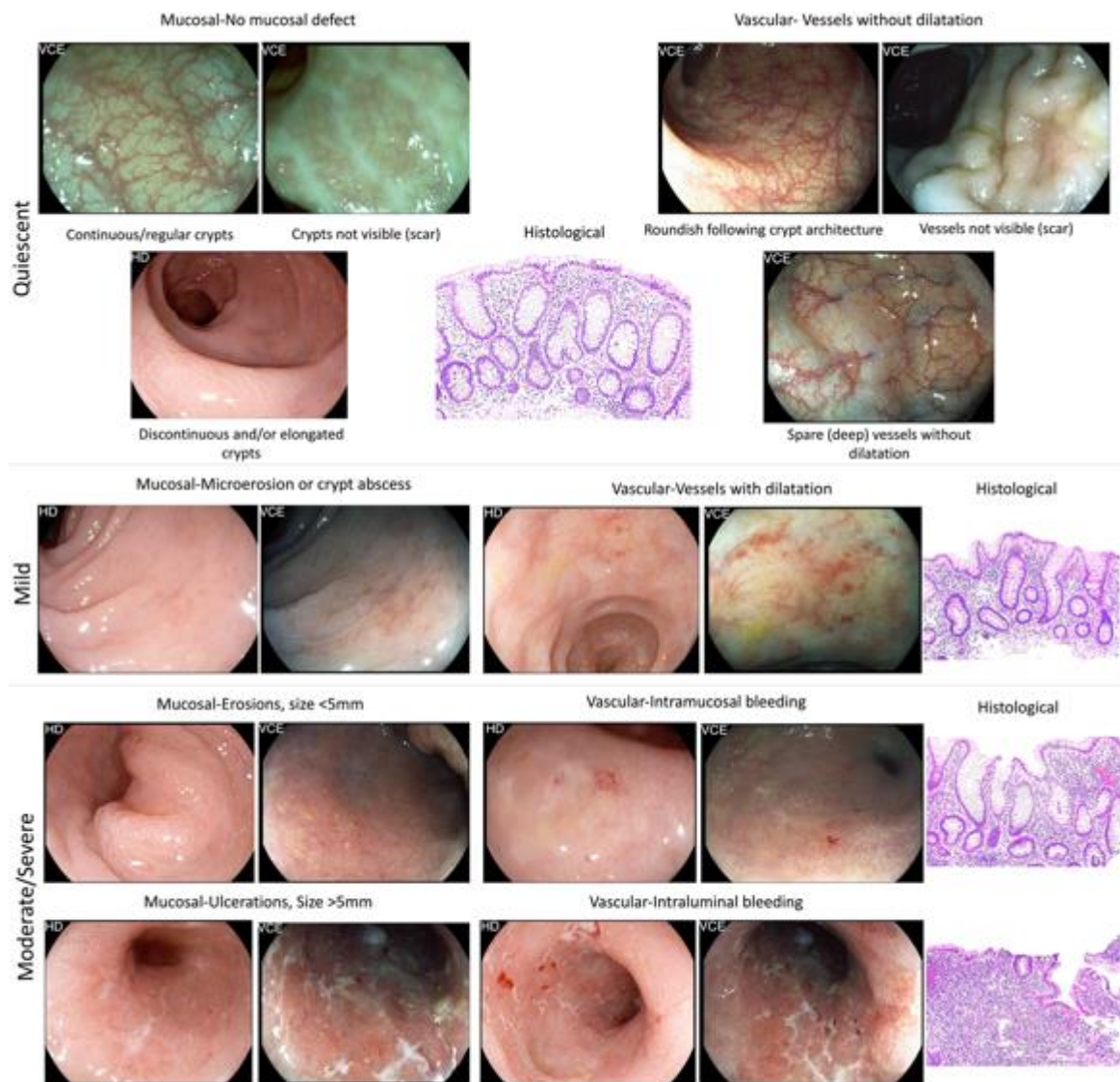


**Figure 3-2 The PICaSSO Score**

Example images of the PICaSSO Score categories, alongside example images of histological activity associated with endoscopic findings

### 3.5.3 Histological assessment

All biopsies were fixed in formalin and processed at institutional pathology laboratories with routine embedding and staining protocols. The haematoxylin-eosin stained sections from the rectum and sigmoid biopsies of patients recruited were digitalised using high-speed slide scanners by participating centres. This allowed histological scoring by five pathologists who were blinded to clinical features and endoscopic findings. Each slide was scored using RHI

Figure 3-3 (Mosli, Feagan et al. 2017), Extent, Chronicity, Activity, Plus Score (ECAP)

Figure 3-4 (Iacucci, Fort Gasia et al. 2015), Nancy Histological Index (NHI)

Figure 3-5 (Marchal-Bressenot, Salleron et al. 2017), Villanacci Simple Score (VSS)

Figure 3-6 (Villanacci, Antonelli et al. 2017) and Geboes Scores (GS) Figure 3-7 (Geboes, Riddell et al. 2000) for comparison with endoscopic grading of inflammatory activity. These histological indices were selected as they represent different combinations of the major histopathologic features of UC, including active inflammation ('activity', i.e., neutrophil infiltration in glandular epithelium and in lamina propria), chronic inflammation (mononuclear cell infiltration in lamina propria), crypt architectural alteration (chronicity).

Component

Chronic inflammatory infiltrate

0 = No increase

1 = Mild but unequivocal increase

2 = Moderate increase

3 = Marked increase

Lamina Propria neutrophils

0 = None

1 = Mild but unequivocal increase

2 = Moderate increase

3 = Marked increase

Neutrophils in epithelium

0 = None

1 = <5% crypts involved

2 = <50% crypts involved

3 = >50% crypts involved

Erosion or ulceration

0 = No erosion, ulceration or granulation tissue

1 = Recovering epithelium + adjacent inflammation

1 = Probable erosion-focally stripped

2= Unequivocal erosion

3 = Ulcer or granulation tissue

**Figure 3-3 The Robarts Histological Index**

|                                                  |                                              |
|--------------------------------------------------|----------------------------------------------|
| Extent (E):                                      | A4. Crypts destruction                       |
| Extent of inflammation                           | 0 = None                                     |
| 0 = None                                         | 1 = Present                                  |
| 1 = Focal                                        | A5. Lamina Propria mononuclear cellularity   |
| 2 = Multifocal (patchy)                          | 0 = Normal                                   |
| 3 = Diffuse                                      | 1 = Mild increase                            |
| Chronicity (C)                                   | 2 = Moderate increase                        |
| C1. Crypt architectural alteration               | 3 = Severe increase                          |
| 0 = None                                         | A6. Basal plasmacytosis                      |
| 1 = Focal alteration                             | 0 = None                                     |
| 2 = Patchy distortion (<50%)                     | 1 = Focal                                    |
| 3 = Diffuse distortion (>50%)                    | 2 = Diffuse                                  |
| C2. Paneth cell metaplasia                       | A7. Lamina Propria neutrophilic infiltration |
| 0 = None                                         | 0 = None                                     |
| 1 = Present                                      | 1 = Rare                                     |
| Activity Inflammation (A)                        | 2 = Scattered                                |
| A1. Surface epithelium                           | 3 = Extensive                                |
| 0 = Normal                                       | Plus. others (P)                             |
| 1 = Reactive changes (mucin depletion villiform) | P1. Lamina Propria eosinophilic infiltration |
| 2 = Neutrophilic infiltration/ probable erosion  | 0 = None                                     |
| 3 = Erosion                                      | 1 = Mild                                     |
| 4 = Ulceration                                   | 2 = Moderate                                 |
| A2. Neutrophilic cryptitis                       | 3 = Severe                                   |
| 0 = None                                         | P2. Lymphoid follicle/aggregates             |
| 1 = <5%                                          | 0 = None                                     |
| 2 = <50%                                         | 1 = Rare                                     |
| 3 = >50%                                         | 2 = Prominent                                |
| A3. Crypt abscess                                |                                              |
| 0 = None                                         |                                              |
| 1 = Present                                      |                                              |

**Figure 3-4 Extent, Chronicity, Activity inflammation and Plus others (ECAP)**

#### **histological index**

**Grade 0-**

Ulceration: None

Acute inflammatory infiltrate: None

Chronic inflammatory infiltrate: None or mild

**Grade 1-**

Ulceration: None

Acute inflammatory infiltrate: None

Chronic inflammatory infiltrate: Moderate or marked increase

**Grade 2-**

Ulceration: None

Acute inflammatory infiltrate: Mild

Chronic inflammatory infiltrate: Moderate or marked increase

**Grade 3-**

Ulceration: None

Acute inflammatory infiltrate: Moderate or severe

Chronic inflammatory infiltrate: Moderate or marked increase

**Grade 4-**

Ulceration: Yes

Acute inflammatory infiltrate: Moderate or severe

Chronic inflammatory infiltrate: Moderate or marked increase

**Figure 3-5 Nancy histological index**

Crypt abscesses (presence of neutrophils aggressive on crypts)

0 = Absent

1 = Present

Erosions/ulceration (presence of granulation tissue)

0 = Absent

1 = Present

Neutrophils in the lamina propria

0 = Absent

1 = Present

Site of involvement

0 = Quiescent

**1 = Active**

**Figure 3-6 Villanacci histological score**

**Grade 0: Structural (architectural change)**

- 0.0 No abnormality
- 0.1 Mild abnormality
- 0.2 Mild or moderate diffuse/multifocal abnormality
- 0.3 Severe diffuse or multifocal abnormality

**Grade 1: Chronic inflammatory infiltrate**

- 0.0 No increase
- 1.1 Mild but unequivocal increase
- 1.2 Moderate increase
- 1.3 Marked increase

**Grade 2: Lamina Propria neutrophils and eosinophils**

- 2A Eosinophils
  - 2A.0 No increase
  - 2A.1 Mild but unequivocal increase
  - 2A.2 Moderate increase
  - 2A.3 Marked increase
- 2B Neutrophils
  - 2B.0 No increase
  - 2B.1 Mild but unequivocal increase
  - 2B.2 Moderate increase
  - 2B.3 Marked increase

**Grade 3: Neutrophils in epithelium**

- 3.0 None
- 3.1 <5% crypts involved
- 3.2 <50% crypts involved
- 3.3 >50% crypts involved

**Grade 4: Crypt destruction**

- 4.0 None
- 4.1 Probable-local excess of neutrophils in part of crypts
- 4.2 Probable- marked increase
- 4.3 Unequivocal crypt destruction

**Grade 5: Erosion or ulceration**

- 5.0 No erosion, ulceration or granulation tissue
- 5.1 Recovering epithelium + adjacent inflammation
- 5.2 Probable erosion- focally stripped
- 5.3 Unequivocal erosion
- 5.4 Ulcer or granulation tissue

**Figure 3-7 Geboes histological index**

The RHI was calculated as the total of the following: 1x chronic inflammatory infiltrate level + 2x lamina propria neutrophils + 3x neutrophils in epithelium + 5x erosion or ulceration (Mosli, Feagan et al. 2017). The ECAP was scored as a total of the following components: Extent of inflammation, Chronicity, Activity inflammation, Plus lamina propria eosinophilic infiltration and lymphoid follicle/aggregates (Iacucci, Fort Gasia et al. 2015). The Nancy Index was graded according to ulceration, acute inflammatory cell infiltrate and chronic inflammatory infiltrate and a grade of 0-4 was given to the sigmoid and rectum (Mosli, Feagan et al. 2017). The Villanacci histological index grades biopsies using categories including crypt abscesses, erosions/ulcerations, neutrophils in lamina propria and the site of involvement (Villanacci, Antonelli et al. 2017). Finally the Geboes histological index is graded according to structural change, chronic inflammation, lamina propria neutrophils, neutrophils in epithelium, crypt destruction and erosions/ulcers and biopsies are graded as 0-5 based on histological findings (Geboes, Riddell et al. 2000).

For the present study, histological healing was defined as RHI  $\leq 3$  plus an absence of neutrophils (Pai, Khanna et al. 2019), ECAP  $\leq 4$ , Nancy index  $\leq 1$ , Villanacci histological index 0, and Geboes grades 0/1.

HR was defined as RHI  $\leq 3$  (with absence of neutrophils (Pai, Khanna et al. 2019)), ECAP  $\leq 4$ , NHI  $\leq 1$ , VSS 0, and GS 0/1. The histological scores were used for correlation with endoscopic scores from the same site as the biopsies.

#### **3.5.4 Interobserver agreement**

To establish the interobserver agreement of the PICASSO score all participating endoscopists were provided with 40 randomly chosen high quality videos taken from the study displaying

quiescent and active disease. Each video clip, 60-90 sec in length, was recorded by using the four different modes of HD, iscan1, iscan2 and iscan3. The endoscopists scored each video using the MES, UCEIS and PICaSSO using CRFs hosted on REDcap (The Vanderbilt University, Nashville, Tennessee, USA).

### **3.5.5 Follow up and outcomes**

Clinical outcomes at follow up were defined as: i) hospitalisation as a result of UC relapse, ii) colectomy, iii) initiation or changes in medical therapy including steroids, immunomodulators and biologics due to UC relapse. The case notes of all patients were reviewed at each centre prospectively to record these clinical outcome events with telephone calls to clarify any uncertainty, at 6 months and 12 months following baseline procedure.

### **3.5.6 Statistical analysis**

The results were exported from REDCap to STATA Version 14 [StataCorp]. Mean  $\pm$  SD and median  $\pm$  interquartile range were determined on continuous variables. Assuming the correlation between PICaSSO- histology and MES- histology to be 0.72 and 0.61 respectively, a sample size of 302 patients was calculated to reach a power of 0.8 with type I error of 0.05 by means of Fisher's Z-transform to address the primary outcome of the study. A 1-sample superiority/non-inferiority test for proportions of relapse rate, and we needed 300 patients to reach power 0.8 with a type I error of 0.05 when comparing specified clinical outcomes of PICaSSO and MES=0 at 6 months follow-up. We accepted a 10% relapse rate for MES 0 (Barreiro-de Acosta, Vallejo et al. 2016) and we assumed 6.4% for PICaSSO score cut-off that best predicted HR.

Pearson correlation between the endoscopic scores and RHI, ECAP, Nancy, Villanacci index and Geboes were calculated. Very strong correlation was considered as a value of 0.8-1.0, strong as 0.6-0.79, moderate as 0.40-0.59 and weak as 0.2-0.39. Fisher Z-transformation by means of {R-package (<http://www.R-project.org/>) cocor} (Diedenhofen and Musch 2015) was performed to compare the correlation of each endoscopic score. A 1000 bootstrap sample was generated to compute the differences between ICCs between PICaSSO and Mayo/UCEIS, calculating the 0.01<sup>th</sup> quantiles of the resulting distributions.

To assess the interrater agreement of the PICASSO score, one-way intra class correlation coefficient by means of R package irr (<http://CRAN.R-project.org/package=irr>) was used. To test the hypothesis of the effect size being larger than 0.1, against the alternative of no effect, we needed a minimum of 30 videos to reach the power of 0.8 with a type I error of 0.05. Very strong correlation was considered as a value of 0.8-1.0, strong as 0.6-0.79, moderate as 0.40-0.59 and weak as 0.2-0.39. Meaningful misclassification rates were calculated by splitting each of the endoscopic scores into three categories: remission, mild, moderate/severe (Mayo 0,1,>1; UCEIS ≤1, 2-4, >4; Picasso ≤3, 4-8, >8 respectively). For each pair of raters the difference between their scorings of each video and recorded misclassification if this difference was greater than zero was calculated. To obtain the overall misclassification rate of the obtained rates over the number of videos and the number of pairs was averaged. Furthermore, to statistically compare the resulting misclassification rates between each other, 1000 values for differences between PICaSSO and UCEIS, PICaSSO and MES misclassification rates were computed, each based on 1000 bootstraps, and derived 0.99<sup>th</sup> quantiles of the obtained distributions.

Receiver Operating Characteristic (ROC) curves were plotted and area under receiver operating characteristic curve (AUROC) was determined to accurately reflect endoscopic scores that predict HR. The AUROCs were compared using DeLong test implemented within an R package pROC (<https://CRAN.R-project.org/package=pROC>) (Robin, Turck et al. 2011). Youden's j-statistics (Youden 1950) was used to determine the point on the curve with the best sensitivity and specificity. Accuracy of the corresponding point was also calculated.

Cox proportional hazard model was used to calculate the probability of survival without clinical relapse for each of the endoscopic scores. Hazard ratios have been presented and p values <0.05 were considered statistically significant. To compare the increment in specified clinical outcomes between MES=0 over MES=1 and PICaSSO  $\leq 3$  over PICaSSO 4-8 10000 bootstrap samples from data with MES  $\leq 1$  and PICaSSO  $\leq 8$  respectively were generated thereby obtaining two sets of incremental values. The p values were computed based on the distribution of the differences between each pair of increments.

To assess the individual weights of each PICaSSO component, multivariate logistic regression was performed, where each component of Picasso was treated as a binary variable. For that R-package net (Venables, Ripley et al. 2002) was used (<https://CRAN.R-project.org/package=nnet>).

### **3.5.7 Ethical approval**

The study was approved by West Midlands Research Ethics Committee (17/WM/0223) for the UK centres. All international sites gained local research ethics committee approval. All patients gave informed consent to participate in the study.

## 3.6 Results

### 3.6.1 Patient demographics

A total of 307 patients with UC from 11 international centres were prospectively included in the study with corresponding endoscopic and histological scores. A total of 302 patients with complete datasets were included in the final analysis. The patients were made up of 182 males (59.3%), with a median age of 48.4 (range 19-77) years and a mean duration of disease of 15.0 (SD  $\pm$  10.8). The majority of patients n=172 (56.0%) had extensive (E3) colitis, n=68 (22.1%) were on immunosuppressants and 118 (38.4%) were on biological therapy at baseline. The patient characteristics are shown in Table 3-1 and the endoscopic grades of activity of the patients included are presented in Table 3-2. At the time of colonoscopy 218 (71.0%) were on 5-ASA, 22 (7.2%) were on corticosteroids, 54 (17.6%) on immunomodulators, 104 (33.9%) on biologics and 29 (9.4%) were not on any IBD treatment. None of the patients were on topical therapies or had E1 disease.

**Table 3-1 Patient characteristics of the real-life prospective validation study of the PICaSSO score**

| <b>Characteristics</b>                 | <b>Patients (n=307)</b> |
|----------------------------------------|-------------------------|
| Age (y) <i>mean ± sd</i>               | 48.4 ± 14.8             |
| Gender Male n (%)                      | 182 (59.3%)             |
| Disease duration (y) <i>mean ± sd</i>  | 15.0 ± 10.8             |
| <b>Extension of disease n (%)</b>      |                         |
| Left-sided colitis                     | 130 (42.3%)             |
| Sub-total or total colitis             | 172 (56.0%)             |
| Unknown                                | 5 (1.6%)                |
| <b>Therapy in last 12 months n (%)</b> |                         |
| No treatment                           | 14 (4.6%)               |
| 5-ASA                                  | 234 (76.2%)             |
| Corticosteroids                        | 74 (24.1%)              |
| Immunomodulators                       | 68 (22.1%)              |
| Biologics                              | 118 (38.4%)             |

**Table 3-2 Endoscopic grading of inflammatory activity of patients included in the real-life prospective validation study of the PICaSSO score**

| <b>Endoscopic Score</b>            | <b>Patients n (%)</b> |
|------------------------------------|-----------------------|
| <b>Mayo Endoscopic Score n (%)</b> |                       |
| MES 0                              | 168 (54.7%)           |
| MES 1                              | 47 (15.3%)            |
| MES 2                              | 56 (18.2%)            |
| MES 3                              | 31 (10.1%)            |
| Missing data*                      | 5 (1.6%)              |
| <b>UCEIS Rectum n (%)</b>          |                       |
| Remission ( $\leq 1$ )             | 209 (68.1%)           |
| Mild (2-4)                         | 62 (20.2%)            |
| Moderate (5-7)                     | 33 (10.7%)            |
| Severe ( $> 7$ )                   | 1 (0.3%)              |
| Missing data*                      | 2 (0.7%)              |
| <b>UCEIS Sigmoid n (%)</b>         |                       |
| Remission ( $\leq 1$ )             | 219 (71.3%)           |
| Mild (2-4)                         | 62 (20.2%)            |
| Moderate (5-7)                     | 21 (6.8%)             |
| Severe ( $> 7$ )                   | 3 (1.0%)              |
| Missing data*                      | 2 (0.7%)              |
| <b>PICASSO Score Rectum n (%)</b>  |                       |
| Remission ( $\leq 3$ )             | 220 (72.7%)           |
| Active ( $> 3$ )                   | 85 (27.7%)            |
| Missing data*                      | 2 (0.7%)              |
| <b>PICASSO Score Sigmoid n (%)</b> |                       |
| Remission ( $\leq 3$ )             | 229 (74.6%)           |
| Active ( $> 3$ )                   | 76 (24.8%)            |
| Missing data*                      | 2 (0.7%)              |

\*Missing data: MES n=5 patients data missing, UCEIS Rectum/Sigmoid n=2 and PICaSSO Rectum/Sigmoid n=2 missing data due to solid stool present precluding the endoscopic scoring of these segments. These patients were not included in the overall analysis.

### **3.6.2 Inter-observer agreement of PICaSSO, UCEIS and MES scores**

Twelve expert endoscopists scored high quality videos independently, and the intra-class correlation coefficient for MES, UCEIS total and PICaSSO total score is shown in

Table 3-3. The one-way ICC between raters was 0.88 (95% CI 0.83-0.92) for the PICaSSO score, as compared with ICC 0.82 (95% CI 0.74-0.88) for MES and 0.84 (95% CI 0.77-0.87) for UCEIS total score. When looking at the vascular sub-groups of the PICaSSO and UCEIS, the PICaSSO vascular component had a good agreement ICC=0.74 (95% CI 0.64-0.83) whereas the vascular component of the UCEIS had a moderate agreement ICC=0.55 (95% CI 0.44-0.64).

The PICaSSO ICC was significantly ( $p<0.01$ ) higher than ICC of either Mayo or UCEIS. In addition, the PICaSSO misclassification rate was significantly lower than that of MES and UCEIS PICaSSO 0.12, UCEIS 0.24, MES 0.29. ( $p<0.01$  PICaSSO Vs MES, UCEIS).

**Table 3-3 Inter-rater correlation of the endoscopic scores**

| <b>Category</b>               | <b>K agreement (95% CI)</b> |
|-------------------------------|-----------------------------|
| <b>Mayo endoscopic score</b>  | <b>0.815 (0.74-0.881)</b>   |
| UCEIS Vascular pattern        | 0.554 (0.44-0.641)          |
| UCEIS Bleeding                | 0.743 (0.651-0.83)          |
| UCEIS Erosion/ulceration      | 0.837 (0.769-0.896)         |
| <b>UCEIS Total Score</b>      | <b>0.837 (0.769-0.896)</b>  |
| PICaSSO Mucosal architecture  | 0.887 (0.836-0.93)          |
| PICaSSO Vascular architecture | 0.737 (0.643-0.826)         |
| <b>PICaSSO Total Score</b>    | <b>0.879 (0.826-0.924)</b>  |

### 3.6.3 Correlation of endoscopic scores with histological scores

The correlations between rectum endoscopic and histological scores are shown in Figure 3-8, which visually shows the distribution of the Pearson correlation coefficients ( $r$ ) across the endoscopic and histological scores in the form of a heatmap. All the endoscopic scores correlated strongly with histological scores. However, PICaSSO correlated more strongly than (a) MES, RHI  $p<0.001$ , ECAP  $p<0.001$ , Nancy  $p<0.01$ , Geboes  $p<0.01$  and Villanacci score  $p<0.01$  and more strongly than (b) UCEIS total score: RHI  $p=0.03$ , ECAP  $p<0.01$ , Nancy  $p<0.001$ , Geboes  $p<0.01$  and Villanacci  $p<0.01$ . This was also the case with a subgroup with the upper range of endoscopy severity scores where PICaSSO correlated significantly better with all histology scores than MES 2/3, UCEIS  $\geq 5$  (all  $p<0.01$ ; Figure 3-9). The corresponding sigmoid data is presented in Figure 3-10 and Figure 3-11.

The PICaSSO mucosal architectural component strongly correlated with RHI ( $r=0.75$  95% CI 0.7-0.8), ECAP ( $r=0.77$  95% CI 0.72-0.81), Nancy index ( $r=0.77$  95% CI 0.72-0.82), Geboes ( $r=0.76$  95% CI 0.71-0.81) and Villanacci score ( $r=0.75$  95% CI 0.69-0.79). The vascular architectural component correlated strongly with RHI ( $r=0.68$  95% CI 0.62-0.74), ECAP ( $r=0.70$  95% CI 0.64-0.76), NHI ( $r=0.70$  0.64-0.76), GS ( $r=0.68$  95% CI 0.62-0.74) and VSS ( $r=0.69$  95% CI 0.62-0.74). The correlation between PICaSSO rectum and sigmoid was 0.66 (95% CI 0.59, 0.72). PICaSSO sigmoid correlated better with all histology scores than MES, but for UCEIS only for the upper end of the range. The different endoscopy scores correlated with each other. PICaSSO score rectum and sigmoid correlated also very strongly with MES rectum 0.81 (95% CI 0.77-0.85); sigmoid 0.82 (95% CI 0.78-0.85) and with UCEIS rectum 0.93 (95% CI 0.91-0.94); sigmoid 0.80 (95% CI 0.76-0.84).

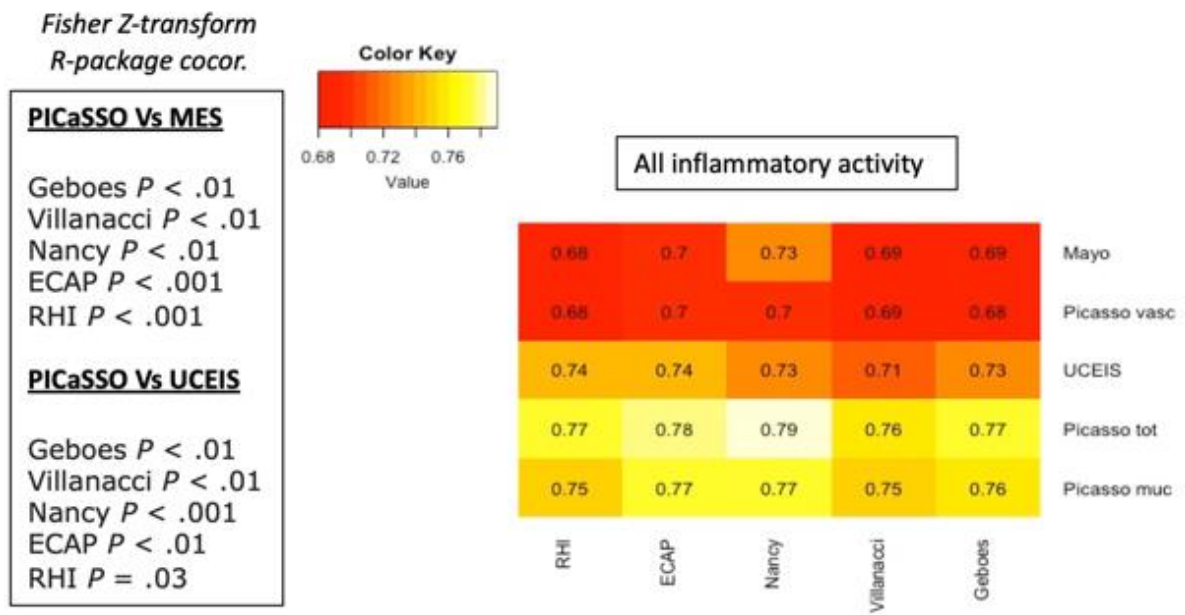


**Figure 3-8 Heatmap demonstrating Pearson correlation in the rectum between endoscopic and histological scores in all patients.**

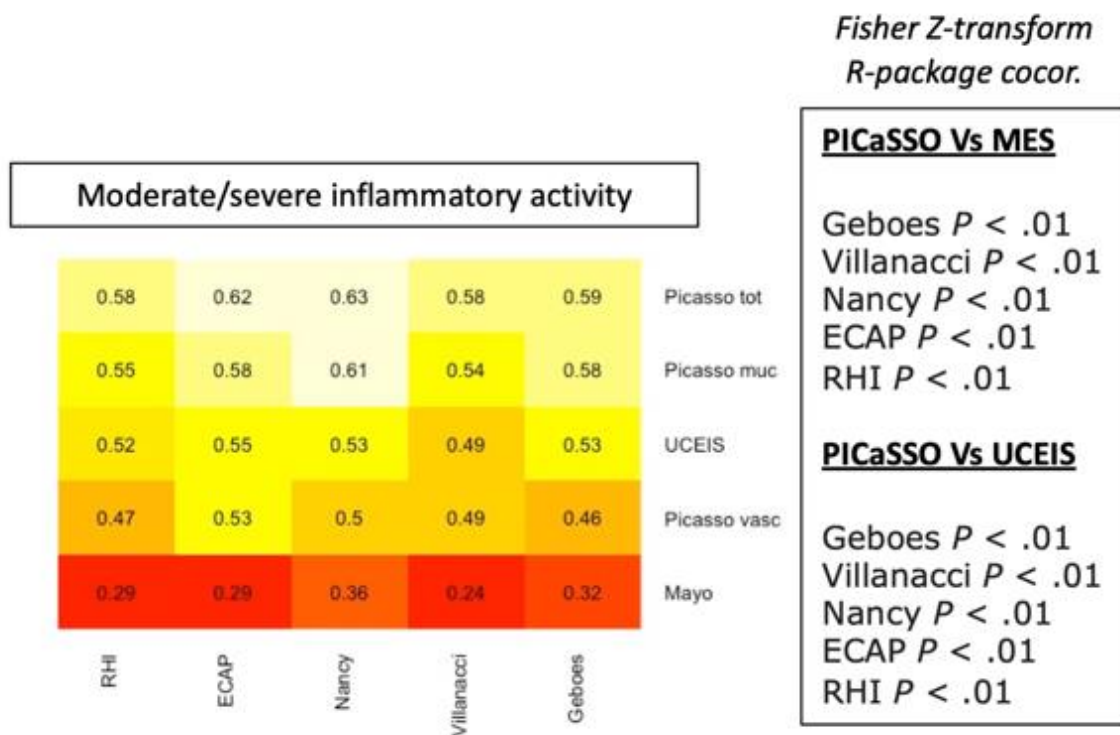


**Figure 3-9 Heatmap demonstrating Pearson correlation in the rectum between endoscopic and histological scores in patients with moderate/severe activity.**

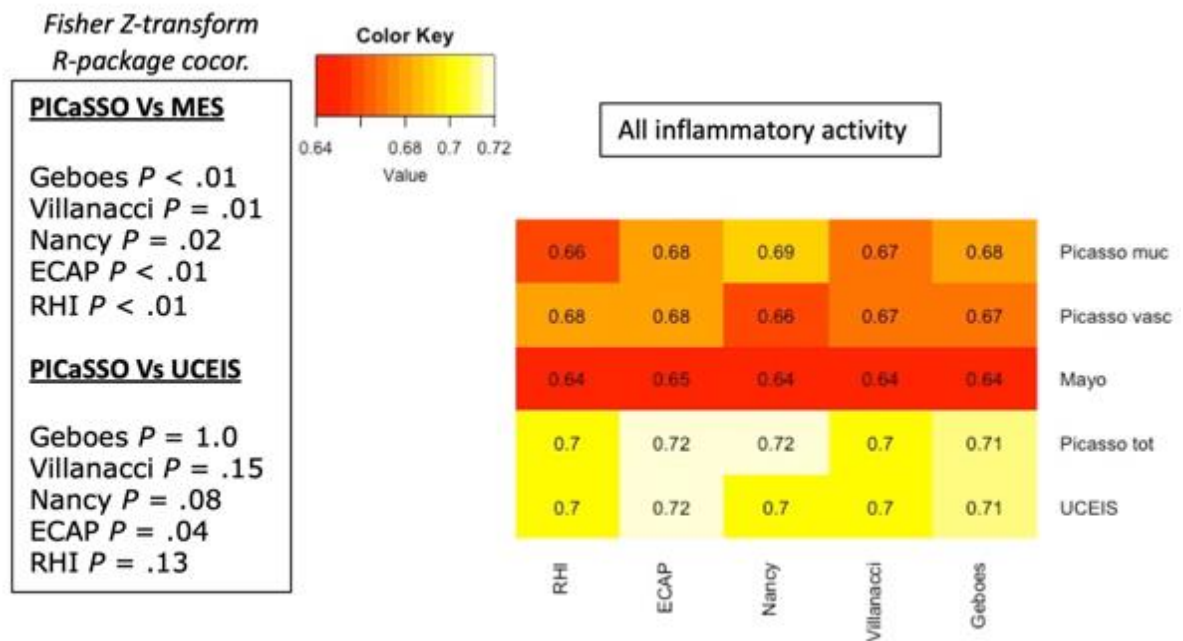


**Figure 3-10 Heatmap demonstrating Pearson correlation in the sigmoid between endoscopic and histological scores in all patients.**

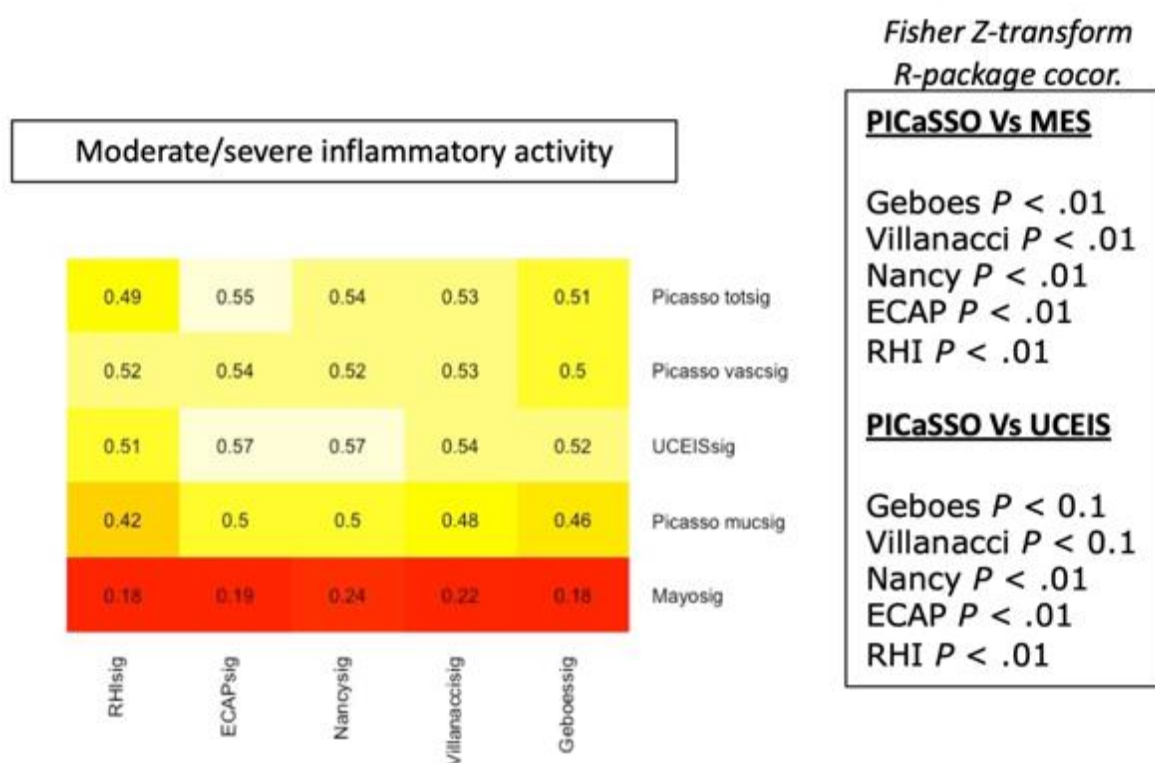


**Figure 3-11 Heatmap demonstrating Pearson correlation in the sigmoid between endoscopic and histological scores in patients with moderate/severe activity.**

### **3.6.4 Relationship between histological remission and endoscopic scores**

95% and 74% of patients with HR by RHI in rectum had PICaSSO  $\leq 3$  and MES =0 (PICaSSO vs MES  $p < 0.00001$ ). 94% and 69% of patients with HR by RHI in sigmoid had PICaSSO  $\leq 3$  and MES=0 (PICaSSO vs MES  $p < 0.00001$ ).

97% and 77% of patients with HR by NHI in rectum had PICaSSO  $\leq 3$  and MES =0 (PICaSSO vs MES  $p < 0.0001$ ). 94% and 73% of patients with HR by NHI in sigmoid had PICaSSO  $\leq 3$  and MES=0 (PICaSSO vs MES  $p < 0.0001$ ). The false negative rate of PICaSSO score for HR by RHI and NHI was significantly lower than the false negative rate for HR determined by MES.

#### *3.6.4.1 The Receiver Operating Characteristics (ROC)*

The Receiver Operating Characteristics (ROC) of the endoscopic scores predicting HR using RHI, Nancy, Geboes, Villanacci and ECAP in the rectum are presented in Figure 3-12. The predictive ability of endoscopic scores at predicting healing in terms of sensitivity, specificity and accuracy are presented in Table 3-4. (Sigmoid data presented in Figure 3-13 and Table 3-5).

##### *3.6.4.1.1 Relationship between RHI $\leq 3$ and endoscopic scores (Figure 3-12a):*

Defining HR as RHI  $\leq 3$  in the absence of neutrophils in the epithelium and lamina propria, a PICaSSO score of  $\leq 3$  was associated with HR with an AUROC of 0.9 (95% CI 0.86-0.94), UCEIS score  $\leq 1$  associated with HR with an AUROC of 0.89 (95% CI 0.85-0.94) and MES 0 was associated with HR with an AUROC of 0.87 (95% CI 0.83-0.91).

##### *3.6.4.1.2 Relationship between ECAP $\leq 4$ and endoscopic scores (Figure 3-12b):*

A PICaSSO total score of  $\leq 3$  was associated with HR defined as ECAP  $\leq 4$  with an AUROC of 0.77 (95% CI 0.72-0.82), UCEIS score of  $\leq 1$  was associated with HR with an AUROC 0.74 (95% CI 0.69-0.80) and MES 0 was associated with HR with an AUROC 0.74 (95% CI 0.69-0.79).

*3.6.4.1.3 Relationship between NHI  $\leq 1$  and endoscopic scores (Figure 3-12c):*

HR defined as a Nancy score of  $\leq 1$  was associated with PICaSSO total score  $\leq 3$  with an AUROC of 0.82 (95% CI 0.77-0.87), UCEIS  $\leq 1$  was associated with HR with an AUROC 0.79 (95% CI 0.73-0.84) and MES 0 was associated with HR with an AUROC 0.81 (95% CI 0.77-0.86).

*3.6.4.1.4 Relationship between GS  $\leq 1$  and endoscopic scores (Figure 3-12d):*

The PICaSSO total score  $\leq 3$  was associated with HR defined by a Geboes score  $\leq 1$  with an AUROC 0.71 (95% CI 0.66-0.76), UCEIS  $\leq 1$  was associated with HR with an AUROC 0.76 (95% CI 0.71-0.81) and MES 0 was associated with HR with an AUROC 0.68 (95% CI 0.63-0.73).

*3.6.4.1.5 Relationship of VSS of 0 with endoscopic scores (Figure 3-12e):*

The PICaSSO total score  $\leq 3$  was associated with HR as defined by Villanacci score of 0, with an AUROC 0.88 (95% CI 0.84-0.93), UCEIS  $\leq 1$  was associated with HR with an AUROC 0.86 (95% CI 0.81-0.91) and MES 0 was associated with HR with an AUROC 0.87 (95% CI 0.82-0.91).

*3.6.4.1.6 Relationship between RHI and Nancy with PICaSSO mucosal and vascular architecture (Figure 3-14):*

A PICaSSO mucosal score of  $\leq 1$  and PICaSSO vascular score of 0 predicted HR as defined by RHI of  $\leq 3$  in the absence of neutrophils in the epithelium and lamina propria, both with an

AUROC 0.88 (95% CI 0.83-0.92). A PICaSSO mucosal score of  $\leq 1$  and PICaSSO vascular score of 0 was associated with HR as defined by NHI of  $\leq 1$  with an AUROC 0.80 (95% CI 0.76-0.86) and 0.80 (95% CI 0.75-0.85) respectively.

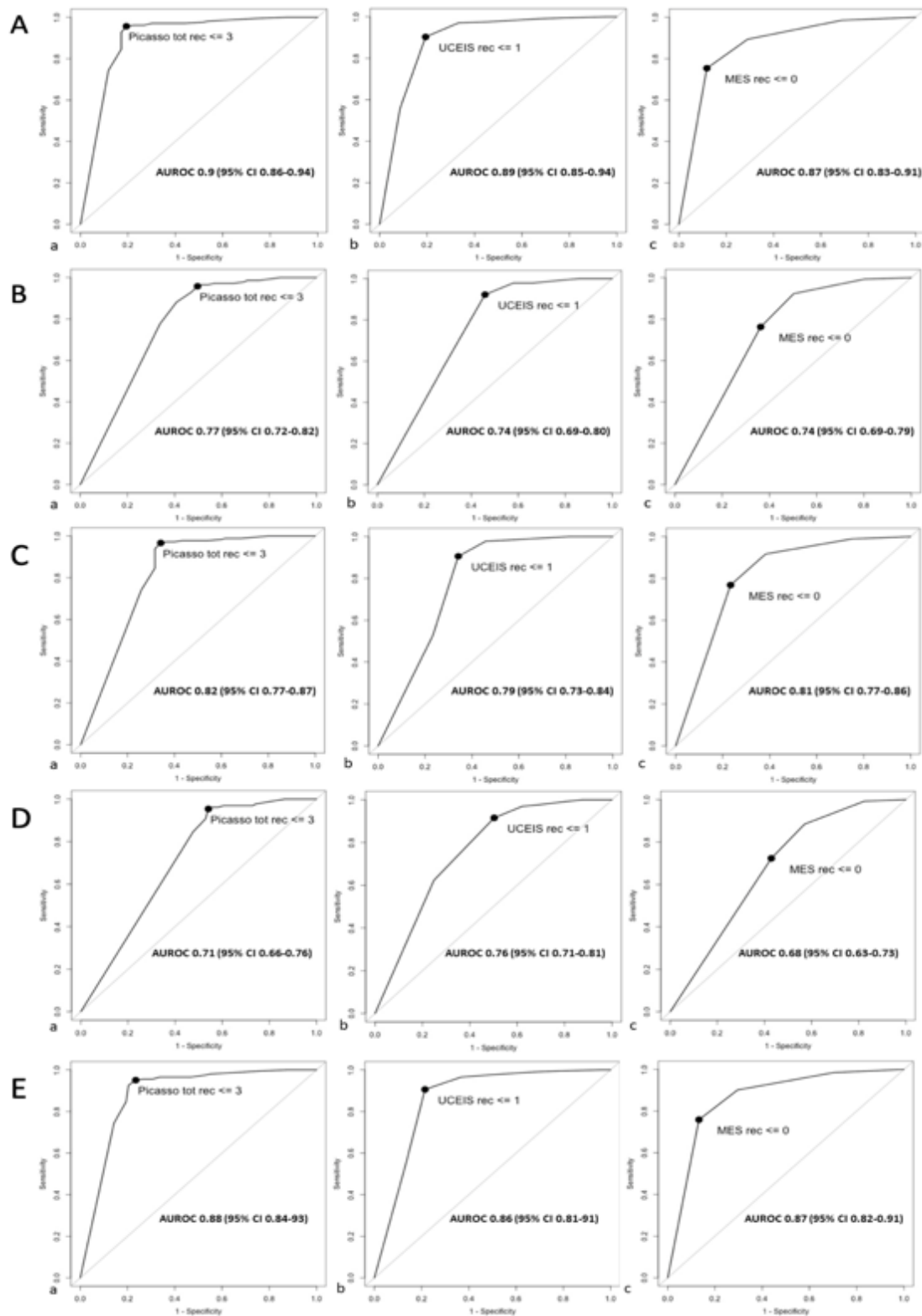


**Figure 3-12 Receiver operator curves for the PICaSSO, UCEIS and MES in predicting histological healing in the rectum**

a) RHI ≤ 3, b) ECAP ≤ 4, c) Nancy ≤ 1, d) Geboes ≤ 1 and e) Villanacci 0

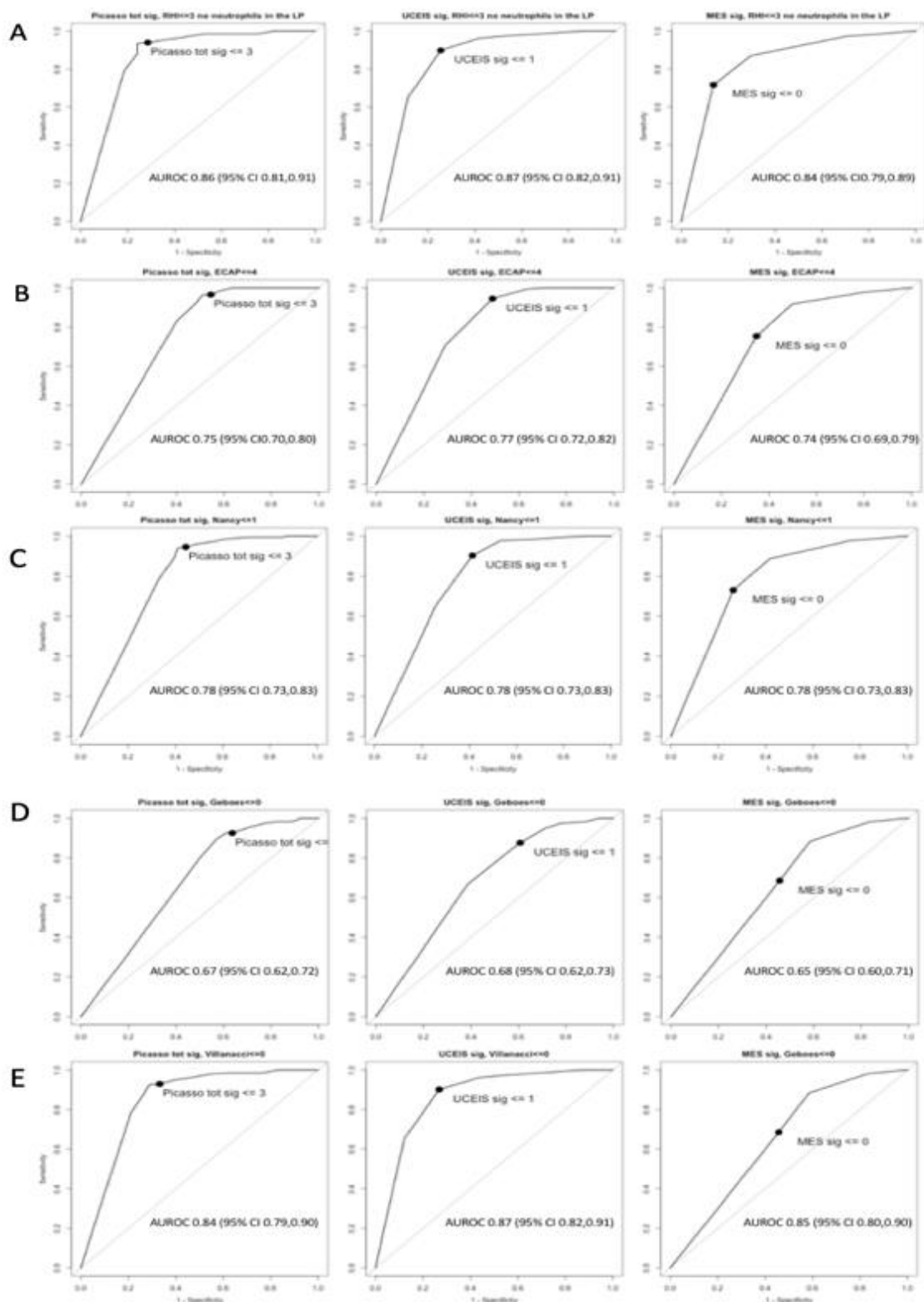


**Figure 3-13 Receiver operator curves for the PICaSSO, UCEIS and MES in predicting histological healing in the sigmoid colon**

a) RHI  $\leq 3$ , b) ECAP  $\leq 4$ , c) Nancy  $\leq 1$ , d) Geboes  $\leq 1$  and e) Villanacci 0

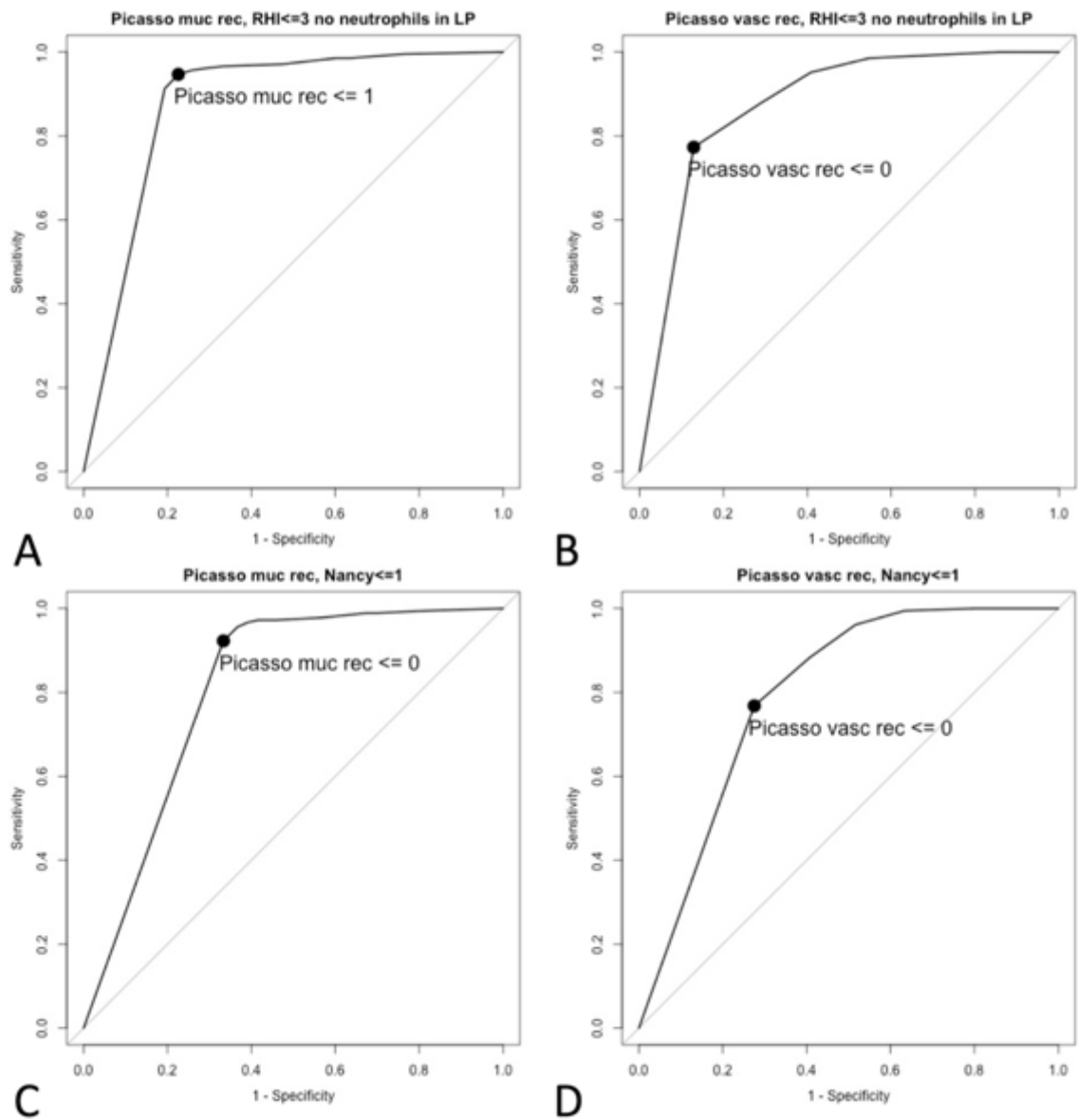


**Figure 3-14 Relationship between RHI and Nancy with PICaSSO mucosal and vascular architecture in the rectum**

a) PICaSSO Mucosal Score vs. RHI  $\leq 3$  b) PICaSSO Vascular Score vs. RHI  $\leq 3$  c) PICaSSO Mucosal Score vs. Nancy  $\leq 1$  d) PICaSSO Vascular Score vs. Nancy  $\leq 1$

**Table 3-4 Sensitivity, specificity and accuracy of endoscopic scores in predicting histological healing**

| <b>Endoscopic score</b>    | <b>RHI <math>\leq 3</math><br/>(95% CI)</b> | <b>ECAP <math>\leq 4</math><br/>(95% CI)</b> | <b>Nancy <math>\leq 1</math><br/>(95% CI)</b> | <b>Geboes <math>\leq 1</math><br/>(95% CI)</b> | <b>Villanacci 0<br/>(95% CI)</b> |
|----------------------------|---------------------------------------------|----------------------------------------------|-----------------------------------------------|------------------------------------------------|----------------------------------|
| <b>PICaSSO Total Score</b> |                                             |                                              |                                               |                                                |                                  |
| Sensitivity                | 0.96 (0.75-0.98)                            | 0.96 (0.87-0.99)                             | 0.97 (0.73-0.99)                              | 0.95 (0.81-0.99)                               | 0.95 (0.77-0.98)                 |
| Specificity                | 0.81 (0.46-0.88)                            | 0.50 (0.28-0.59)                             | 0.66 (0.38-0.74)                              | 0.46 (0.23-0.53)                               | 0.76 (0.46-0.85)                 |
| Accuracy                   | 0.91 (0.80-0.93)                            | 0.72 (0.60-0.76)                             | 0.84 (0.73-0.88)                              | 0.67 (0.54-0.71)                               | 0.89 (0.79-0.92)                 |
| <b>MES</b>                 |                                             |                                              |                                               |                                                |                                  |
| Sensitivity                | 0.75 (0.48-0.83)                            | 0.76 (0.61-0.86)                             | 0.77 (0.57-0.86)                              | 0.72 (0.59-0.83)                               | 0.76 (0.49-0.84)                 |
| Specificity                | 0.88 (0.79-0.94)                            | 0.64 (0.55-0.72)                             | 0.77 (0.67-0.84)                              | 0.57 (0.48-0.65)                               | 0.87 (0.78-0.93)                 |
| Accuracy                   | 0.79 (0.77-0.81)                            | 0.70 (0.65-0.74)                             | 0.77 (0.73-0.80)                              | 0.64 (0.59-0.68)                               | 0.79 (0.76-0.81)                 |
| <b>UCEIS Total Score</b>   |                                             |                                              |                                               |                                                |                                  |
| Sensitivity                | 0.90 (0.70-0.95)                            | 0.92 (0.78-0.97)                             | 0.91 (0.63-0.97)                              | 0.92 (0.83-0.97)                               | 0.91 (0.60-0.95)                 |
| Specificity                | 0.81 (0.69-0.88)                            | 0.54 (0.44-0.62)                             | 0.66 (0.55-0.74)                              | 0.50 (0.39-0.58)                               | 0.79 (0.67-0.86)                 |
| Accuracy                   | 0.87 (0.84-0.90)                            | 0.72 (0.67-0.76)                             | 0.81 (0.77-0.84)                              | 0.68 (0.62-0.73)                               | 0.87 (0.83-0.89)                 |

**Table 3-5 Sensitivity, specificity and accuracy of endoscopic scores in predicting histological healing in the sigmoid colon**

| <b>Endoscopic score</b>    | <b>RHI <math>\leq 3</math><br/>(95% CI)</b> | <b>ECAP <math>\leq 4</math><br/>(95% CI)</b> | <b>Nancy <math>\leq 1</math><br/>(95% CI)</b> | <b>Geboes <math>\leq 1</math><br/>(95% CI)</b> | <b>Villanacci 0<br/>(95% CI)</b> |
|----------------------------|---------------------------------------------|----------------------------------------------|-----------------------------------------------|------------------------------------------------|----------------------------------|
| <b>PICaSSO Total Score</b> |                                             |                                              |                                               |                                                |                                  |
| Sensitivity                | 0.94 (0.81-0.97)                            | 0.97 (0.90-1.0)                              | 0.95 (0.82-0.98)                              | 0.93 (0.86-0.98)                               | 0.93 (0.84-0.97)                 |
| Specificity                | 0.71 (0.55-0.83)                            | 0.45 (0.37-0.57)                             | 0.56 (0.41-0.67)                              | 0.36 (0.25-0.48)                               | 0.67 (0.53-0.79)                 |
| Accuracy                   | 0.86 (0.83-0.91)                            | 0.72 (0.66-0.76)                             | 0.80 (0.74-0.84)                              | 0.61 (0.52-0.66)                               | 0.85 (0.81-0.89)                 |
| <b>MES</b>                 |                                             |                                              |                                               |                                                |                                  |
| Sensitivity                | 0.72 (0.46-0.80)                            | 0.76 (0.60-0.85)                             | 0.73 (0.55-0.83)                              | 0.69 (0.57-0.81)                               | 0.73 (0.46-0.80)                 |
| Specificity                | 0.86 (0.77-0.93)                            | 0.65 (0.56-0.73)                             | 0.74 (0.64-0.81)                              | 0.54 (0.46-0.63)                               | 0.87 (0.77-0.93)                 |
| Accuracy                   | 0.76 (0.73-0.78)                            | 0.70 (0.66-0.74)                             | 0.73 (0.70-0.76)                              | 0.6 (0.55-0.65)                                | 0.77 (0.74-0.79)                 |
| <b>UCEIS Total Score</b>   |                                             |                                              |                                               |                                                |                                  |
| Sensitivity                | 0.90 (0.76-0.95)                            | 0.95 (0.85-0.98)                             | 0.90 (0.77-0.96)                              | 0.88 (0.79-0.95)                               | 0.90 (0.78-0.95)                 |
| Specificity                | 0.74 (0.63-0.83)                            | 0.51 (0.40-0.59)                             | 0.59 (0.49-0.68)                              | 0.39 (0.3-0.49)                                | 0.73 (0.60-0.82)                 |
| Accuracy                   | 0.85 (0.82-0.88)                            | 0.73 (0.67-0.77)                             | 0.78 (0.74-0.82)                              | 0.59 (0.53-0.64)                               | 0.85 (0.81-0.88)                 |

### 3.6.5 Follow up results and prediction of adverse outcomes

A total of 289 and 270 patients completed 6 and 12 months follow up respectively following baseline survival benefit of clinical remission without specified clinical outcomes with a PICaSSO colonoscopy. A total of 66 (22.8%) and 78 (28.9%) patients had a specified clinical outcome at 6 and 12 months respectively. Figure 3-15 and Figure 3-16 presents the Cox proportional hazards curves for the PICaSSO, UCEIS and MES scores (rectum) at 6 and 12 months, respectively. There was a score of  $\leq 3$  vs  $> 3$ , UCEIS  $\leq 1$  vs.  $> 1$  and MES 0 vs. MES  $\geq 1$  (Figure 3-15 and Figure 3-16). MES 0 had a significantly better outcome over MES 1 at both 6 and 12 months (Figure 3-15 and Figure 3-16). In addition, Cox proportional hazard curve analysis of PICaSSO  $\leq 3$  vs PICaSSO  $> 3 \leq 8$  demonstrated significant survival benefit without specified clinical outcomes at 6 and 12 months (Figure 3-15 and Figure 3-16). The sigmoid data is shown in Figure 3-17.

A further sub-analysis to assess the ability of the PICaSSO mucosal and PICaSSO vascular architectural scores separately in the rectum to predict specified clinical outcomes was performed using Cox proportional hazards curves (Figure 3-18). There was a significant increased survival benefit without specified clinical outcomes with a vascular score of 0 vs.  $\geq 1$  and mucosal score of  $< 2$  vs.  $\geq 2$  at 6 and 12 months ( $P < 0.0001$ ). The sigmoid data is shown in Figure 3-19. Based on the bootstrap sample the increment in specified clinical outcomes (event rates) of MES=1 over MES=0 was significantly greater than increment of events of PICaSSO 4 – 8 over PICaSSO  $\leq 3$  at 6 months follow-up ( $p < 0.05$  for both rectum and sigmoid). The difference between MES 0 and 1 is therefore significantly bigger than PICaSSO  $\leq 3$  and 4-8.

In patients with PICaSSO  $\leq 3$  (rectum), the specified clinical outcomes at 6 and 12 months were 12.7% and 16.8% respectively. This was not significantly different from PICaSSO =0 (10.8% and 14.8% at 6 and 12 months). In patients with RHI  $\leq 3$  and no neutrophils the specified clinical outcomes at 6 and 12 months were 11.7% and 15.9% respectively. Specified clinical outcome rates for NHI  $\leq 1$  at 6 and 12 months were similar (12.2% and 17.1%). Therefore, PICaSSO and histology specified clinical outcome rates after remission were similar.

The weightage of each component of the PICaSSO score (mucosal and vascular) was also analysed by means of multivariate logistic regression to predict adverse outcome at 12-month follow-up, which is shown in Table 3-6. The mucosal component I-B (patchy micro-erosion) and vascular component II-B (intra-mucosal bleeding, crowded or tortuous superficial vessels with dilatation) contributed relatively less to the regression score.

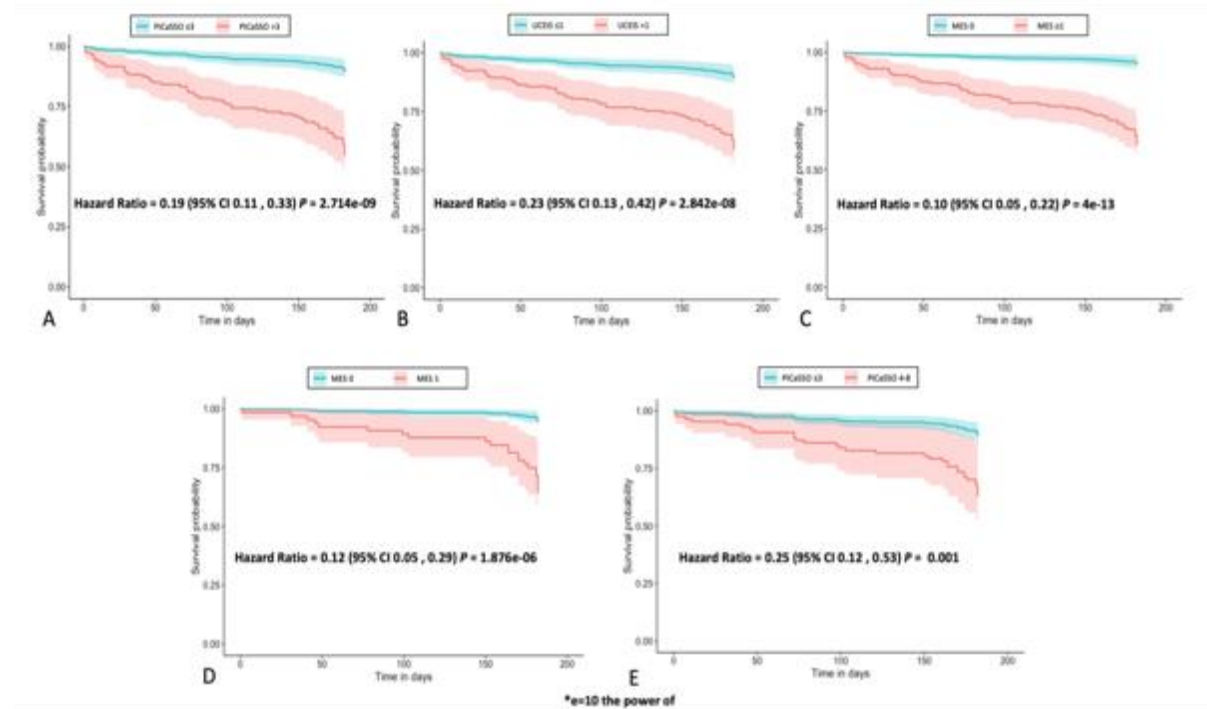


**Figure 3-15 Cox proportional hazard curves in predicting likelihood of specified clinical outcomes at 6 months**

a) PICaSSO  $\leq 3$  (blue) vs.  $>3$  (red) b) UCEIS  $\leq 1$  (blue) vs.  $>1$  (red) C) MES 0 vs.  $\geq 1$  d) MES 0 (blue) vs. MES 1 (red) and e) PICaSSO  $\leq 3$  (blue) vs. 4-8 (red)

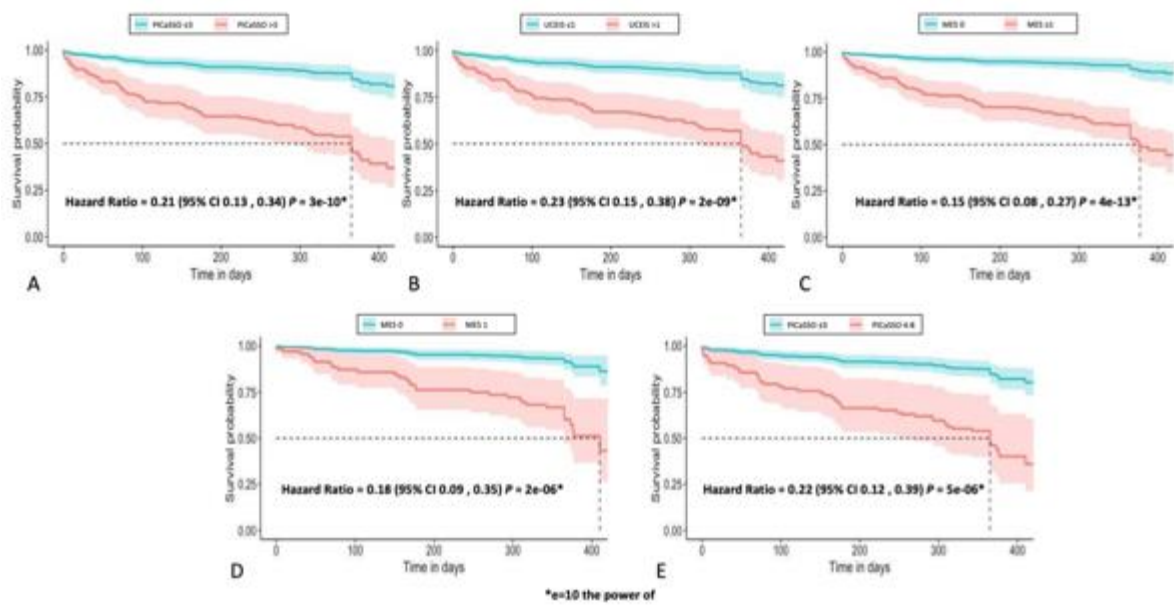


**Figure 3-16 Cox proportional hazard curves in predicting likelihood of specified clinical outcomes at 12 months**

a) PICaSSO  $\leq 3$  (blue) vs.  $> 3$  (red) b) UCEIS  $\leq 1$  (blue) vs.  $> 1$  (red) C) MES 0 vs.  $\geq 1$  d)

MES 0 (blue) vs. MES 1 (red) and e) PICaSSO  $\leq 3$  (blue) vs.  $4 \leq \text{PICaSSO} \leq 8$

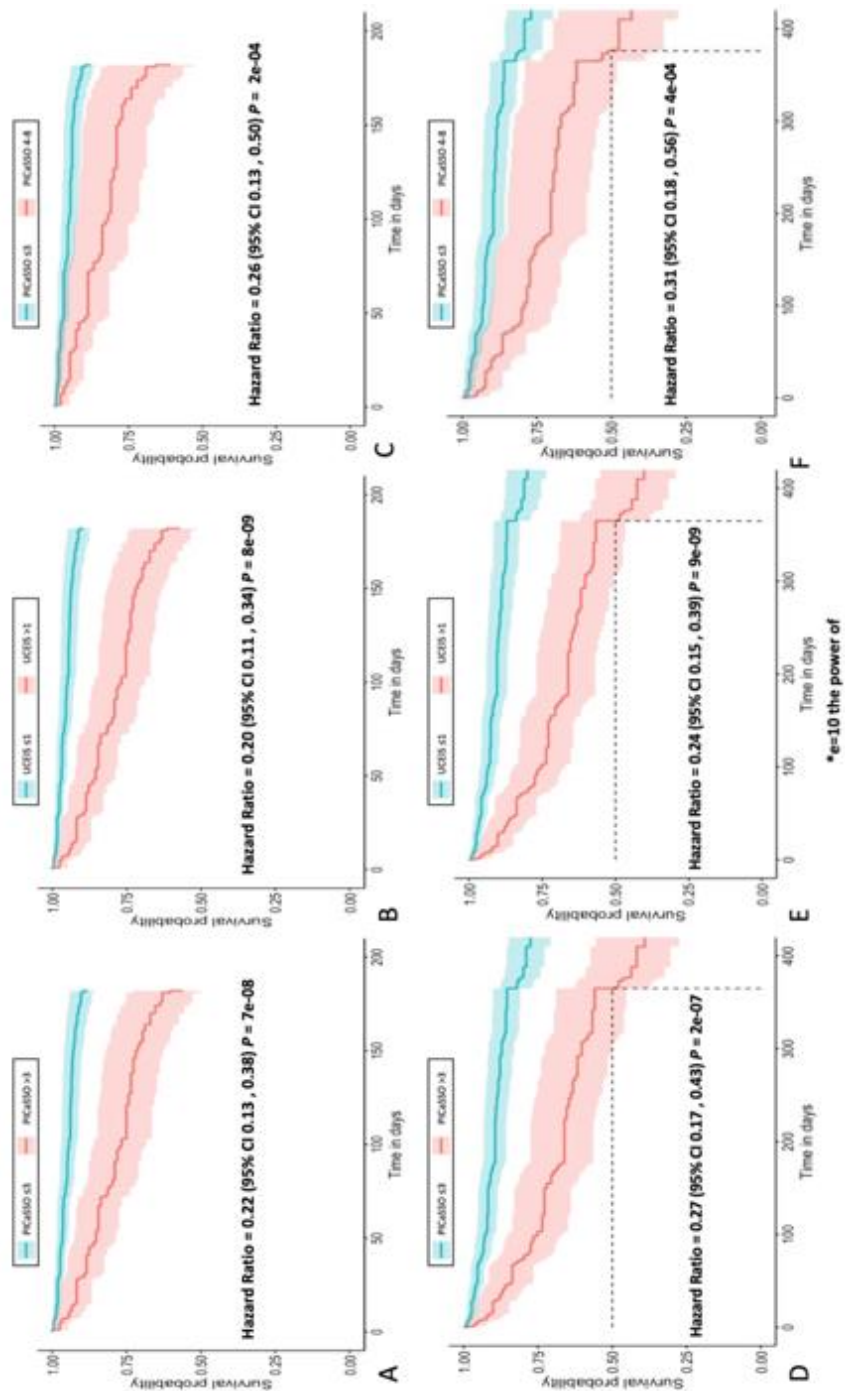


**Figure 3-17 Cox proportional hazard curves in predicting likelihood of specified clinical outcomes at 6 months**

a) PICaSSO  $\leq 3$  (blue) vs.  $>3$  (red) b) UCEIS  $\leq 1$  (blue) vs.  $>1$  (red) c) PICaSSO  $\leq 3$  (blue) vs. 4--8 (red); and 12 months d) PICaSSO  $\leq 3$  (blue) vs.  $>3$  (red) e) UCEIS  $\leq 1$  (blue) vs.  $>1$  (red) f) PICaSSO  $\leq 3$  (blue) vs. 4-8 (red) calculated in the sigmoid colon

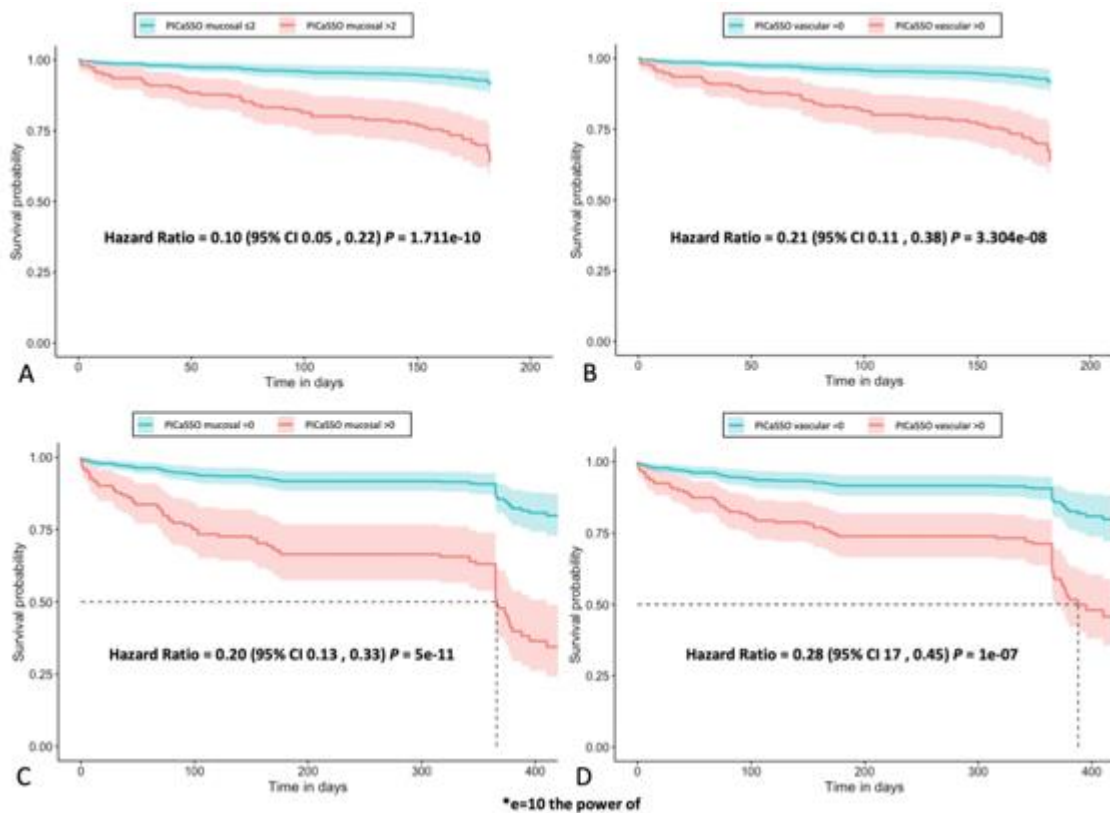


**Figure 3-18 Cox proportional hazard curves of PICaSSO mucosal score and PICaSSO vascular architectural scores predicting specified clinical outcomes**

Six months (A and B) and twelve months (C and D) in the rectum

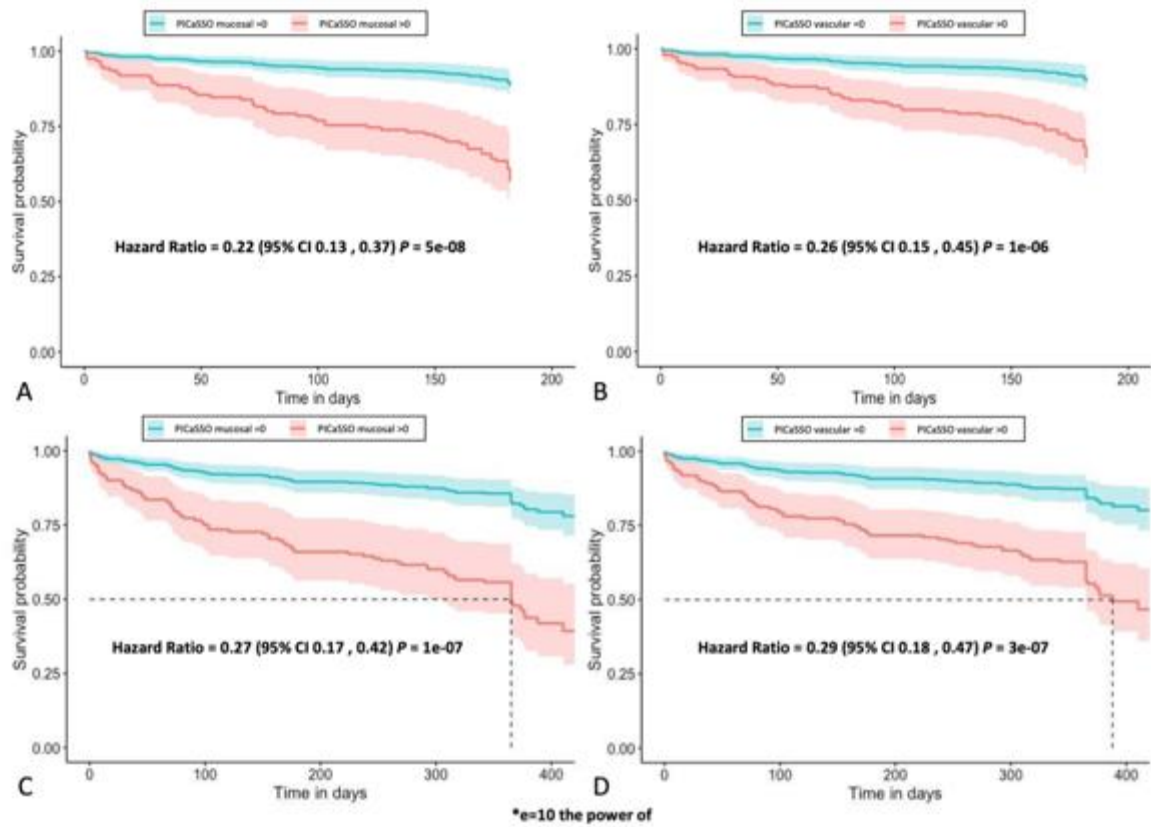


**Figure 3-19 Cox proportional hazard curves of PICaSSO mucosal score and PICaSSO vascular architectural scores predicting clinical outcomes**

Six (A and B) and twelve months (C and D) in the sigmoid colon

**Table 3-6 Multivariate regression analysis of PICaSSO components in predicting adverse outcome at 12 months**

| <b>PICaSSO Mucosal Architecture</b>             | <b>Multivariate regression (95% CI)</b> | <b>PICaSSO Vascular Architecture</b>                       | <b>Multivariate regression (95% CI)</b> |
|-------------------------------------------------|-----------------------------------------|------------------------------------------------------------|-----------------------------------------|
| <b>0-No mucosal defect</b>                      |                                         | <b>0- Vessels without dilatation</b>                       |                                         |
| A.Continuous/regular crypts                     | 0.19 (0.11-0.35)                        | A. Roundish following crypt architecture                   | 0.18 (0.11-0.31)                        |
| B.Crypts not visible (scar)                     |                                         | B. Vessels not visible (scar)                              |                                         |
| C.Discontinuous and/or dilated/elongated crypts |                                         | C. Spare (deep) vessels without dilatation                 |                                         |
| <b>I: Microerosion or cryptal abscess</b>       |                                         | <b>I: Vessels with dilatation</b>                          |                                         |
| A- Discrete                                     | 0.43 (0.12-1.59)                        | A- Roundish with dilatation                                | 0.68 (0.32-1.41)                        |
| B- Patchy                                       | 1.00 (0.16-6.26)                        | B- Crowded or tortuous superficial vessels with dilatation | 0.61 (0.27-1.37)                        |
| C- Diffuse                                      | 0.50 (0.05-4.61)                        | <b>II: Intramucosal bleeding</b>                           |                                         |
| <b>II: Erosions size &lt;5mm</b>                |                                         | A- Roundish with dilatation                                | 1.13 (0.47-2.68)                        |
| A- Discrete                                     | 3.99 (0.52-30.67)                       | B- Crowded or tortuous superficial vessels with dilatation | 1.02 (0.42-2.46)                        |
| B-Patchy                                        | 2.66 (0.73-9.68)                        | <b>III: Luminal bleeding</b>                               |                                         |
| C-Diffuse                                       | 1.16 (0.39-3.48)                        | A- Roundish with dilatation                                | 1.27 (0.38-4.21)                        |
| <b>III: Ulcerations size &gt;5mm</b>            |                                         | B- Crowded or tortuous superficial vessels with dilatation | 9.14 (1.47-56.90)                       |
| A- Discrete                                     | 0.50 (0.05-4.61)                        |                                                            |                                         |
| B- Patchy                                       | 1.40 (0.45-4.38)                        |                                                            |                                         |
| C- Diffuse                                      | 2.80 (0.99-7.90)                        |                                                            |                                         |

### 3.7 Discussion

An endoscopic scoring system that accurately encompasses the elements of mucosal healing, namely ER and HR is needed. This study represents for the first time, a real-life prospective multicentre, multirater study of the PICaSSO score in UC for assessment of ER and HR. A novel aspect to the study is that biopsies were taken at the site of endoscopic scoring with blinded histological scoring providing precise histological data from the endoscopic findings. It is the largest prospective validation study of any endoscopic score in UC. The prospective multi-centre design exhibits the reproducibility of PICaSSO across many centres. The inclusion of five histological scores, many of which are used in clinical trials, also provided robust information on the performance of the PICaSSO score.

The results show that PICaSSO scores correlate strongly with multiple histological indices, and the correlation coefficients were statistically superior to both MES (rectum and sigmoid) and UCEIS (rectum only); this also applied to the upper end of the endoscopy disease spectrum. We have also shown that the PICaSSO score has a high inter-rater agreement, which is consistent with a previous study (Iacucci, Daperno et al. 2017). Some of the collaborators in the present study were not involved in the conception of the PICaSSO score, and therefore the fact that through limited training a high inter-rater agreement can be achieved, raises the possibility of the PICaSSO score used globally. The ICC of PICaSSO was significantly better than MES or UCEIS, with less misclassification rate. It has been previously shown that the PICaSSO score could be reproduced in less experienced and trainee endoscopists after a short training module (Trivedi, Kiesslich et al. 2018).

Current definitions of MH using existing endoscopic scores have limitations, not least that they do not truly reflect ER (Iacucci, Fort Gasia et al. 2015) and HR (Bessissow, Lemmens et al. 2012, Bryant, Winer et al. 2014, Kanazawa, Takahashi et al. 2019). HR is a predictor of favourable patients' outcomes, (Park, Abdi et al. 2016, Calafat, Lobaton et al. 2017, Christensen, Hanauer et al. 2017) and there is a current unmet need of an endoscopic score that can accurately predict HR accurately. This study has demonstrated that the PICaSSO score predicts HR accurately irrespective of which of the five histological scores are used. A PICaSSO score of  $\leq 3$ , not only describes endoscopically quiescent disease but is also associated with HR with a high degree of accuracy and low false negative rate. This confirms recent findings with a single operator using multiple endoscopic platforms (Iacucci, Cannatelli et al. 2020), with clinical and fecal calprotectin endpoints (Cannatelli, Bazarova et al. 2020). Therefore, the PICaSSO score can be used to define MH in a more robust manner than alternative scores reinforcing the concept that endoscopic assessment is becoming closer to histology using current generation endoscopes. The corresponding cut-offs for MES was 0 and UCEIS  $\leq 1$  which, in contrast to several results in the past (Limdi, Picco et al. 2020) also predicted HR well. This may be the consequence of assessment by HD, experienced operators and training prior to start of the study.

The most commonly used endoscopic scores of inflammatory activity in UC are the MES, and UCEIS; however, these scores were developed in an era before high definition and VEC. Studies have shown the problem with MES 0 as a definition of MH since inflammatory activity is frequently present (Bessissow, Lemmens et al. 2012, Iacucci, Fort Gasia et al. 2015, Iacucci, Cannatelli et al. 2020). These results of the present study and previous studies with PICaSSO suggest that, in the era of HD and VEC, the discrepancy between endoscopic

and histological grading of inflammatory activity has become much smaller (Iacucci, Fort Gasia et al. 2015, Iacucci, Kiesslich et al. 2017, Iacucci, Cannatelli et al. 2020). The UCEIS incorporates three components (vascular pattern, bleeding, erosions/ulceration), which was developed using endoscopic videos (Travis, Schnell et al. 2012). The interinvestigator reliability of the score was strong and a moderate interinvestigator *kappa* agreement was found (Travis, Schnell et al. 2013). The UCEIS correlated strongly with the Nancy index and RHI, however this study was limited by its single centre, single and single pathologist design (Irani, Wang et al. 2018). The score also does not include endoscopic features that are now visible with VEC using latest generation endoscopes and has yet to be widely adopted in clinical trials.

VEC enables endoscopists to identify subtle changes reflecting histological activity, which is a technology that is already widely available at a push of a button, and the PICaSSO score includes features that reflect MH accurately as well as active inflammation. It can help to define ER and HR which guides UC management (Cannatelli, Bazarova et al. 2020). This will assist decision-making on when to escalate treatment when inflammation is present and in de-escalating therapy, which is an area of concern for many clinicians. All current endoscopy platforms are equipped with VEC and are routinely used for better definition of GI cancer (Iacucci, Furfaro et al. 2019). PICaSSO score is compatible with other VEC platforms such as NBI and BLI (Trivedi, Kiesslich et al. 2018, Cannatelli, Bazarova et al. 2020). VEC can help detect subtle mucosal and vascular changes and these are reflected in the PICaSSO score (Iacucci, Daperno et al. 2017). The mucosal and vascular items of total PICaSSO score appeared to contribute near-equal weight to the final score by multiple regression analysis. It would be important to conduct a future study to determine responsiveness of PICaSSO to

specific therapeutic intervention to understand the overall value of individual items. This will provide insight into future simplification of the PICASSO score in clinical practice.

For an endoscopic score of inflammatory activity to be used in everyday clinical practice it must reflect histological activity and the PICaSSO score correlates strongly with histology. MH is important in guiding management of patients with UC in terms of medical therapy changes, discontinuation of treatment and predicting risk of relapse and adverse outcome. Achieving MH positively as defined by PICaSSO may give more confidence when stopping therapy (Doherty, Katsanos et al. 2017). A score cut-off to indicate likelihood of long-term outcome of remission is also desirable. The present study has shown a significant favorable outcome of PICaSSO score  $\leq 3$  over higher scores at both 6 and 12 months. In addition, there is an outcome benefit of patients with PICaSSO  $\leq 3$  vs. PICaSSO 4-8, which allows a greater bandwidth when defining ER. Clinicians may closely monitor patients with PICaSSO  $> 3$  to promptly treat the first sign of relapse, and/or increase medical therapy to achieve MH. The mucosal and the vascular components of the PICaSSO score were independently able to predict outcomes on follow-up and therefore both have significant role in the assessment of inflammatory activity in UC.

MES and UCEIS performed very well in our study as well as these were also assessed using HD imaging and scored by experienced operators providing detailed mucosal, vascular images with targeted biopsies. The main strengths in favor of the PICaSSO score are the following: strong correlation with all histology scores (better than MES and UCEIS), accurate correlation with HR, excellent reliability as demonstrated by its consistently strong inter-rater correlation with lowest misclassification rate, low false negative rate of PICaSSO  $\leq 3$  for histologic remission (better than MES and UCEIS) and it can predict

outcomes accurately. The biopsies were targeted contributing to strong endoscopy-histology correlations (Cannatelli, Bazarova et al. 2020, Iacucci, Cannatelli et al. 2020). Specified clinical outcome rates were similar for patients in remission by PICaSSO and by RHI or NHI. The endoscopy scores performed similarly for outcomes. Of note, specified clinical outcomes rates had a larger increase between MES 0 and 1 compared with PICaSSO  $\leq 3$  and 4-8, a limitation of a 4-point MES and the challenge of considering MES 0 and MES 1 as endoscopic MH.

A summary of the characteristics of PICaSSO against other endoscopic scores is presented in Table 3-7. A future use of the PICaSSO score could be in clinical trials, whereby precise assessment of inflammation and MH from endoscopic videos in central read-out (Feagan, Sandborn et al. 2013) is vital in determining response to novel treatments. Unlike MES which is a composite score with several descriptors for each score item, PICaSSO mucosal and vascular scoring items each describe a single feature. These results show that the PICaSSO score correlates superiorly with histology over MES and UCEIS as well as having the advantage of predicting MH accurately. For this reason, the PICaSSO score could be considered as an endpoint for clinical trials and as a central read-out tool and may be adapted conveniently to future development of artificial intelligence.

One limitation of the study is the inclusion of endoscopists experienced in optical diagnosis. On one hand this could be seen as a strength as it allows a better understanding of the possibility of PICaSSO and VEC, but it cannot be assumed the same levels of performance would be replicated amongst non-experts and those not accustomed to VEC in a real-life setting. However, it has been shown that, with adequate training, the level of performance can be reproduced in non-experts and trainees (Trivedi, Kiesslich et al. 2018). Since endoscopists completed scoring having examined the rectum and sigmoid with high definition white light

and VEC, it is possible that the diagnostic performances and histological correlation for the MES and UCEIS were artificially high. All raters underwent refresher training for PICaSSO using a structured and previously validated training module and this may have improved performance also for MES and UCEIS. In this study we did not follow-up patients using patient reported outcomes (as is the case in other studies (Barreiro-de Acosta, Vallejo et al. 2016, Bryant, Burger et al. 2016) as symptoms do not relate well to endoscopy and histology. In the case of 6 and 12 months follow-up the required sample size has not been reached and even though initially 307 patients were recruited, only 289 and 270 completed the respective follow-up periods.

In conclusion, this large real life prospective multicentre study shows the first VEC PICaSSO score can predict ER and HR accurately and has a strong correlation across the range of histological scores outperforming the MES and UCEIS. The PICaSSO score should be considered in clinical practice and clinical trials to grade endoscopic inflammatory activity and define MH and further prospective studies in the context of specific therapies are necessary.

The section above is taken from the following publication:

Iacucci M, **Smith SCL**, Bazarova A, Shivaji UN, Bhandari P, Cannatelli R, Daperno M, Ferraz J, Goetz M, Gui X, Hayee B, Hertogh G, Lazarev M, Li J, Nardone OM, Parra-Blanco A, Pastorelli L, Panaccione R, Occhipinti V, Rath T, Tontini GE, Vieth M, Villanacci V, Zardo D, Bisschops R, Kiesslich R, Ghosh S. An International Multicentre Real-Life Prospective Study of Electronic Chromoendoscopy Score PICaSSO in Ulcerative Colitis. *Gastroenterology* 2021; 160: 1558-1569

**Table 3-7 Comparison of commonly used endoscopic scoring systems in UC**

| Score          | Endoscopic findings                                                                                | Apply to HD WLE? | Apply to VEC? | Describe mucosal healing? | Validated in an artificial setting? | Validated in a clinical setting? | Training intervention validated? | Strengths                                                                                                                                                                                   | Weaknesses                                                                                                                                                                          |
|----------------|----------------------------------------------------------------------------------------------------|------------------|---------------|---------------------------|-------------------------------------|----------------------------------|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>PICaSSO</b> | Mucosal architecture and vascular architecture                                                     | Y                | Y             | Y                         | Y                                   | Y                                | Y                                | Validated<br>Defines endoscopic remission<br>Strong correlation with histology<br>Strong inter-rater correlation<br>Establish cut-off for endoscopic remission<br>Predicts clinical outcome | Not reproduced in community gastroenterologists                                                                                                                                     |
| <b>MES</b>     | Erythema, vascular pattern, erosions, friability, bleeding and ulceration                          | Y                | N             | N                         | N                                   | N                                | N                                | Easy to use<br>Used in clinical trials                                                                                                                                                      | Partially validated<br>No definition of endoscopic remission<br>Subjective findings<br>Developed prior to HD and VEC                                                                |
| <b>UCEIS</b>   | Erosions and ulceration (three levels), bleeding (four levels) and vascular pattern (three levels) | Y                | N             | N                         | Y                                   | Y                                | N                                | Easy to use<br>Used in clinical trials<br>Validated<br>Strong correlation with histology                                                                                                    | No definition of endoscopic remission<br>Developed prior to HD and VEC<br>No definition of mild, moderate and severe thresholds                                                     |
| <b>UCCIS</b>   | Vascular pattern, granularity, ulceration and bleeding/friability                                  | Y                | N             | N                         | N                                   | Y                                | N                                | Easy to use<br>Strong correlation with clinical activity                                                                                                                                    | Partially validated<br>No definition of endoscopic remission<br>Developed prior to HD and VEC<br>No definition of mild, moderate and severe thresholds<br>Requires full colonoscopy |

## 4 SECTION TWO: RAMAN SPECTROSCOPY IN INFLAMMATORY BOWEL DISEASE

### 4.1 Overview

In this chapter I present Raman Spectroscopy, what it is and how it can be utilised in clinical practice. I then discuss the current existing evidence for its use in IBD. I then present my research into Raman Spectroscopy and mucosal healing in IBD and the spectrum changes following successful treatment of IBD.

### 4.2 Background to Raman Spectroscopy and its application in IBD

#### 4.2.1 What is Raman Spectroscopy?

Given the limitations of existing technology and despite patients displaying endoscopic and histological healing, patients still suffer short-term relapse, which has been highlighted in Chapter 3. Therefore, there is increasing interest in alternative advanced imaging technologies, to give greater information on the mucosal status of the patient. Raman Spectroscopy is an example of one such technology.

Raman Spectroscopy was first discovered in 1928 (Raman 1928) and describes the phenomenon when light strikes a tissue causing vibrational energy change of individual molecules. When molecules return from this excited state, there is scattering of inelastic light, and the frequency difference between incident and scattered light can be detected and measured. This is detected as shifts in light wavelength due to the energy transfer (Committee 2013). This is displayed as a spectrum and is highly specific for individual molecules thereby

giving a “biomolecular fingerprint”, which reveals the biochemistry make-up of tissues (Robles, Singh et al. 2015).

Since Raman Spectroscopy can be used to give an immediate interpretation of the “make-up” of biochemical tissue, it can potentially give a real-time optical diagnosis. Healthy tissue has a different molecular composition compared with diseased tissue, and therefore will have different vibrations and different shifts in light, which is visualised as differences in spectra (Committee 2013).

The shifts that occur are rare meaning the resultant signal is considered weak (Committee 2013). Since the advent of laser technology, the development of charge coupled devices (CCD) and personal computers (PC), Raman spectroscopy can be increasingly considered for biomedical use (Bocklitz, Guo et al. 2016). It can provide immediate analysis and is increasingly being considered in the detection of clinical biomarkers (Banbury, Mason et al. 2019). Increasingly Raman Spectroscopy can be utilised to give rapid disease assessments, whereby specific spectral changes consistent with disease sub-type can be identified. A further advantage is that it is non-destructive and can be considered for *in vivo* use.

Raman Spectroscopy has been investigated previously in the field of gastrointestinal diseases, predominantly in the field of Barrett’s oesophagus. A study completed in an *ex vivo* setting found that Raman Spectroscopy can identify neoplasia associated with Barrett’s oesophagus with sensitivities of 73-100% and specificities 90-100% (Kendall, Stone et al. 2003). A Raman probe designed for *in vivo* use was investigated in an *ex vivo* setting by Almond *et al*, who further confirmed the potential of Raman Spectroscopy by demonstrating it can detect high grade dysplasia and adenocarcinoma associated with Barrett’s oesophagus with a sensitivity of 86% and specificity of 88% (Almond, Hutchings et al. 2014).

The next section will cover the existing literature of Raman Spectroscopy in Inflammatory Bowel Disease.

#### **4.2.2 Raman Spectroscopy in differentiating UC from non-IBD**

A paediatric study whereby surgical specimens from four UC patients and four patients without IBD were collected. The authors measured the Raman spectra from the mucosal surface and the serosal surface, using comparisons with histopathological analysis (Veenstra, Palyvoda et al. 2015). Using principal component analysis (PCA) and discriminant function analysis on the spectra from the mucosal surface, 85% of UC spectra were correctly classified according with histopathology, and 91% of the non-IBD tissue, which gave a sensitivity of 82% and specificity of 89% (Veenstra, Palyvoda et al. 2015). Similar results were seen when comparing the spectra measured from the serosa however, since endoscopic biopsies only sample the mucosal surface, this is less clinically relevant. *In vivo* probes applied via the accessory channel of an endoscope only sample the mucosal surface, therefore the serosal surface is only relevant to surgical specimens. This study is limited by its small number of patients (n=8) although they did attempt to compensate for this by increasing the number of spectra obtained per patient (total n=801). It also cannot be generalised to adult populations nor CD and the use of resected colon is not a viable option for routine clinical application.

A later study also sought to establish the ability of Raman Spectroscopy to differentiate UC from healthy control (Wood, Kendall et al. 2014). The authors recruited 18 patients with UC and 30 patients with non-IBD (healthy control) and took biopsies for *ex vivo* Raman Spectroscopy, which were frozen immediately with liquid nitrogen. Once the samples were allowed to thaw, they measured the Raman Spectra using both 10-s and 1-s acquisition times (Wood, Kendall et al. 2014). An expert histopathologist established the diagnosis and activity

of UC or healthy control. Principal component analysis (PCA) and linear discriminant analysis (LDA) were employed to analysis the Raman Spectra datasets, and models were tested with cross-validation analysis. When comparing UC with healthy control the modelling was able to differentiate with an accuracy of 95.9% with no discernible difference between the 1-s and 10-s acquisition times. This was an important finding since *in vivo* use will rely on rapid measurements.

#### **4.2.3 Raman Spectroscopy to differentiate CD from UC**

The first study to examine Raman Spectroscopy in the differentiation of CD (n=9) from UC (n=12), utilised a Raman probe *in vitro* whereby it made contact with tissue for 3 seconds, capturing 5 spectra per biopsy (Bi, Walsh et al. 2011). The authors found differences in the spectrum corresponding to nucleic acid, phenylalanine and lipids between CD and UC. They found the spectra collected from UC displayed higher lipid content and lower amounts of phenylalanine and nucleic acid (Bi, Walsh et al. 2011). This study is limited by a lack of a healthy control comparison and given Raman Spectroscopy has a degree of intra- and inter-sample variability 5 spectra per sample may not be enough to be confident in identifying real differences.

*Bielecki et al* aimed to use Raman Spectroscopy to classify CD (n=14), UC (n=13) and healthy control patients (n=11) compared with histopathological analysis (Bielecki, Bocklitz et al. 2012). Biopsies, which were snap-frozen in liquid nitrogen and later thawed, were taken for *ex vivo* Raman analysis. To analyse and compare spectral features, the authors used PCA followed by an unsupervised learning regime and modelling based on spectral differences. They found the spectra between UC, CD and HC were similar but found small differences, which is explained by the fact that biochemical composition is similar in the three different

groups (Bielecki, Bocklitz et al. 2012). When using the classification modelling, the correct classification rate was 98.9%, with a sensitivity and specificity for detecting HC of 99.1% and 98.8% respectively (Bielecki, Bocklitz et al. 2012). This is a large study in terms of patient numbers but also in terms of number of spectra analysed (n=80,356). The IBD patients included in this study had biopsies taken from normal or moderately inflamed areas, however the authors did not complete sub-group analyses comparing the different inflammatory groups. It was also not stated as to how the inflammatory status was assessed i.e., which endoscopic and histological score; therefore, it is hard to deduce how accurate the assessments of “normal or moderately inflamed” were.

*Pence et al* recruited 23 patients (CD n=15 and UC n=8) with a known diagnosis of IBD who were undergoing endoscopic evaluation as well as patients without IBD (healthy control n=8) (Pence, Beaulieu et al. 2017). Using an *in vivo* Raman probe, which passes down the accessory channel of the endoscope, measurements were recorded. The authors used a machine-learning algorithm to determine separation of different tissue types and independent training and cross validation to ascertain if the algorithm can predict tissue type. They found a sensitivity of 86.2% and specificity 39.7% at differentiating healthy control from IBD. When comparing inactive CD and inactive UC a sensitivity and specificity of 62% and 22.9% was seen respectively, which improved when there is inflammatory activity (83.3% and 55.9% respectively). This has important clinical relevance, particularly at index diagnoses of IBD (i.e., when there is likely to be activity) as it is not uncommon for a formal diagnosis of UC or CD endoscopically or histologically to be challenging. However, the consistently disappointing specificity means more work needs to be done before Raman Spectroscopy can be used in routine clinical practice for this purpose. Interestingly, there were spectral

differences noted between controls and quiescent IBD (Sensitivity 81.1% Specificity 44.4%). This is also clinically relevant given the relapse rate of patients with quiescent IBD, and perhaps there is a role for Raman Spectroscopy is to identify patients who have molecular changes that may predict relapse. This study also does not give robust details as to how inflammatory activity was graded (endoscopic and histological score), which is a limitation.

#### **4.2.4 Raman Spectroscopy to assess inflammatory activity in UC**

*Addis et al* attempted to establish if there are Raman Spectral differences between active and inactive UC (Addis, Mohammed et al. 2016). They recruited 39 patients with UC, giving 60 biopsies for *ex vivo* Raman analysis in total (n=24 endoscopic healing and n=36 active inflammation). Additional biopsies from the same site were taken for histological analysis using the Geboes score (Mosli, Feagan et al. 2015). A Geboes score of <3.1 was taken as histological healing (n=27). They used single scans of 10-s acquisition at 4 different points on each biopsy and through multivariate analysis a final model was created. The authors found the average peak intensity of inflamed tissue was at  $1003\text{cm}^{-1}$ , which was significantly higher than healed tissue. The peak intensity at  $1155\text{cm}^{-1}$  and  $1518\text{cm}^{-1}$  were much higher in healed tissue. Phospholipid concentrations were higher in those samples with mucosal healing (defined as Mayo 0), which is consistent with the fact that phospholipids are a key component of colonic cell membranes (Addis, Mohammed et al. 2016).

*Ding et al* (Ding, Dupont et al. 2017) used an *in vivo* probe to capture Raman Spectra in patients with UC (n=18) and healthy control (n=31), which was followed by biopsies for histological confirmation. The modified Riley score was used as a histological comparator (Lemmens, Arijis et al. 2013). Overall healthy control spectra (n=233) were compared with inactive (n=27), mild (n=6), moderate (n=89) and severe (n=56) colitis. They found

reductions in the content of phosphatidylcholines and lipids in active UC. Using support vector machine multivariate classifications in UC spectra with some degree of activity ( $\geq 1$  modified Riley score (Lemmens, Arijns et al. 2013)) could be differentiated from the healthy control and inactive colitis with a sensitivity of 83.5% and 97.1% respectively (Ding, Dupont et al. 2017). This gives a total AUC of 0.92 between active UC and healthy control, and 0.83 between inactive and active inflammation (Ding, Dupont et al. 2017). The strengths of this study are that a formal histological scoring system was used, in addition to having a healthy control group as a comparator. However, the majority of patients in the UC cohort had moderate-severely active UC, therefore the number of patients and spectra collected in the quiescent/mild end may not be sufficient to establish if Raman Spectroscopy can differentiate mucosal healing.

As previously mentioned in Chapter two, the traditional assessment of inflammation in IBD relies upon histological assessment. Endoscopic assessments of inflammatory activity are becoming increasingly accurate and can approximate histology (Iacucci, Fort Gasia et al. 2015). The limitations of existing scoring systems have been discussed in Chapter two, in particular in the assessment of mucosal healing. Histological scores rely upon on accurate sampling of the mucosa, with endoscopists aiming to sample a portion of colon which represents the overall inflammatory activity of the patient. This is limited when patchy inflammation is present, which may mean the biopsies taken do not reflect the inflammatory activity present. Objective, safe and rapid measures of biochemical make-up of tissue is desirable to give more accurate measures of inflammatory activity and healing: Raman Spectroscopy has the potential to meet these needs.

### 4.3 Assessment of inflammatory activity using Raman Spectroscopy

As previously mentioned, the current recommended treatment strategy of inflammatory bowel disease (IBD) is “treat to target”, whereby clinicians aim to achieve clearly defined outcomes (Peyrin-Biroulet, Sandborn et al. 2015, Turner, Ricciuto et al. 2021). I have previously presented the evidence that clinical symptoms in isolation is not a reliable measure of underlying inflammatory activity since it does not correlate well with endoscopic findings. For this reason, a key target is that of mucosal healing (MH) (Peyrin-Biroulet, Sandborn et al. 2015, Turner, Ricciuto et al. 2021). MH is associated with improved patient outcomes with fewer episodes of hospitalisation, surgical intervention and corticosteroids (Shah, Colombel et al. 2016).

However, MH is poorly defined, the assessment of which is open to debate. Endoscopic remission has traditionally been used to define MH, but inflammatory activity at histologic level can remain (Iacucci, Fort Gasia et al. 2015). It is increasingly recognised that histological remission (HR) also predicts better outcomes (Park, Abdi et al. 2016, Christensen, Hanauer et al. 2017), therefore definitions of MH is starting to incorporate endoscopic and histological remission.

The PICaSSO score, which was presented in Chapter 3, was developed to better define subtle mucosal and vascular changes of UC and for the first time introduced endoscopic findings of MH by using virtual electronic chromoendoscopy (VEC) (Iacucci, Daperno et al. 2017, Trivedi, Kiesslich et al. 2018). Despite apparent endoscopic MH (Mayo Endoscopic Score (MES) 0) (Schroeder, Tremaine et al. 1987), up to 9.4% of patients relapse within 6 months

(Barreiro-de Acosta, Vallejo et al. 2016). It is possible, that there are subtle molecular changes present in those patients that suffer early relapse, which have yet to manifest at the endoscopic level and histological level. Raman spectroscopy is one technology that can potentially provide additional information on inflammatory properties of tissue. As mentioned in this chapter, studies have demonstrated that Raman Spectroscopy is able to differentiate histologically active and inactive UC, however, to date there have been no studies that aimed to assess MH (Addis, Mohammed et al. 2016, Ding, Dupont et al. 2017). Studies have shown spectral changes between active and inactive UC (Addis, Mohammed et al. 2016, Ding, Dupont et al. 2017), however no study to date has demonstrated spectral changes before and after treatment, which would further confirm spectral changes seen between active and quiescent disease.

The potential of subtle biomolecular signals of low-grade inflammation before mucosal defects appear, to predict relapse could be an important finding but the threshold at which subtle inflammation can be detected has first to be established. However, a consistent drawback of Raman Spectroscopy is that, due to its weak signal, the signal is complex making analysis challenging (Banbury, Mason et al. 2019). Traditional methods of analysing Raman Spectroscopy data relied upon PCA (principal component analysis), which has several limitations (Cheriyadat and Bruce 2003). A novel method utilising self-organising maps, the Self-Optimising Kohonen Index Network (SKiNET), was developed which has shown promising results in studies involving eye tissue (Banbury, Mason et al. 2019) and has the potential to revolutionise the clinical application of Raman Spectroscopy.

### **4.3.1 Aims and objectives**

The aim of this chapter is to assess the ability of Raman Spectroscopy to differentiate between active IBD and MH, using formal endoscopic and histological definitions of activity and healing. A further aim was to demonstrate spectral changes from before and after biological treatment in IBD and compare those changes with spectra from non-IBD (healthy control). The primary objective was to establish spectral features of active IBD and MH in IBD and train a machine learning technique to identify the two groups. A further primary objective was to show the spectral changes from before (active IBD) and after biological therapy (endoscopic response) in IBD. A secondary objective was to establish the diagnostic performance of the trained model at differentiating the MH from active IBD.

### **4.3.2 Study design**

#### *4.3.2.1 Patients*

Patients with IBD were prospectively recruited from a single tertiary centre, University Hospitals Birmingham NHS Foundation Trust between August 2018 and October 2019, who underwent colonoscopy to assess inflammatory activity of disease. Demographic characteristics, IBD history and formal endoscopic scoring of inflammatory activity were collected. To better define spectral changes associated with MH, a sub-group of patients underwent colonoscopic examination before and 12-16 weeks following biological therapy to assess response to treatment were also included. Inclusion criteria included a proven diagnosis of IBD and age 18-80. Exclusion criteria included inability to consent, coagulopathy and pregnancy/breast-feeding. To allow comparisons with patients who do not have IBD, patients without IBD who underwent colonoscopy for non-IBD polyp surveillance were

recruited to provide a healthy control group and were excluded if macroscopic or microscopic inflammation was present.

#### *4.3.2.2 Endoscopic procedure and tissue collection*

All procedures were performed by a single endoscopist, experienced in optical diagnosis and IBD using Pentax 7010 processor (Pentax, Tokyo, Japan) and Olympus 290 series endoscopes (Olympus, Tokyo, Japan). Colonic mucosa was inspected, and precise assessments of inflammatory activity took place using high-definition white light endoscopy and VEC (Optical Enhancement iScan mode/Narrow-Band Imaging (NBI)) with and without magnification to define the most inflamed areas (Figure 4-1). Inflammatory activity was graded using the MES (Schroeder, Tremaine et al. 1987), UCEIS (Travis, Schnell et al. 2012) and the PICaSSO score (Iacucci, Daperno et al. 2017, Iacucci, Smith et al. 2021) in UC and SES-CD (Daperno, D'Haens et al. 2004) in CD. Using these advanced endoscopic technologies and endoscopic scores, targeted biopsies of inflamed/healed areas, which best reflected the overall inflammatory activity irrespective of colonic segment, were collected using biopsy forceps down the accessory channel of the endoscope. Biopsies were taken for *ex vivo* RS analysis and confirmatory histological analysis at the same site with one colonic site per patient selected. The biopsies taken for Raman Spectroscopy were placed in an empty aliquot and immediately snap frozen in liquid nitrogen and kept at -80°C, a preparation technique used in previous studies (Bielecki, Bocklitz et al. 2012, Wood, Kendall et al. 2014).



**Figure 4-1 Example of endoscopic assessment of IBD.**

Colonic mucosa was examined using high-definition white light and virtual electronic chromoendoscopy demonstrating vascular and mucosal architecture of healing using a) iScan 1 b) iScan 2 c) iScan Optical Enhancement

#### 4.3.2.3 *Definition of MH and response to treatment*

UC patients were grouped as MH if all the following stringent criteria were present:

Endoscopic scores:

- $MES \leq 1$
- UCEIS score  $\leq 1$
- PICaSSO score  $\leq 3$

Histological score:

- RHI score  $\leq 3$  without neutrophils in the Lamina Propria (Pai, Khanna et al. 2019).

CD patients were grouped as MH if all the following were present:

Endoscopic score:

- $SES-CD \leq 2$

Histological score:

- Modified Riley score 0

Patients who did not meet these criteria were classified as having active inflammation.

For those patients who underwent endoscopic examination before and after biological therapy, a response to treatment was described as a reduction in endoscopic inflammatory activity score and in histological inflammatory score.

#### 4.3.2.4 *Raman Spectroscopy measurement protocol*

The biopsies for Raman Spectroscopy analysis were placed on aluminium foil covered glass slides and were analysed at room temperature. This is a standard method for analysing Raman Spectra from biological tissue since aluminium foil has a near-featureless signal, therefore

allowing clear measurements of Raman Spectra from the target tissue (Cui, Butler et al. 2016). Measurements were collected using an InVia Qontor (Renishaw plc, Wotton-under-Edge, UK) with a 785nm (10mW) laser, 1s acquisition, 10 accumulations for each point on a 10x10 surface map scan, resulting in 100 spectra per patient. The spectrometer was calibrated prior to each use with a silicon wafer ( $520.7\text{cm}^{-1}$ ). A 50x Leica objective, 1200l/mm grating were used for all scans in the range  $550\text{-}1670\text{cm}^{-1}$ . Wire 5.1 (Renishaw plc) software was used to process the measurements, baseline subtraction and cosmic ray removal was applied to all spectra (Figure 4-2).

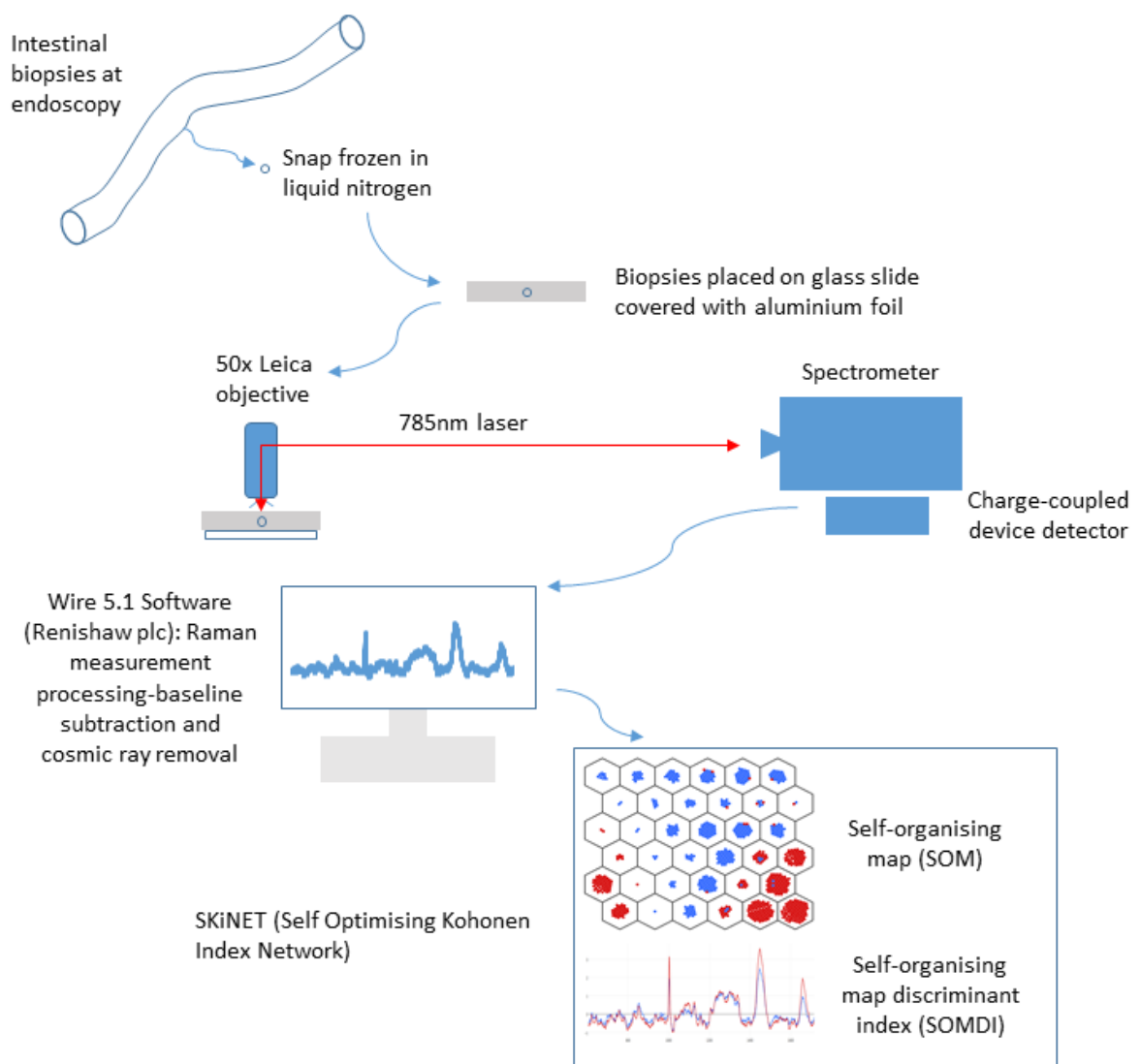


**Figure 4-2 Overview diagram demonstrating the process of sample acquisition and preparation for Raman Spectral measurement and data analysis**

#### 4.3.2.5 *Statistical analysis*

At the end of recruitment, artificial neural networks (Self-organising maps-SOM) and a supervised learning model were developed to demonstrate spectral differences and build predictive modelling (Banbury, Mason et al. 2019). A class is defined as a group of data to identify and discriminate, for example: mucosal healing and active inflammation or pre-biological therapy and post-biological therapy. To define a classification model, a pool of training data was used where the class of the data to “teach” a machine (80% of data) was known, based on the endoscopic and histological findings. This data was pulled at random. This process has been established and previously published (Banbury, Mason et al. 2019). The model was validated by presenting test data (20% of data) that the model has not seen before.

Raman Spectroscopy is multi-dimensional and has a high degree of intra- and inter-class variability for complex samples, such as tissue, meaning spectral analysis can be challenging (Butler, Ashton et al. 2016). The Self-organising map (SOM) is a method that can reduce the hyper-dimensional nature of RS to a two-dimensional model. A SOM represents an artificial neural network arranged as a map of hexagonal neurons (Kohonen 1982), with each individual spectrum represented as a dot on the map. Each neuron activates for certain inputs (spectral features) and not others. Neurons activate on similar inputs to their neighbours, therefore clustering occurs when multiple neurons group together when activated by the same characteristic, similar to a human brain (Banbury, Mason et al. 2019). A self-organising map discriminant index (SOMDI) (Banbury, Mason et al. 2019) is extracted from this process, which allows an understanding of which features of the training data causes the spatial clustering and separation of classes observed in the SOM. The magnitude in the SOMDI relates to how strongly a variable is considered a discriminator for a given class. This results

in a representation similar to a spectrum, where we can compare SOMDI for each group, and identify Raman bands that are more important to one class from another. I exploited the newly developed SKiNET (Self Optimising Kohonen Index Network) approach, which provides a means of supervised learning to further optimise the result (Banbury, Mason et al. 2019).

The model was optimised by 10-fold stratified cross-validation and oversampling to account for class imbalance in the training data. Since the present study was a pilot study, using for the first time SKiNET (Banbury, Mason et al. 2019) in IBD, a sample size calculation was not possible. To further confirm the difference between classes we ran the analysis 10 times and took an average sensitivity, specificity, PPV, NPV and accuracy plus standard deviation.

#### *4.3.2.6 Ethical approval*

Ethical approval was granted by the Office for Research Ethics Committees Northern Ireland (17/NI/0148) and the South West-Frenchay Research Ethics Committees 19/SW/0010.

Written informed consent was obtained for all participants.

## 4.4 Results

### **4.4.1 Raman Spectroscopy in the assessment of inflammatory activity and mucosal healing:**

#### *4.4.1.1 4.3.3.1 Patient characteristics*

A total of 74 patients were included (42 UC and 32 CD) (Table 4-1) with 7400 Raman Spectra acquired overall. The median age of patients recruited was 41.2 years ( $\pm 13.4$ , range 19-66), and 34 (45.9%) of patients were male. The majority (88.1%) of UC patients were on 5-ASA therapy, and 35.7% of UC patients and 53.1% of CD patients were on biological therapy at the time of recruitment.

**Table 4-1 Patient characteristics of the Raman Spectroscopy in the differentiation of MH from active inflammation in IBD study**

| Characteristics            | Patients        |
|----------------------------|-----------------|
| Age, median $\pm$ sd (yrs) | 41.2 $\pm$ 13.4 |
| Gender                     |                 |
| Male (%)                   | 34 (45.9%)      |
| Female (%)                 | 40 (54.1%)      |
| IBD type                   |                 |
| UC (%)                     | 42 (56.8%)      |
| CD (%)                     | 32 (43.2%)      |
| Therapy, n (%)             |                 |
| UC                         |                 |
| NSAIDs                     | 6 (14.3%)       |
| 5-ASA                      | 37 (88.1%)      |
| Steroids                   | 26 (61.9%)      |
| Thiopurine                 | 28 (66.7%)      |
| Biologic                   | 15 (35.7%)      |
| CD                         |                 |
| NSAIDs                     | 6 (18.8%)       |
| 5-ASA                      | 5 (15.6%)       |
| Steroids                   | 18 (56.3%)      |
| Thiopurine                 | 22 (68.8%)      |
| Biologic                   | 17 (53.1%)      |

#### 4.4.1.2 *Differentiating MH from active inflammation in UC*

A total of nine (21.4%) UC patients had evidence of MH whilst 33 (78.6%) had active inflammation. Figure 4-3 demonstrates the SOM, showing the MH (blue) vs. active inflammation (red) groups are well separated. The key spectral differences from the SOMDI (Figure 4-4 )are presented in Table 4-2 (cross-validation accuracy 0.95). The trained artificial neural network was able to differentiate between MH and active inflammation in 20% of the testing dataset with a sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy of 96.29% (sd 0.94), 95.03% (sd 1.52), 94.89% (sd 1.59), 96.33 (sd 0.97) and 95.65 (sd 0.99) respectively (Table 4-3).

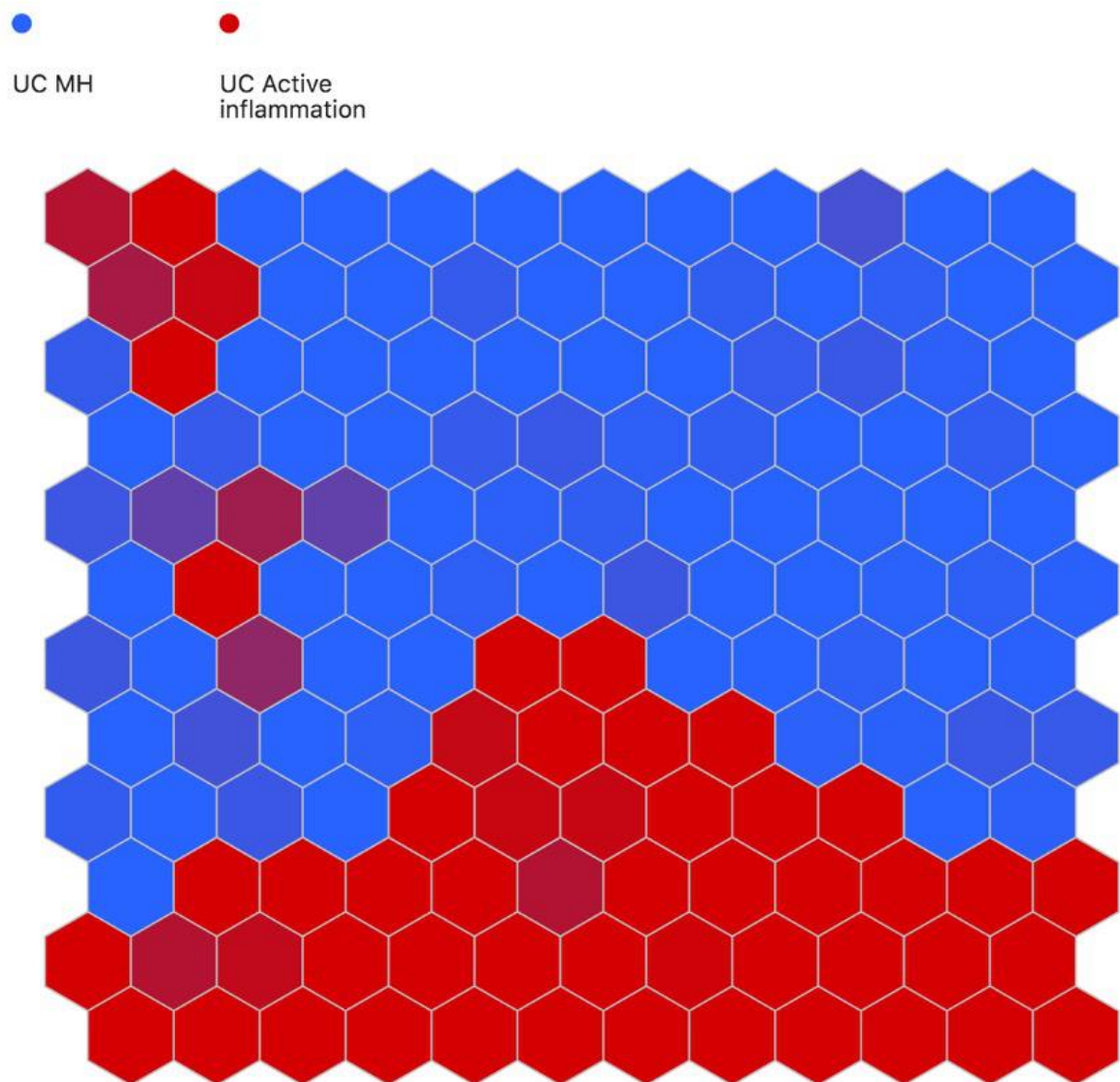


**Figure 4-3 Self-organising map (SOM) of UC mucosal healing (Blue) vs. active inflammation (Red)**

SOM-each dot on the map represents a neuron, and each neuron represents a spectrum. Each new neuron will be drawn to its nearest vector, thereby causing clustering of neurons that have similar characteristics

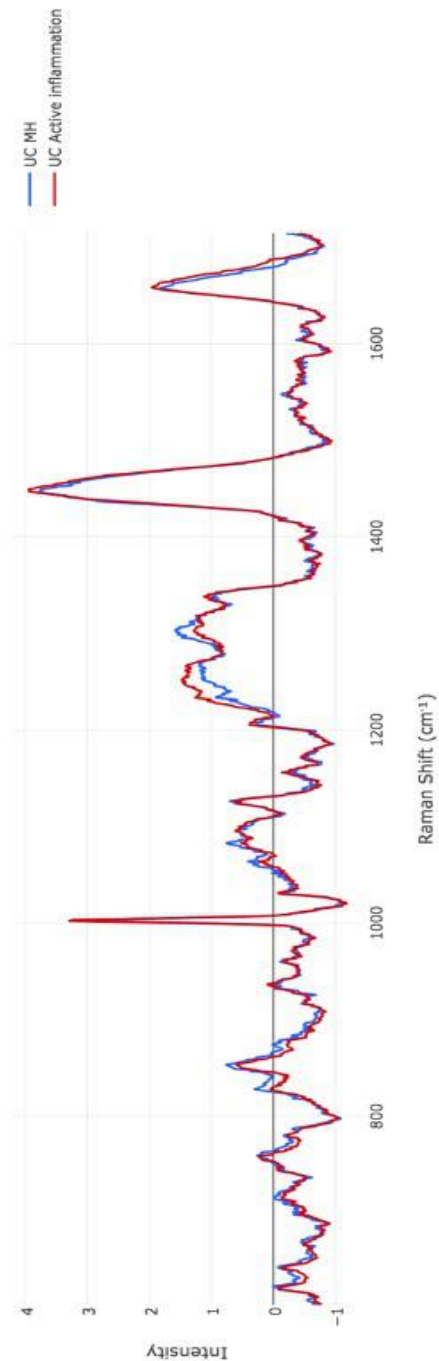


**Figure 4-4 Self-organising map discriminant index (SOMDI) of UC MH (blue) vs active inflammation (red)**

The SOMDI gives a visual representation of the SOM data, highlighting the differences in spectra

**Table 4-2 Relative intensity of Raman shifts between UC MH vs active inflammation  
taken from SOMDI**

| <b>Raman Shift (cm<sup>-1</sup>)</b> | <b>UC Active inflammation (Intensity)</b> | <b>UC Mucosal Healing (Intensity)</b> |
|--------------------------------------|-------------------------------------------|---------------------------------------|
| <b>1003</b>                          | 3.30                                      | 2.69                                  |
| <b>1252</b>                          | 1.44                                      | 1.11                                  |
| <b>1304</b>                          | 1.27                                      | 1.54                                  |
| <b>1449</b>                          | 3.92                                      | 3.73                                  |
| <b>1657</b>                          | 1.93                                      | 1.83                                  |

**Table 4-3 Example confusion matrix of the trained predictive model at differentiating  
MH from active UC**

|                               | <b>UC MH</b> | <b>UC Active inflammation</b> |
|-------------------------------|--------------|-------------------------------|
| <b>UC MH</b>                  | 170          | 10                            |
| <b>UC Active inflammation</b> | 7            | 173                           |

#### 4.4.1.3 *Differentiating MH from active inflammation in CD*

**A total of ten (31.3%) CD patients had evidence of endoscopic and histological healing and 22 (68.8%) active inflammation. The separation between MH CD (blue) and active inflammation CD (red) is demonstrated by the SOM in Figure 4-5. The spectral differences presented in SOMDI (Figure 4-6) are demonstrated in Table 4-4. The artificial neural network was able to differentiate the remaining 20% of the data with a sensitivity, specificity, PPV, NPV and accuracy of 96.19% (sd 1.46), 88% (sd 4.20), 86.60% (sd 5.39), 96.55% (sd 1.32) and 91.6% (sd 2.75) respectively (**

Table 4-5).

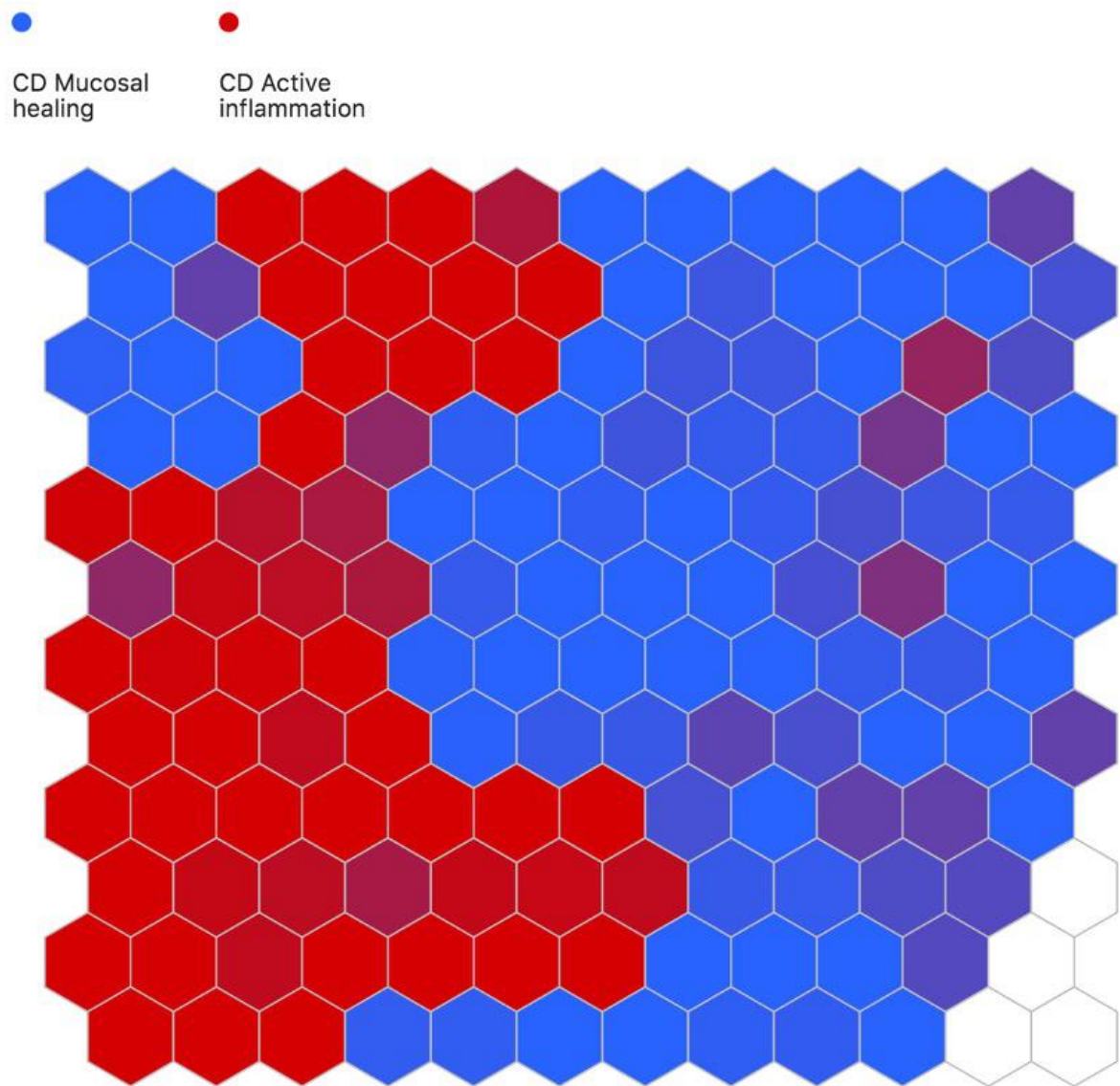


**Figure 4-5 SOM of CD MH (Blue) vs. active inflammation (red)**

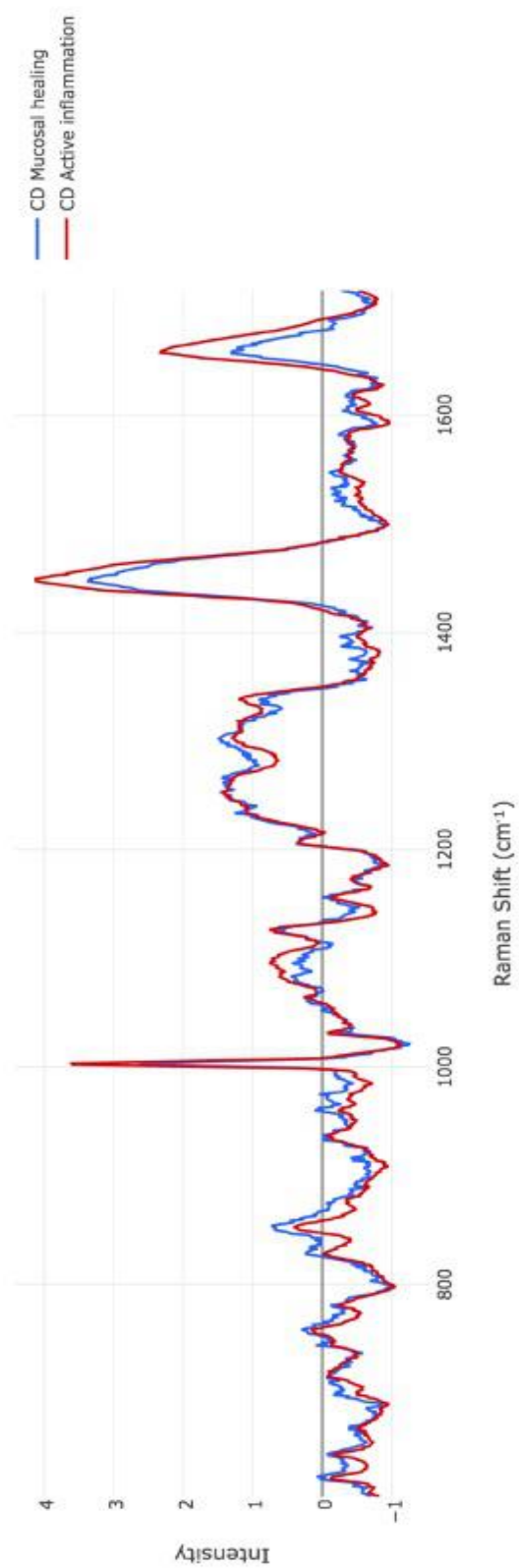


**Figure 4-6 SOMDI of CD MH (Blue) vs active inflammation (red) highlighting the spectral differences from the SOM**

**Table 4-4 Relative intensity of Raman shifts between CD healing vs non-healing taken from SOMDI**

| <b>Raman Shift (cm<sup>-1</sup>)</b> | <b>CD Active inflammation (Intensity)</b> | <b>CD Mucosal Healing (Intensity)</b> |
|--------------------------------------|-------------------------------------------|---------------------------------------|
| <b>1003</b>                          | 3.60                                      | 2.63                                  |
| <b>1252</b>                          | 1.40                                      | 1.40                                  |
| <b>1304</b>                          | 1.30                                      | 1.46                                  |
| <b>1449</b>                          | 4.11                                      | 3.30                                  |
| <b>1657</b>                          | 2.32                                      | 1.30                                  |

**Table 4-5 Example confusion matrix of the trained predictive model at differentiating MH from active CD**

|                               | <b>CD MH</b> | <b>CD Active inflammation</b> |
|-------------------------------|--------------|-------------------------------|
| <b>CD MH</b>                  | 181          | 19                            |
| <b>CD Active inflammation</b> | 7            | 193                           |

#### **4.4.2 Biochemical changes following response to biological therapy**

##### *4.4.2.1 Patient demographics of pre- vs. post-biologic therapy*

A total of 23 patients were included (6 UC, 8CD and 9 HC) in the exploratory analysis comparing spectral features in patients with IBD before and after biological therapy with patients without IBD (Table 4-6). The median age of patients in this analysis was 44.0 (+/- 14.8, range 25-79), nine were male (39.1%) and all UC patients all but one CD patients (7 out of 8) were biologic naïve.

**Table 4-6 Patient characteristics of the pre- vs. post-biological therapy study**

| Characteristics                              | Patients        |
|----------------------------------------------|-----------------|
| <b>Age, median <math>\pm</math> sd (yrs)</b> | 44.0 $\pm$ 14.8 |
| <b>Gender</b>                                |                 |
| <b>Male (%)</b>                              | 9 (39.1%)       |
| <b>Female (%)</b>                            | 14 (60.9%)      |
| <b>IBD type</b>                              |                 |
| <b>UC (%)</b>                                | 6 (26.1%)       |
| <b>CD (%)</b>                                | 8 (34.8%)       |
| <b>Non-IBD</b>                               | 9 (39.1%)       |
| <b>Therapy, n (%)</b>                        |                 |
| <b>UC</b>                                    |                 |
| NSAIDs                                       | 1 (16.7%)       |
| 5-ASA                                        | 6 (100.0%)      |
| Steroids                                     | 4 (66.7%)       |
| Thiopurine                                   | 3 (50.0%)       |
| Biologic                                     | 0 (0.0%)        |
| <b>CD</b>                                    |                 |
| NSAIDs                                       | 1 (12.5%)       |
| 5-ASA                                        | 1 (12.5%)       |
| Steroids                                     | 6 (75.0%)       |
| Thiopurine                                   | 7 (87.5%)       |
| Biologic                                     | 1 (12.5%)       |

#### *4.4.2.2 Pre- vs. post-treatment in UC and healthy controls*

The separation between pre- (red) vs. post-biologic (blue) in UC patients and healthy control (black) is shown by the SOM in Figure 4-7 and the spectral differences are shown in the SOMDI in Figure 4-8. The key spectral differences that differentiate the 3 groups are presented in Table 4-7 (cross-validation 0.95).



**Figure 4-7 SOM of the UC pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black)**

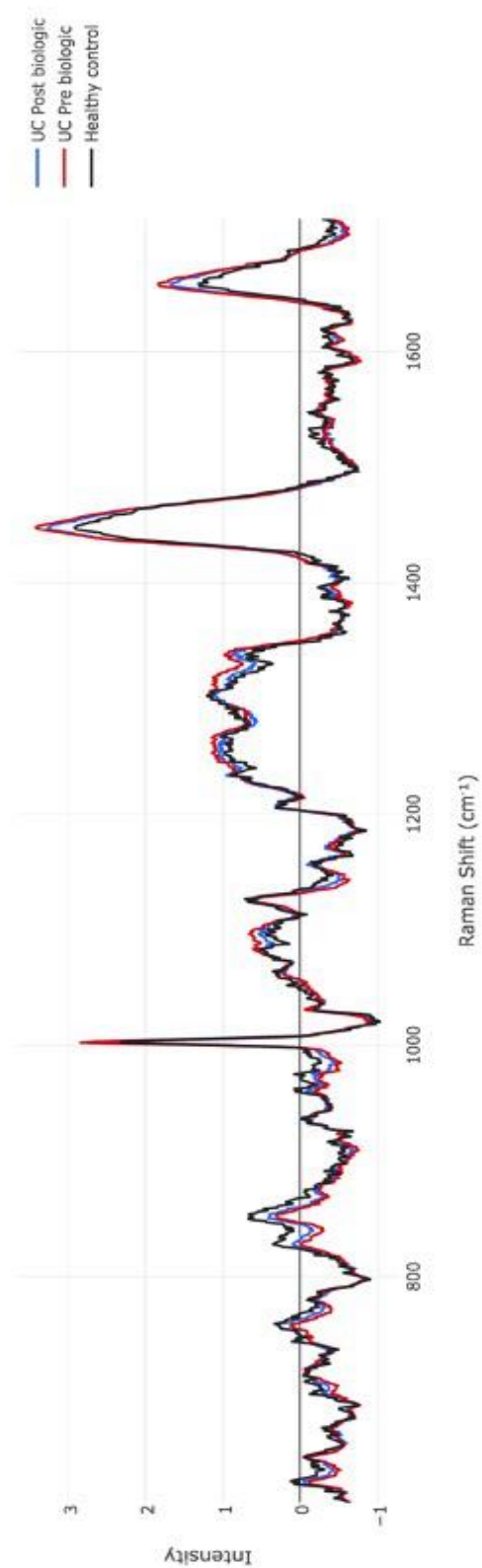


**Figure 4-8 SOMDI highlighting the spectral differences from the SOM of UC pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black)**

**Table 4-7 Relative intensity of Raman shifts between UC pre biologic and post biologic and HC taken from SOMDI.**

| <b>Raman Shift (cm-1)</b> | <b>UC Pre Biologic (Intensity)</b> | <b>UC Post Biologic (Intensity)</b> | <b>Healthy control (Intensity)</b> |
|---------------------------|------------------------------------|-------------------------------------|------------------------------------|
| <b>1003</b>               | 2.85                               | 2.72                                | 2.32                               |
| <b>1252</b>               | 1.13                               | 1.04                                | 0.99                               |
| <b>1304</b>               | 1.17                               | 1.10                                | 1.16                               |
| <b>1449</b>               | 3.42                               | 3.24                                | 2.91                               |
| <b>1657</b>               | 1.84                               | 1.68                                | 1.32                               |

#### *4.4.2.3 Pre- vs. post-biological therapy spectra in CD patients and healthy controls*

The separation of pre-biologic CD (red), post-biologic CD (blue) and HC (black) is presented in the SOM in Figure 4-9, and SOMDI demonstrating the spectral features is shown in Figure 4-10 (cross-validation 0.83). The spectral differences that differentiate the groups taken from the SOMDI are presented in Table 4-8. This exploratory analysis provided additional information on the spectral features to focus on in the differentiation between MH and active inflammation in IBD.



**Figure 4-9 SOM demonstrating the groupings of CD pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black)**

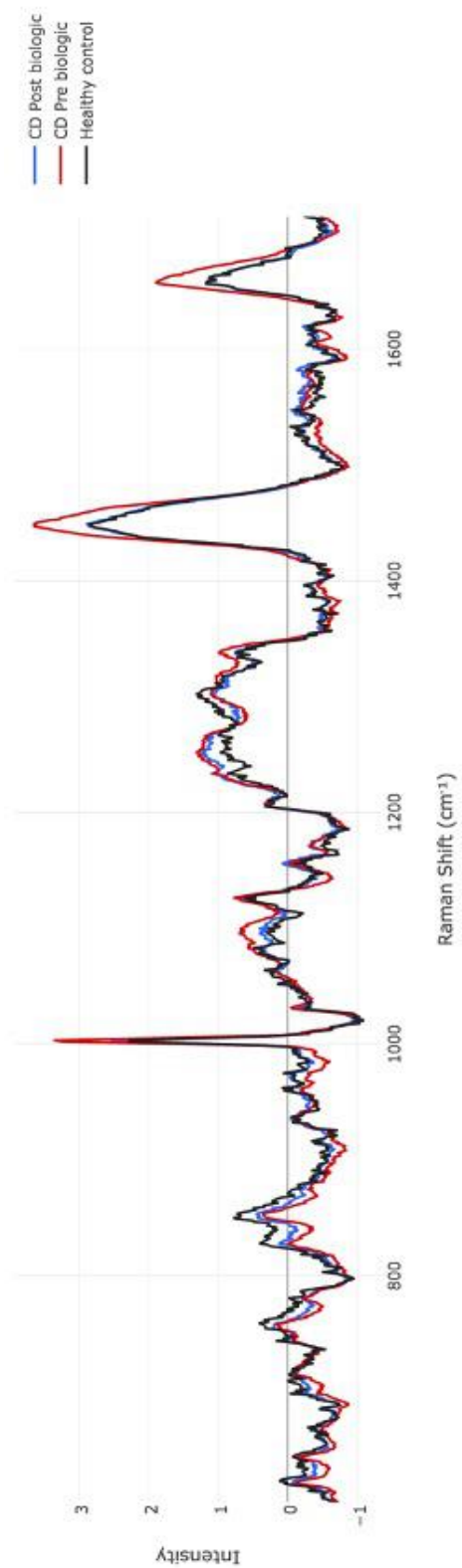


**Figure 4-10 SOMDI highlighting the spectral differences from the SOM of CD pre- (Red) vs. post-biologic therapy (Blue) vs. healthy control (black)**

**Table 4-8 Relative intensity of Raman shifts between CD pre-biologic vs. CD post-biologic vs. healthy control taken from SOMDI.**

| <b>Raman Shift<br/>(cm-1)</b> | <b>CD Pre Biologic</b> | <b>CD Post Biologic</b> | <b>Healthy control</b> |
|-------------------------------|------------------------|-------------------------|------------------------|
| <b>1003</b>                   | 3.38                   | 2.49                    | 2.29                   |
| <b>1252</b>                   | 1.32                   | 1.24                    | 0.98                   |
| <b>1304</b>                   | 1.10                   | 1.05                    | 1.27                   |
| <b>1449</b>                   | 3.66                   | 2.90                    | 2.87                   |
| <b>1657</b>                   | 1.89                   | 1.84                    | 1.17                   |

#### 4.4.2.4 *Biomolecular explanation for differences in Raman Spectra*

The spectral distribution was very similar between active inflammation and MH in both UC and CD, however there are subtle differences demonstrated. The Raman band at  $1003\text{cm}^{-1}$  is associated with the amino acid phenylalanine (Krafft, Knetschke et al. 2003, Talari, Movasaghi et al. 2015), and our results show a relative reduction at this band in MH and when there is a reduction in inflammatory activity and a lower level in healthy controls. A higher intensity at Raman band  $1252\text{cm}^{-1}$  in UC when inflammation is present compared with MH, which may represent amide III (Krafft, Knetschke et al. 2003, Bielecki, Bocklitz et al. 2012, Talari, Movasaghi et al. 2015) and lower intensity was seen in healthy controls. An increase in intensity at  $1304\text{cm}^{-1}$ , seen in both UC and CD when MH is present, likely corresponds to phospholipids (Omberg, Osborn et al. 2002, Talari, Movasaghi et al. 2015, Ding, Dupont et al. 2017). A reduction in intensity at  $1657\text{cm}^{-1}$  in UC and CD patients in MH, following treatment with biological therapy and healthy controls which could be attributed to C=C stretching in lipids (Talari, Movasaghi et al. 2015, Ding, Dupont et al. 2017, Pence, Beaulieu et al. 2017).

## 4.5 Discussion

The assessment of MH is increasingly important in the management of IBD, given the association with favourable outcomes for patients (Peyrin-Biroulet, Sandborn et al. 2015, Turner, Ricciuto et al. 2021), which was presented in Chapter 2. Endoscopic assessments of MH can be inaccurate, particularly using older scoring systems such as the MES (Iacucci, Fort Gasia et al. 2015). Using a multi-step process and a novel spectral analysis software, Raman spectral changes before and after biological therapy in IBD were established for the first time and compared MH vs. active inflammation in UC and CD. Relative Raman spectral changes in keeping with MH in CD and UC were identified, defined in histological and endoscopic terms-so-called endo-histological MH. An artificial neural network can differentiate between MH and active disease with high accuracy in both UC and CD.

The discrepancy between endoscopic and histological assessment of MH is subject of considerable controversy as it may be relevant to the disease course in patients with IBD. Significant proportions of patients who appear in endoscopic remission still have evidence of microscopic inflammation and suffer early relapse (Bryant, Winer et al. 2014). Whilst histological assessment of inflammation is considered gold standard and may help predict prognosis (Calafat, Lobaton et al. 2017, Narang, Kaur et al. 2018, Kanazawa, Takahashi et al. 2019), it is reliant upon accurate biopsy sampling and incurs significant healthcare costs. Raman Spectroscopy has been emerging as a candidate technique to provide the crucial missing information since it is known to be able to give a rapid assessment of the biochemical “make-up” of biological molecular tissue. It can give instant biomolecular assessments and offer optical diagnosis without the need for biopsies (Ding, Dupont et al. 2017, Pence,

Beaulieu et al. 2017), which would be safer for patients as well as offering significant cost benefits. Raman Spectroscopy does not require specialist tissue preparation, labelling, is non-destructive and provides an overview of biochemical composition rather than an individual biomarker (Pence and Mahadevan-Jansen 2016, Shi, Chen et al. 2016). For these reasons it can be considered for *in vivo* use (Ding, Dupont et al. 2017, Pence, Beaulieu et al. 2017).

The main aim of this chapter was to explore the differences in spectra between MH and active inflammation. Therefore, the tallest peaks were focused on, since the higher the peak, the more important the corresponding wavelength is in differentiating classes. The Raman Spectroscopy distribution were very similar between CD, UC and HC, but subtle differences in intensities of various Raman bands is noted between inflammation and healing, which is consistent with previous work (Morasso, Truffi et al. 2020).

The reduction in Raman band at  $1003\text{cm}^{-1}$ , which is associated with the amino acid phenylalanine (Krafft, Knetschke et al. 2003, Talari, Movasaghi et al. 2015), in MH and following successful treatment with biological therapy is consistent with previous work by *Addis et al* (Addis, Mohammed et al. 2016). Increases at Raman band  $1003\text{cm}^{-1}$  were noted when inflammation is present in both UC and CD. Higher levels of Phenylalanine has been reported to be associated with immune activation and inflammation (Murr, Grammer et al. 2014), therefore when there is active inflammation present in both UC and CD and pre-biological therapy treatment there are greater infiltration of inflammatory cells thereby increasing the quantity of phenylalanine, which is reflected in the results of this study.

Likewise, when there is reduced inflammatory activity or healing present, there is less inflammatory cell infiltrate, as reflected by histological scores, thereby reducing the quantity of amino acids such as phenylalanine. The reduction at this band following biological therapy, brought intensities that were more in line with healthy control patients.

The higher intensity at Raman band  $1252\text{cm}^{-1}$ , which is likely a constituent of haem (Krafft, Knetschke et al. 2003, Bielecki, Bocklitz et al. 2012), is consistent with inflammatory activity (Bielecki, Bocklitz et al. 2012). However, this was not found when comparing active inflammation and MH directly in CD. This lack of consistency may reflect the fact that UC is a mucosal disease and is associated with bleeding when inflammation is present on mucosal surface, more so than CD.

The increase in intensity at  $1304\text{cm}^{-1}$ , seen in both UC and CD when MH is present, likely corresponds to phospholipids, which are key to the integrity of the cell wall membrane of colonic membrane (Omberg, Osborn et al. 2002, Ding, Dupont et al. 2017). This has potential clinical significance, as it demonstrates a potential positive biomarker suggesting MH is present. Given the limitations of endoscopic and histological evaluations of MH, this may provide an additional measure to confirm the presence of MH.

The hyper-spectral nature can make RS analysis challenging. Traditionally principal component analysis (PCA) has been used, however this has several limitations (Cheriyadat and Bruce 2003). Although a simple and popular method, it is not considered the best method of feature extraction (Cheriyadat and Bruce 2003). Also in recent years, the advancement of technology and machine learning has made PCA relatively outdated (Brereton 2012). A method that is increasingly being used, particularly for multidimensional data such as Raman

Spectroscopy, is self-organising maps (SOM) (Kohonen 1982), which is thought to have significant advantages over PCA, in particular in showing data in a clear manner allowing feature extraction and classification (Brereton 2012, Banbury, Mason et al. 2019).

With the aid of SKiNET (Banbury, Mason et al. 2019), subtle differences can be detected. This has a potential application in clinical practice when assessing response to treatment, particularly when considering de-escalation of biological therapy. Stopping therapy in IBD in order to minimise long term side effects and risks of treatments, the “Exit strategy”, is a hot topic in IBD management (Doherty, Katsanos et al. 2017) and is a difficult decision for clinicians largely due to a lack of robust guidance (Doherty, Katsanos et al. 2018). The risk of relapse within 1 year of biological cessation is as high as 38-44% (Gisbert, Marin et al. 2016). In those patients who are at risk of short-term relapse, it is probable that biochemical changes are present which have yet to manifest as mucosal changes, and Raman Spectroscopy could have a role in identifying those patients.

Early studies aiming to differentiate active from inactive UC give insight into the potential of Raman Spectroscopy. *Addis et al* included 39 patients with UC and identified Raman band differences between active and healed UC however this study was limited by single scan measurements on each biopsy and by the use of the MES (Schroeder, Tremaine et al. 1987) as a definition of healing (Addis, Mohammed et al. 2016). *Ding et al* recruited 18 patients with UC and found a machine-learning algorithm can differentiate between inactive UC from inflamed UC with an area under the curve of 0.83 (Ding, Dupont et al. 2017). Whilst the healing patients were identified using the modified Riley score, there was no endoscopic scoring of inflammatory activity.

I have demonstrated for the first time that Raman Spectroscopy can be highly accurate in differentiating between active inflammation and MH using robust endoscopic and histological definitions, in both UC and CD. By including 97 patients (88 IBD, 9 HC), measuring 100 spectra per patient (giving 9900 spectra in total), the present study is the largest in terms of patient number and spectra in IBD (Bielecki, Bocklitz et al. 2012, Wood, Kendall et al. 2014, Ding, Dupont et al. 2017). This study is also the first to demonstrate spectral changes following treatment of IBD and compare MH and active inflammation in CD. This study is also the first to group patients as having MH, which is more clinically relevant to modern IBD management. It is the first to compare MH and active inflammation in CD. A further strength of the present study is in using the latest optical enhancement endoscopic technology, we ensured accurate assessments of inflammatory activity.

A limitation to the study is that results were acquired in the *ex vivo* setting, thereby laying an important foundation for further *in vivo* application. For this purpose, laser power, total acquisition time and auto fluorescence issues will need to be carefully considered.

Nevertheless, our aim was to explore spectral differences, and a potentially more focused map scan only including those wavelengths of interest could take place *in vivo*, which would take less time. Each patient was represented by 100 spectral maps in this analysis. Ideally these 100 maps would be averaged and 1 spectrum per patient would be analysed.

However, large patient numbers would be required for this otherwise would be subject to bias from outlier patients. A further limitation is that one colonic segment and one Raman

Spectroscopy biopsy was taken per patient, therefore we are unable to assess whether similar results of healing vs. active inflammation can be seen within the same patient if patchy inflammation is seen. Biopsies taken from mucosa in endoscopic remission which later had evidence of histological activity were included in the active inflammation groups. A further area of interest would be in identifying subtle spectral differences in patients with endoscopic MH but histological inflammation from patients with endo-histological remission, as this would help shape decision making in IBD.

Although these results are promising, there is reservation regarding the clinical relevance in CD. As mentioned in Chapter 1, CD is a transmural disease, as opposed to UC which is a mucosal disease. Raman Spectroscopy is very much a surface imaging tool, and I have demonstrated it can provide biomolecular information on the target tissue. However, given CD is transmural, it is possible that a mucosal surface in CD may have an absence of inflammation endoscopically and histologically, but inflammation may be present more deeply in the bowel wall. Since Raman Spectroscopy gives surface information, it may be of limited use in assessing quiescent disease in CD. However, the STRIDE (Selecting Therapeutic Targets in Inflammatory Bowel Disease) recommendations for the target of treatment in CD is the resolution of ulceration at ileocolonoscopy (Peyrin-Biroulet, Sandborn et al. 2015), which reflects the mucosal status of the disease. For those patients that cannot be assessed via ileocolonoscopy, resolution of inflammatory findings at cross-sectional imaging is considered the target (Peyrin-Biroulet, Sandborn et al. 2015). Therefore, whilst Raman Spectroscopy cannot give detailed inflammatory data on possible deep inflammation in CD, it can give additional information on mucosal inflammation and MH based on current recommendations.

Identifying fibrosis in CD, can be challenging, with regards to differentiating fibrotic from inflammatory strictures using existing assessment modalities. It is probable the biochemical make-up of fibrosis differs from inflammation and therefore RS could be a useful management adjunct. RS probes that can be used via endoscopy channels are being designed for colonoscopes and our study shows the basis for exploring further such probes for assessment of response to therapy and mucosal healing at a molecular level in IBD.

## 4.6 Conclusion

Overall, I successfully differentiated between active inflammation and histo-endoscopic MH in UC and CD using Raman Spectroscopy and SKiNET. I. Established spectral changes before and after biological therapy in IBD and Raman Spectroscopy may be able to give more detailed information to clinicians on the presence or absence of MH. For many years the focus of IBD management has been on achieving endoscopic MH, and in more recent insight into the potential role of an additional mode of healing: molecular healing, which could further ensure robust assessment of healing. Further multicentre studies in this field needs to take place before this can be used in everyday clinical practice.

The section above is taken from the following publication:

**Smith SCL**, Banbury C, Zardo D, Cannatelli R, Nardone OM, Shivaji UN, Ghosh S, Oppenheimer PG, Iacucci M. Raman Spectroscopy can accurately differentiate mucosal healing from non-healing in inflammatory bowel disease. PLoS ONE 2021 (Epub)

## 5 SECTION THREE: TRAINING IN ADVANCED ENDOSCOPY

### 5.1 Overview

I have produced evidence of the application of novel advanced endoscopy in clinical practice, however this will be of limited value if training in this field is not robust. In the following chapter I present the evidence of how advanced endoscopy can be utilised in the field of small/diminutive colorectal polyps and present a randomised controlled study that compares traditional didactic training with computer-based self-learning. I present a systematic review and meta-analysis in the available optical diagnosis of small/diminutive polyp training methods. Finally, I discuss how advanced technologies could be utilised in the field of IBD surveillance and present an observational study of a training intervention on the quality of IBD surveillance at the University Hospitals Birmingham NHS Trust.

### 5.2 Small/diminutive colorectal polyps

#### 5.2.1 Optical diagnosis training of small/diminutive colorectal polyps

Gastrointestinal endoscopy is integral to the diagnosis, treatment and management of colorectal polyps. It is thought that bowel cancers originate as polyps, and therefore early detection and removal of polyps can reduce the risk and incidence of bowel cancer. Optical diagnosis is the process whereby predictions of histology are made based on endoscopic appearances and is increasingly possible through developments in endoscopic technology. Accurate optical diagnosis allows small/diminutive colorectal polyps (<10mm) to be removed and discarded without the need for formal histological assessment: the “Resect and Discard” strategy (Abu Dayyeh, Thosani et al. 2015). Small/diminutive polyps are selected in this strategy since the risk of malignancy or high grade dysplasia is very small (<1%) (Lieberman,

Moravec et al. 2008, Rex, Overhiser et al. 2009). This strategy also includes the identification of hyperplastic polyps in the rectum and sigmoid colon, which can be left in situ given the low risk of transformation into advanced histology. This will mean huge savings to healthcare services as fewer specimens are sent for formal histological analysis, and lower risk to patients with fewer polypectomies required. The incorporation of optical diagnosis of small/diminutive polyps has been endorsed by The American Society of Gastrointestinal Endoscopy (ASGE) as well as The National Institute for Health and Care Excellence (NICE) in the UK ([www.nice.org.uk/guidance/dg28](http://www.nice.org.uk/guidance/dg28)). If implemented there would be fewer specimens sent for histological analysis with associated cost savings (Abu Dayyeh, Thosani et al. 2015, Solon, Klausnitzer et al. 2016).

Effective optical diagnosis enables endoscopists to accurately identify malignant appearances within lesions, which would then activate an alternative management pathway. Since malignant polyps are those described as having invasion beyond the muscularis, endoscopic removal should not be attempted as its risk bowel perforation and incomplete resection. Accurate optical diagnosis can also help identify lesion borders, and together with recognising malignancy will help improve patient management. Sessile serrated lesions (SSL) are regarded as subtle lesions that can be easily missed or misinterpreted as hyperplastic polyps (Murakami, Sakamoto et al. 2019), however enhanced optical diagnosis ability may give endoscopists confidence in the detection and characterisation of SSL (Bisschops, Hassan et al. 2018, Iacucci, Trovato et al. 2018). This is important since SSL can have dysplasia present (Murakami, Sakamoto et al. 2019), and therefore recognition as SSL as opposed to hyperplastic is vital to ensure prompt removal improving patient management.

Endoscopists who are knowledgeable and can practice high quality optical diagnosis can only benefit patients. However, to equip endoscopists in optical diagnosis, training will be important, which was recognised in a recent evidence-based consensus (Bhandari, Thayalasekaran et al. 2018). Possible training methods in optical diagnosis include traditional didactic training and computer-based self-learning. The following sections covers the published evidence in this field of optical diagnosis training of small/diminutive colorectal polyps in trainees and experienced endoscopists. For the purposes of this review, which is focused on the effect of training, I have only included those studies which publish pre- and post-training data.

#### *5.2.1.1 Optical diagnosis of small/diminutive colorectal polyps in trainee endoscopists studies*

A total of 7 studies included trainee endoscopists (n=62), 4 of which used didactic training and 3 computer-based. All included studies were prospective observational studies. Didactic training sessions of 60-90 minutes have been shown to improve the accuracy of optical diagnosis from 61% (95% CI 54-68) to 71% (95% CI 68-74 p0.037) (Basford, Longcroft-Wheaton et al. 2017), which was also seen when shorter 20 minute didactic training was provided, improving accuracy from 60.0% (sd  $\pm$  24.1) to 94.0% (sd  $\pm$  4.7 p=0.0009) (Raghavendra, Hewett et al. 2010). However, the improvement was not consistently seen during the validation study of the SIMPLE (Simplified Identification Method for Polyp Labelling during Endoscopy) classification, whereby there was no significant improvement in optical diagnosis accuracy following a 1 hour didactic training session, 72% (95% CI 64-78) vs. 78% (95% CI 71-84 p=NS), when trainees scored NBI polyp videos, yet there was an improvement when reviewing iScan OE (optical enhancement) videos (57% (95% CI 52-63) vs. 68% (95% CI 63-73) p=0.004) (Iacucci, Trovato et al. 2018). The use of a validated polyp

classification to aid optical diagnosis is recommended (Bisschops, Dekker et al. 2019), however despite this, a study using BASIC (BLI Adenoma Serrated International Classification) failed to see an improvement in trainee accuracy, 88% (95% CI 83-92) pre-training vs. 91% (95% CI 85-93  $p=0.26$ ) post-training (Subramaniam, Hayee et al. 2019). One potential explanation behind inconsistency of accuracies may be because of high variability of trainers' ability to teach. Therefore, an alternative mode of training is computer-based training, which is more easily standardised and reproducible. The first study in this area showed improvements in accuracy can be achieved, 75% (95% CI 67-82) vs. 90% (95% CI 84-93  $p<0.001$ ), through training which took only 20 minutes (Ignjatovic, Thomas-Gibson et al. 2011). This was an effect that was consistently seen in subsequent studies, including an NBI-based study (86.0% (95% CI 82.6-89.2) vs. 92.0% (95% CI 89.3-94.7)  $p<0.001$ ) (Aihara, Kumar et al. 2018) and an iScan-based study (65.7% (95% CI 60.8-70.7) vs. 71.1% (95% CI 67.4-74.9)  $p=0.013$ ) (Basford, Brown et al. 2019).

#### 5.2.1.2 *Experienced/fully trained endoscopists*

The most studied form of training to experienced endoscopists is didactic training with a total of seven studies, followed by six computer-based training studies. The first study to assess didactic training amongst experienced endoscopists showed no improvement in optical diagnosis accuracy following training (90% vs. 95%  $p0.55$ ) (Rogart, Jain et al. 2008), which was also found in a subsequent study when using conventional white light endoscopy, however when using NBI there was an improvement in accuracy (65% (95% CI 58-72) vs. 78% (95% CI 72-84)  $p0.001$ ) (Higashi, Uraoka et al. 2010). The lack of improvement was further confirmed in a study that delivered two 1 hour didactic training sessions to experienced endoscopists, and the authors found an insignificant impact of training (OR 1.22

95% CI 0.88-1.71  $p=0.23$ ) (Coe, Thomas et al. 2012). A study including experienced endoscopists did show didactic training can improve accuracy in optical diagnosis, pre-training 45.4% (SD  $\pm$  25.3) vs. post-training 90.8% (SD  $\pm$  7.5)  $p=0.0001$  (Raghavendra, Hewett et al. 2010). Resource intensive didactic training consisting of monthly 1 hour sessions for 3 months can lead to a statistically significant improvement in optical diagnosis accuracy amongst experienced endoscopists, 79.54% (SD  $\pm$  4.76%) pre-training vs. 96.46% (SD  $\pm$  1.50%)  $p=0.001$  post-training training (Sikong, Lin et al. 2016). The studies mentioned thus far utilise the NBI platform, however positive effects from training can be seen when using high definition white light endoscopy and chromoendoscopy, with accuracy improving from 68% (95% CI 61-74) to 73% (95% CI 67-79)  $p<0.001$ , following a one-off training session (Basford, Longcroft-Wheaton et al. 2017). Improvements were also seen using BLI, with experienced endoscopists achieving a post-training accuracy of 97% (95% CI 95-100) from a baseline of 87% (95% CI 82-91)  $p<0.001$  (Subramaniam, Hayee et al. 2019).

Two studies that assessed the impact of computer-based learning in trainee endoscopists, also included experienced endoscopists and both showed statistically significant improvements in accuracy, 68% (95% CI 68-74) vs. 84% (95% CI 76-89)  $p<0.001$  (Ignjatovic, Thomas-Gibson et al. 2011) and 62.4% (95% CI 54.0-70.7) vs. 71.3% (95% CI 64.7-77.9)  $p<0.001$  (Basford, Brown et al. 2019). Improvements were not seen amongst experts in a USA based study, but importantly there were improvements amongst non-experts from an academic background (64% (56.9-70.6) vs. 80.5% (74.3-85.7)  $p<0.001$ ) and community background (64% (56.9-70.6) vs. 80.5% (74.3-85.7)  $p<0.001$ ) (Rastogi, Rao et al. 2014). A consistent improvement amongst community experienced endoscopists was also seen in a further study, achieving a post-training accuracy of 86.0% (95% CI 83.2-88.8) from 64.0% (95% CI 60.1-67.8)  $p<0.001$  (Sinh, Gupta et al. 2015). During the development and validation of the WASP (Workgroup

serrated polyps and Polyposis) classification, ten experienced endoscopists, most of which were accredited for bowel cancer screening, saw an improvement in optical diagnosis accuracy from 73% (95% CI 64-82) to 86% (95% CI 80-95)  $p<0.01$  (JE, Bastiaansen et al. 2016). Computer-based training on this polyp classification was also shown to increase accuracy in optical diagnosis in experienced endoscopists from 70.3% (95% CI 65-73.8) to 78.2% (95% CI 76.3-80.9) (Bae, Lee et al. 2019).

There has been one study comparing didactic and computer-based self-training, however there was no comparison between pre- and post-training performance (Khan, Cinnor et al. 2017). Despite this the results suggests there is no difference between the two training methods. To thoroughly assess whether one method is superior to the other, I will now present a randomised controlled study comparing didactic vs. computer-based self-training in the optical diagnosis of small/diminutive colorectal polyps using NBI and iScan OE platforms.

## 5.3 A randomised controlled study of the prediction of diminutive/small colorectal polyp histology using didactic vs. computer based self-learning module in gastroenterology trainees

### 5.3.1 Aims

The aim of this study was to compare the diagnostic performances of gastroenterology trainees at predicting histology of small/diminutive colonic polyps, following either face-to-face didactic training with an expert or computer-based self-learning, using different endoscopic platforms and endoscopic polyp classification scoring systems.

### 5.3.2 Methods

#### 5.3.2.1 Study design

Participants were randomised in a non-inferiority randomized controlled study comparing didactic vs. self-learning on diagnostic performances of trainees at predicting histology of diminutive/small polyps. The study was approved by the research ethics committee at the University of Birmingham, UK (ERN\_17-1370A).

#### 5.3.2.2 Participants

Participants from 6 centres in the Midlands, UK were invited to take part in the study that met the following eligibility criteria:

- Doctors in training without any endoscopic experience
- Gastroenterology trainees who have not yet completed training
- Ability to consent.

The training was completed at the University of Birmingham Medical School, UK.

#### 5.3.2.3 Video collection

Seventy-eight high quality videos (48 iScan-OE/30 NBI) of small/diminutive colonic polyps were selected from an existing video library, which were used in an earlier study whereby expert endoscopists achieved a NPV of 91% (78-98) using the SIMPLE classification following training (Iacucci, Trovato et al. 2018). The iScan videos showed individual polyps in high definition white light (HD-WL) and iScan-OE in different modes (iScan 1-3) without magnification. The NBI videos showed polyps in HD-WL and NBI, both without magnification. Each video was 30-90 seconds in duration and allowed individuals to pause the video to assess polyps in detail replicating real-life practice with the freeze button on the handpiece of endoscopic equipment. Two endoscopic platforms were chosen since trainees often encounter more than one endoscopic platform during their training and career. Therefore, training needs to reflect this and be validated in more than one platform.

#### *5.3.2.4 Pre-training assessment phase*

Prior to training participants viewed the 78 videos of small/diminutive colonic polyps and recorded the following observations on an Excel document (Microsoft Inc., Redmond, Washington, USA): Quality of Video (High/Low), NICE classification (Type 1, Type 2 or Type 3) (Figure 5-1), Confidence level (High/Low), SIMPLE classification (Type 1, Type 2a or Type 2b) (Figure 5-2) and Confidence level (High/Low).

#### *5.3.2.5 Intervention*

##### *5.3.2.5.1 Didactic training*

Training was conducted in a classroom for those participants randomised to receive didactic training, with training provided via a PowerPoint (Microsoft Inc., Redmond, Washington, USA) presentation by an expert endoscopist. The same expert endoscopist with extensive experience in optical characterisation in NBI and iScan platforms reviewed all teaching

material. Included in the presentation was an overview of the “Resect and Discard” strategy mentioned earlier in the chapter, the technology behind the endoscopic platforms NBI and iScan, NICE classification (Figure 5-1), SIMPLE classification (Figure 5-2) and example still images (n=43) and videos (n=8) of both classifications in use (Figure 5-3). A large number of still images were used to ensure participants had the best opportunity to observe and learn Kudo Pit Patterns (Kudo, Rubio et al. 2001) and other polyp features without movement artefact before observing videos, which are more challenging to interpret. Participants within this group had opportunity to ask questions and receive feedback in an interactive fashion. The trainer then discussed several videos demonstrating the use of NICE and SIMPLE in both NBI and i-Scan platforms with the histology being revealed to the participants. All lesions demonstrated in the videos and images had been resected and sent for histological confirmation. The training took approximately 1 hour to complete.

#### *5.3.2.5.2 Computer-based self-learning*

Participants randomised to the computer-based self-learning group were provided the same PowerPoint materials (Microsoft Inc., Redmond, Washington, USA) as the didactic group and completed the training in a separate room. Participants completed training without feedback interaction. They reviewed the same number of videos as the didactic group, which had guidance on the polyp features using the NICE and SIMPLE classifications.

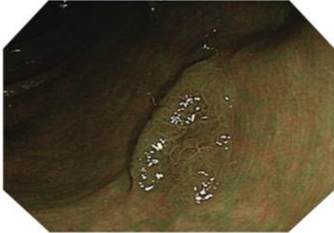
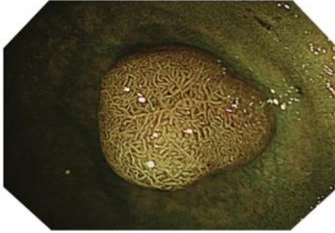
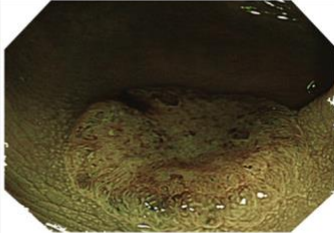
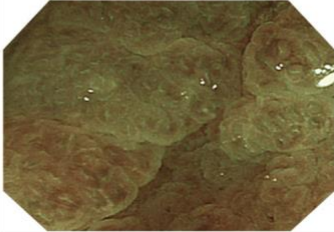
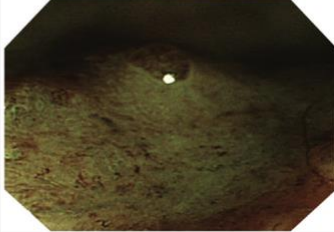
|                              | <b>Type 1</b>                                                                      | <b>Type 2</b>                                                                       | <b>Type 3</b>                                                                        |
|------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| <b>Color</b>                 | Same or lighter than background                                                    | Browner relative to background (verify color arises from vessels)                   | Brown to dark brown relative to background; sometimes patchy whiter areas            |
| <b>Vessels</b>               | None, or isolated lacy vessels coursing across the lesion                          | Brown vessels surrounding white structures**                                        | Has area(s) of disrupted or missing vessels                                          |
| <b>Surface Pattern</b>       | Dark or white spots of uniform size, or homogeneous absence of pattern             | Oval, tubular or branched white structure surrounded by brown vessels**             | Amorphous or absent surface pattern                                                  |
| <b>Most likely pathology</b> | <b>Hyperplastic</b>                                                                | <b>Adenoma***</b>                                                                   | <b>Deep submucosal invasive cancer</b>                                               |
| <b>Examples</b>              |   |   |   |
|                              |  |  |  |

**Figure 5-1 NICE (NBI International Colorectal Endoscopic) Classification**

Taken from (Hewett, Kaltenbach et al. 2012)

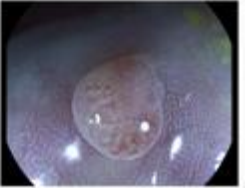
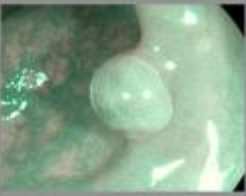
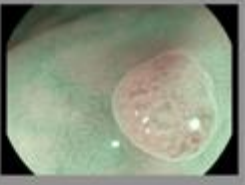
| Non-neoplastic features |                                                                                                         | Neoplastic features                                                                                            |                                                        |                                                                                      |
|-------------------------|---------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------------------------------------------|
|                         | HYPERPLASTIC POLYP                                                                                      | sessile serrated adenoma/ lesion                                                                               | ADENOMA                                                |                                                                                      |
| SURFACE PATTERN         | ROUND PIT<br>          | OPEN/DILATED (DARK) PIT<br>  | NOT ROUND STRUCTURE:<br>-OVAL<br>-TUBULAR<br>-BRANCHED |   |
| VESSEL PATTERN          | NONE ISOLATED LACY<br> | NONE ISOLATED LACY<br>       | THICK VESSELS                                          |   |
| LESION BORDER           | REGULAR<br>           | IRREGULAR/INDISTINCTIVE<br> | REGULAR                                                |  |

**Figure 5-2 SIMPLE (Simplified Identification Method for Polyp Labelling during the Endoscopy) classification**

Taken from (Iacucci, Trovato et al. 2018)

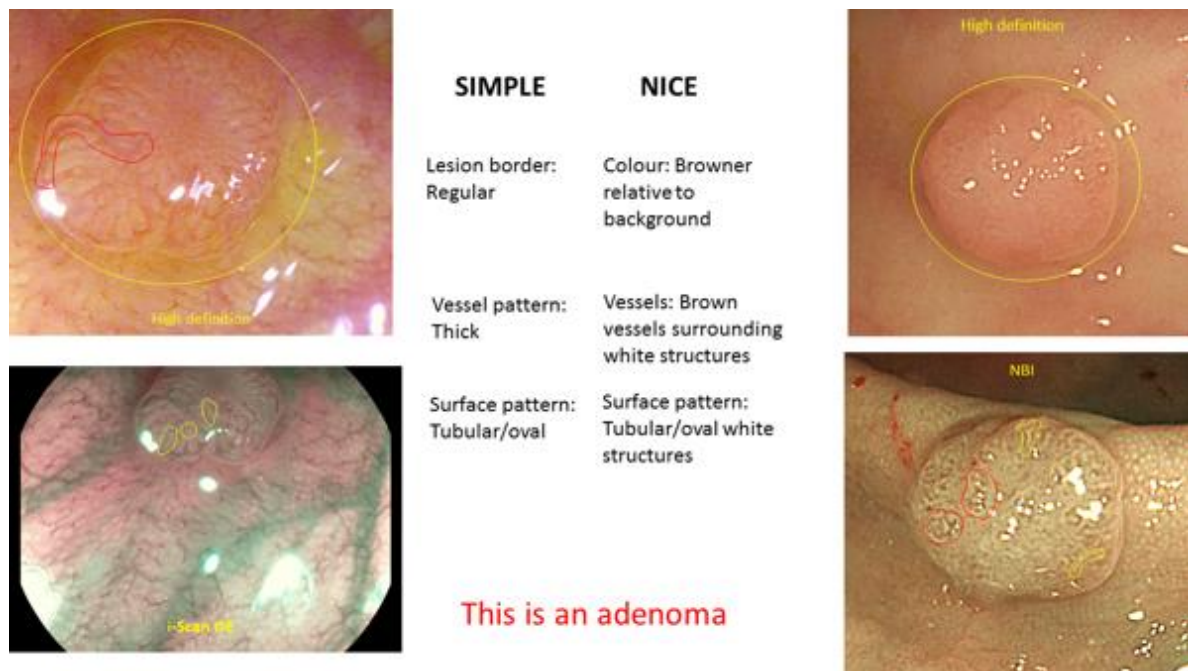


**Figure 5-3 Example of training material used in the training module illustrating polyp features included in the NICE and SIMPLE classifications**

#### 5.3.2.6 Post-training assessment

Following training, participants completed a post-training assessment on the same day. These were the same 78 videos as the pre-training assessment in a different random order to reduce recall bias. Participants completed the same observations as per the pre-training assessment.

#### 5.3.2.7 Randomisation

Each participant was allocated a computer-generated random number on Microsoft Excel (Microsoft Inc., Redmond, Washington, USA) following which computerised randomisation to either computer-based self-learning or didactic training took place at a 1:1 ratio. Due to the nature of the study blinding of participants was not possible.

#### *5.3.2.8 Study outcomes*

The outcome measures included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy of polyp histology predictions. In addition, other outcome measures included proportions of high confidence predictions and inter-observer agreement.

#### *5.3.2.9 Sample size*

Assuming a non-inferiority trial with one-sided distribution (face-to-face training vs. computer-based training) and power of 90% to detect a 5% difference in accuracy, the sample size required is 375 observations (one video=one observation per participant). In using 78 videos, a minimum of 5 participants in each arm was required.

If calculating sample size independently for either modality, NBI and iScan-OE, a sample size of 750 would be needed. Sixteen trainees were recruited giving a total number of 1248 observations, therefore achieving necessary sample size to reach 90% power.

#### *5.3.2.10 Statistical analysis*

All data was collected on Microsoft Excel (Microsoft Inc., Redmond, Washington, USA), and participants were allocated a study identification code to allow tracking of results from pre- to post-training. Predictions of polyp histology were compared with histological results as gold standard. Comparison between groups were made using Fisher's Exact Test. Inter-observer agreement was quantified using Fleiss' Kappa. This is an analogue to Cohen's Kappa for when more than two raters are used. Confidence intervals and p values were calculated using a bootstrap approach with 1000 iterations. Participant characteristics were analysed using a Wilcoxon-Rank sum test and Fisher's exact test. A p-value of <0.05 was considered statistically significant.

### 5.3.3 Results

Sixteen trainees (12 gastroenterology trainees and 4 endoscopically-naïve trainees) participated in the study with 8 trainees (6 gastroenterology trainees and 2 naïve trainees) randomised to receive computer-based self-learning and 8 trainees (6 gastroenterology trainees and 2 naïve trainees) to receive didactic training. Baseline characteristics for participants are shown in Table 5-1.

**Table 5-1 Baseline characteristics of participants in the randomisation-controlled study of optical diagnosis training**

|                                               | Didactic training | Computer-based self-training | P-value |
|-----------------------------------------------|-------------------|------------------------------|---------|
| Gastroenterology years of experience (median) | 3 (0-5)           | 3 (0-5)                      | 0.705   |
| Number of colonoscopies in lifetime (median)  | 145 (0-360)       | 105 (0-600)                  | 0.958   |
| NBI experience (%)                            | 50% (4/8)         | 37.5% (3/8)                  | 1       |
| iScan experience (%)                          | 12.5% (1/8)       | 12.5 (1/8)                   | 1       |

There was no statistically significant difference in the prior endoscopic experience of each group. No participants were withdrawn from the study, and all completed the pre- and post-training assessments.

Performances of the naïve and trainee endoscopists pre- and post-training when using NICE and SIMPLE classification are shown in Table 5-2.

**Table 5-2 Pre-training vs. Post-training optical diagnosis performance using the NICE and SIMPLE classification**

| SIMPLE Pre-Training |                      |                        |         | SIMPLE Post-Training |                        |         |
|---------------------|----------------------|------------------------|---------|----------------------|------------------------|---------|
|                     | Naïve<br>(%, 95% CI) | Trainee<br>(%, 95% CI) | P-Value | Naïve<br>(%, 95% CI) | Trainee<br>(%, 95% CI) | P-Value |
| Sensitivity         | 72 (63-78)           | 74 (70-78)             | 0.479   | 76 (68-82)           | 82 (78-85)             | 0.093   |
| Specificity         | 50 (42-58)           | 60 (55-65)             | 0.056   | 57 (49-66)           | 50 (45-54)             | 0.115   |
| PPV                 | 63 (56-70)           | 69 (65-73)             | 0.144   | 69 (61-75)           | 67 (63-70)             | 0.657   |
| NPV                 | 59 (49-68)           | 65 (60-70)             | 0.223   | 66 (56-74)           | 69 (63-74)             | 0.489   |
| Accuracy            | 62 (56-67)           | 68 (64-71)             | 0.067   | 67 (62-73)           | 67 (64-70)             | 1       |
| NICE Pre-Training   |                      |                        |         | NICE Post-Training   |                        |         |
| Sensitivity         | 63 (56-70)           | 71 (67-75)             | 0.057   | 74 (67-81)           | 79 (75-82)             | 0.248   |
| Specificity         | 64 (55-72)           | 62 (57-66)             | 0.687   | 55 (46-63)           | 55 (51-60)             | 0.922   |
| PPV                 | 69 (61-76)           | 69 (65-73)             | 0.845   | 67 (60-74)           | 69 (65-72)             | 0.721   |
| NPV                 | 58 (50-66)           | 64 (59-68)             | 0.280   | 64 (54-72)           | 68 (62-73)             | 0.433   |
| Accuracy            | 64 (58-69)           | 67 (64-70)             | 0.299   | 66 (60-71)           | 68 (65-71)             | 0.402   |

The performance in predicting histology in both groups (didactic and computer-based self-learning) are shown in Table 5-3 and Table 5-4. Following training the proportion of predictions made with high confidence was higher in the computer-based self-learning group when using both the NICE 77% (73.2-80.4% 95% CI) and SIMPLE classifications 77.1% (73.4-80.3 95% CI) than the didactic group 69.8% (66-73.4% 95% CI;  $p < 0.05$ ) and 69.9% (66.1-73.5% 95% CI;  $p < 0.05$ ) respectively. When comparing performances, the didactic group demonstrated a higher sensitivity of 83.1% (78.7-86.9% 95% CI) over the computer

group 72% (66.9-76.6% 95%CI) when using the NICE classification. There was no statistical difference between the two groups in other performance measures.

When comparing the inter-observer agreement (table 5.5), it was clear that training improves the agreement when using the SIMPLE classification, from 0.35 (0.29-0.42 95% CI) to a moderate agreement of 0.52 (0.45-0.61 95% CI;  $p < 0.0001$ ).

**Table 5-3 Comparison of the performance at predicting small/diminutive polyp histology between didactic and computer-based training**

|        |                             | Didactic training %<br>(95% CI) | Computer-based self-training %<br>(95% CI) | P value |
|--------|-----------------------------|---------------------------------|--------------------------------------------|---------|
| SIMPLE | Sensitivity                 | 83 (78-86)                      | 78 (73-82)                                 | 0.148   |
|        | Specificity                 | 52 (46-58)                      | 51 (48-57)                                 | 0.735   |
|        | PPV                         | 68 (63-72)                      | 66 (61-71)                                 | 0.551   |
|        | NPV                         | 71 (64-77)                      | 65 (58-72)                                 | 0.249   |
|        | Accuracy                    | 69 (65-73)                      | 66 (62-69)                                 | 0.225   |
|        | High confidence predictions | 70 (66-74)                      | 77 (73-80)                                 | 0.005   |
| NICE   | Sensitivity                 | 83 (79-87)                      | 72 (67-77)                                 | 0.0005  |
|        | Specificity                 | 52 (46-58)                      | 59 (53-65)                                 | 0.106   |
|        | PPV                         | 68 (63-73)                      | 68 (63-73)                                 | 0.939   |
|        | NPV                         | 71 (65-77)                      | 63 (57-69)                                 | 0.059   |
|        | Accuracy                    | 69 (65-73)                      | 66 (62-69)                                 | 0.275   |
|        | High confidence predictions | 70 (66-73)                      | 77 (73-80)                                 | 0.006   |

**Table 5-4 Performances using NICE and SIMPLE classification in the didactic training and computer-based training groups**

| SIMPLE-Didactic Training            |                         |                          |         |
|-------------------------------------|-------------------------|--------------------------|---------|
|                                     | Pre-training % (95% CI) | Post-training % (95% CI) | P-value |
| Sensitivity                         | 75 (70-80)              | 83 (78-86)               | 0.018   |
| Specificity                         | 61 (55-67)              | 52 (46-58)               | 0.031   |
| PPV                                 | 70 (65-75)              | 68 (63-73)               | 0.532   |
| NPV                                 | 68 (60-72)              | 71 (64-77)               | 0.361   |
| Accuracy                            | 69 (65-73)              | 69 (65-73)               | 1       |
| High confidence                     | 46 (41-50)              | 70 (66-74)               | <0.001  |
| NICE-Didactic Training              |                         |                          |         |
| Sensitivity                         | 76 (71-80)              | 83 (79-87)               | 0.018   |
| Specificity                         | 60 (54-66)              | 52 (46-58)               | 0.073   |
| PPV                                 | 70 (65-75)              | 68 (63-73)               | 0.591   |
| NPV                                 | 66 (60-72)              | 71 (65-73)               | 0.264   |
| Accuracy                            | 69 (65-72)              | 69 (65-73)               | 0.807   |
| High confidence                     | 55 (49-60)              | 70 (66-73)               | <0.001  |
| SIMPLE-Computer-based self-training |                         |                          |         |
| Sensitivity                         | 72 (66-76)              | 78 (73-82)               | 0.063   |
| Specificity                         | 54 (48-60)              | 51 (45-57)               | 0.498   |
| PPV                                 | 65 (60-70)              | 66 (61-71)               | 0.880   |
| NPV                                 | 61 (54-67)              | 65 (59-72)               | 0.334   |
| Accuracy                            | 64 (60-67)              | 66 (62-69)               | 0.439   |
| High confidence                     | 50 (46-54)              | 77 (73-80)               | <0.001  |
| NICE- Computer-based self-training  |                         |                          |         |
| Sensitivity                         | 63 (58-68)              | 72 (67-77)               | 0.012   |
| Specificity                         | 65 (59-70)              | 59 (53-65)               | 0.164   |
| PPV                                 | 69 (63-74)              | 68 (63-73)               | 1       |
| NPV                                 | 59 (53-64)              | 63 (57-69)               | 0.343   |
| Accuracy                            | 64 (60-67)              | 66 (62-70)               | 0.373   |
| High confidence                     | 61 (57-65)              | 77 (73-80)               | <0.001  |

**Table 5-5 Inter-observer agreement comparison between SIMPLE and NICE pre- vs. post-training**

|        | Pre-training | 95% CI      | Post-training | 95% CI      | P-value |
|--------|--------------|-------------|---------------|-------------|---------|
| SIMPLE | 0.349        | 0.286-0.417 | 0.523         | 0.447-0.612 | <0.0001 |
| NICE   | 0.295        | 0.231-0.354 | 0.346         | 0.298-0.464 | 0.168   |

### 5.3.4 Discussion

This study demonstrates that self-learning training can be effective for the prediction of diminutive/small polyp histology. This training method can achieve results similar to the more labour intensive and expensive didactic training method. This study is the first in the literature to compare two classification systems, NICE and SIMPLE, and the impact of a training module on both in terms of diagnostic performance. Secondly, using two endoscopic platforms (NBI and iScan-OE), is a first in the literature and differs from the Khan *et al* paper (Khan, Cinnor et al. 2017). This is the first study comparing didactic training with computer-based training using iScan-OE platform and the newly developed SIMPLE classification of small/diminutive colonic polyps. This is particularly important, as clinicians will have access to different endoscopic platforms (Olympus, Pentax and Fujifilm). Therefore, when designing a training module, it needs to be effective for several platforms and restricting to one platform means results cannot be generalised. Another strength of the present study was the number of observations made, both in the pre-training and post-training assessments. Sixteen participants assessed 78 videos giving a total of 1248 observations, allowing sufficient power to investigate for any difference between the two groups, and also independently for NBI and iScan-OE platforms, as the two platforms may have similarities as well as differences in the operating characteristics of training and inter-observer agreement. This is significantly more than the 680 observations made by the Khan et al paper (Khan, Cinnor et al. 2017). Further to this, participants completed a pre-training assessment before receiving either computer-based self-learning or didactic training, followed by a post-training assessment, which allowed the true impact of the training module in both modalities on the performance of participants to be demonstrated. Videos of polyps were used which allowed participants to pause the video,

similar to holding or taking a picture during a real colonoscopy examination thereby allowing assessments from several angles, reproducing real-world performance.

Interactive still images (annotated with arrows and circles) gave trainees the best opportunity to observe and learn mucosal and vascular patterns and polyp characteristics using NICE and SIMPLE classifications systems without movement artefact. This was used to gain a baseline level of knowledge before testing on videos, which is more challenging with the polyp moving and more difficult to standardize. However, videos were not solely used in the training as it takes time to observe videos and the aim was to ensure training could be delivered within 1 hour to ensure maximum effective learning and efficiency.

Didactic training is an attractive method since it allows the opportunity to ask questions to an expert endoscopist and receive feedback, with studies demonstrating it can be effective (Higashi, Uraoka et al. 2010, Bouwens, de Ridder et al. 2013, Patel, Rastogi et al. 2013).

However, it is resource intensive, time consuming and expensive, which means this method will unlikely be able to train significant numbers of non-experts. Computer-based learning is a common method of training in healthcare and is relatively inexpensive, not resource intensive and can be delivered to a large number of participants in multiple countries. There is growing evidence demonstrating it can be an effective method of training in optical diagnosis (Ignjatovic, Thomas-Gibson et al. 2011, Ladabaum, Fioritto et al. 2013, Rastogi, Rao et al. 2014). The drawbacks to this method are the lack of feedback possible and the inability to ask questions.

We demonstrated that the computer-based self-learning group predicted histology with higher confidence, using both NICE and SIMPLE classifications with the number of high confidence observations increasing following training in both classifications (SIMPLE and NICE). There may be an element of self-satisfaction associated with self-learning, whereas having direct

feedback on polyp characteristics that participants may not have acknowledged may reduce confidence levels, as may be the case in the didactic group. There were elements of feedback in the self-training group in that histology was revealed with explanations using the NICE and SIMPLE classification. The fact that performances were similar in both groups highlights that the role of direct feedback face-to-face is less pivotal as was once anticipated. This will be incorporated into self-training as tested in this study.

In terms of diagnostic performance, the didactic group demonstrated a higher sensitivity at differentiating small/diminutive polyps when using the NICE classification. Otherwise, there was no statistical difference between the two groups. This further supports the findings from *Khan et al* (Khan, Cinnor et al. 2017), and shows promise that computer-based self-learning can have a role in training. Importantly, the NPV in both groups failed to reach the PIVI threshold, demonstrating that training modules, whilst having a role, should not be used in isolation but should be considered an important component of teaching. The inter-observer agreement improved following training when using both NICE and SIMPLE classifications, with SIMPLE having a higher kappa agreement over NICE classification.

There are some limitations to this study. Firstly, the same videos were used in the pre- and post-training assessment in a random order to reduce recall bias. However, in using 78 videos the impact of this would be minimal as it allowed us to increase the number of observations made. In order to minimise this, different sets of videos matched for histology and endoscopic platform would need to be used which would need a large library of videos. These results cannot be generalised to Blue-light Laser Imaging (BLI) and other classifications systems such as BASIC (BLI Adenoma Serrated International Classification) (Bisschops, Hassan et al. 2018). We have not used the NBI Expert Team (JNET) classification (Sano, Tanaka et al. 2016) which has been demonstrated to characterize polyps using magnification with high

accuracy (Sumimoto, Tanaka et al. 2017). However optical zoom magnifying endoscopes are not widely used in clinical practice in Western countries. Therefore, we have not used magnifying images/videos in our training materials to replicate the endoscopic platforms that are likely to be encountered on a daily basis. The newly developed near focus with electronic zoom endoscope system (Exera III and Lucera Elite, Olympus) can now provide similar images and have been increasingly adopted in Western countries. This will enable implementation of the use of the JNET classification in the future and training modules will need adaptation.

In conclusion, computer-based self-training is as effective as didactic training. This gives promise to the widespread delivery of effective training to colonoscopists, improving the prospect of a “resect and discard” strategy. Computer-based self-learning is a training method that many trainees are familiar with its use in training. Its main advantages are that it is low cost and its ease of delivery. While individual feedback cannot be delivered as per didactic training, well-constructed explanations of lesions and the use of classification systems can allow for this. Further studies should investigate if a combination of training modules in a stepwise approach might be the right future strategy into how to best achieve the PIVI thresholds, which may include training using live endoscopy.

The section above is taken from the following publication:

**Smith SCL**, Saltzman J, Shivaji UN, Lethebe BC, Cannatelli R, Birmingham Colonic Polyp Characterisation Group, Ghosh S, Iacucci M. A randomised controlled study of the prediction of diminutive/small colorectal polyp histology using didactic vs. computer based self-learning module in gastroenterology trainees. *Dig Endoscopy* 2019; 31: 535-543.

## 5.4 Training methods in optical diagnosis and characterisation of colorectal polyps: a systematic review and meta-analysis

### 5.4.1 Introduction:

In the previous section I have presented findings that would suggest that there is very little difference between didactic and computer-based training on the optical diagnosis performance amongst trainees and naïve endoscopists. The importance of endoscopy training has become increasingly relevant following the coronavirus disease 2019 (COVID-19) pandemic, which has had an adverse impact on endoscopy training for trainees, particularly hands-on training, with a reduction in procedures of up to 96% (Siau, Iacucci et al. 2020). Societies are now recommending trainees utilise alternative learning opportunities such as cognitive-based learning (Keswani, Sethi et al. 2020).

Therefore, it is important that optical diagnosis training is effective. The aim of this systematic review and meta-analysis is to provide an overview of training in optical diagnosis of colorectal polyps and in view of the COVID-19 pandemic complete sub-group analysis of computer-based self-training.

#### 5.4.2 METHODS:

Methodology of this analysis, inclusion criteria and reporting were in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations (Moher, Liberati et al. 2009) and the Meta-analyses Of Observational Studies in Epidemiology (MOOSE) checklist (Stroup, Berlin et al. 2000). This systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO) with registration number CRD42020167486.

##### 5.4.2.1 Search strategy:

An electronic database search of PUBMED/Medline, Cochrane and SCOPUS databases was performed in addition to grey literature (scanning reference lists), to identify studies reporting training for the optical diagnosis of colorectal polyps between 1980 and October 2019. The search strategy is outlined below:

Search strategy in PUBMED/Medline:

*(Optical Diagnosis OR optical biopsy OR optical characterisation) AND (colorectal polyp OR colorectal adenoma OR colorectal hyperplastic OR colorectal sessile serrated lesion OR colorectal sessile serrated adenoma), (Optical Diagnosis OR optical biopsy OR optical characterisation) AND (colorectal polyp OR colorectal adenoma OR colorectal hyperplastic OR colorectal sessile serrated lesion OR colorectal sessile serrated adenoma) AND (training OR education), (Education OR Training) AND Colonoscopy AND (Colorectal polyp OR Colorectal adenoma OR Colorectal hyperplastic OR Colorectal sessile serrated lesion).*

Search strategy in the other databases followed a similar but simplified strategy.

#### *5.4.2.2 Inclusion and exclusion criteria:*

Studies were deemed eligible for inclusion according to the PICO statement (P, endoscopists undergoing assessment of optical diagnosis accuracy of small/diminutive colorectal polyps; I, endoscopists receiving optical diagnosis training; C, optical diagnosis as compared with histological result as gold standard; O, pre- vs. post-training accuracy of optical diagnosis).

Studies not reporting pre- vs. post-training performance and not published in English language were excluded. Randomised-controlled trials, observational and cohort studies and abstracts were all included for analysis.

#### *5.4.2.3 Study selection:*

Titles and abstracts of all identified articles were independently screened by two authors to exclude studies not related to the topic or not meeting inclusion criteria. Potentially relevant studies were screened for eligibility by analysis of the full text. Disagreements between the two authors were referred to and discussed with the senior author and resolved with consensus.

#### *5.4.2.4 Data extraction and quality assessment:*

A standardized form was used to extract the data from each study. Data extracted included: a) Author name, b) Year of publication, c) Country, d) Training method (didactic, computer-based), e) Number of participants, f) Setting of study (*in vivo/ex vivo*), g) Number and size of polyps, h) Number of correct histology predictions (accuracy), i) Training material, j) Endoscopic platform and k) Duration of training.

The risk of bias of included studies was assessed using the Cochrane Collaboration's tool (Higgins, Altman et al. 2011). Each study was assessed for risk of bias through study design in selection bias (random sequence generation and methods to conceal allocation), performance bias (blinding of participants), detection bias (blinding of outcomes), attrition bias (completeness of outcome data) and reporting bias (selective reporting). The overall quality of

evidence was summarized using The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, categorizing the evidence as very low, low, moderate or high-quality.

#### *5.4.2.5 Study definitions:*

A small/diminutive colorectal polyp was defined as  $\leq 10$ mm in size. Optical diagnosis training was defined as an intervention designed to educate participants on optical diagnosis methodology. Experienced endoscopists were participants who are independent endoscopists who have completed endoscopy training but who are not considered experts. Trainee endoscopists were defined as participants who are practicing within a designated gastroenterology training programme.

#### *5.4.2.6 Study outcomes:*

The primary outcome of our systematic review and meta-analysis was optical diagnosis accuracy compared with histological analysis before and after training intervention. We aimed to complete sub-group analysis by endoscopists experience (experienced, trainee endoscopists), training method (didactic, computer-based) and polyp size (small/diminutive polyps).

#### *5.4.2.7 Statistical analysis:*

A random effect meta-analysis was performed to investigate the effect of training on the accuracy of optical diagnosis. For each study, accuracy was compared between post-training and pre-training stages and expressed as a risk ratio (RR) with 95% confidence intervals (CIs), which were pooled using a random-effects Mantel-Haenszel model. Forest plots were generated for all studies, followed by level of experience (trainee vs experienced endoscopists), and then for the subgroups of didactic vs. computer-based training. 95% confidence intervals were also calculated to determine the variation in effect between studies (Riley, Higgins et al. 2011, Deeks 2020). Statistical heterogeneity was assessed using  $I^2$

statistics, with a value of 0-40% accepted as not important, 30-60% as moderate, >60% as substantial and >90% as considerable heterogeneity (Higgins, Altman et al. 2011).

Publication bias was assessed by observing asymmetry in funnel plots and sensitivity analyses performed by excluding outliers and then by year of publication.

Meta-analyses were performed using RevMan v5.3 (Cochrane Collaboration, Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014) and pooled effects of didactic and computer-based training subjected to a random effects meta-regression model using Open Meta-Analyst (Brown University). All P-values were considered significant at <0.05.

### 5.4.3 RESULTS:

#### 5.4.3.1 Study selection and characteristics:

The PRISMA (Moher, Liberati et al. 2009) (Table 5-6), MOOSE (Stroup, Berlin et al. 2000) checklist (Table 5-7) and PRISMA flow chart (Figure 5-4) were followed to ensure compliance. The literature search yielded 1237 results. After preliminary screening of titles and abstracts, 113 were selected for full text review. Of these, 16 papers (Rogart, Jain et al. 2008, Higashi, Uraoka et al. 2010, Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Coe, Thomas et al. 2012, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Sikong, Lin et al. 2016, Basford, Longcroft-Wheaton et al. 2017, Aihara, Kumar et al. 2018, Iacucci, Trovato et al. 2018, Bae, Lee et al. 2019, Basford, Brown et al. 2019, Smith, Saltzman et al. 2019, Subramaniam, Hayee et al. 2019) (before 2010 n=1, 2010-2015 n=7, 2016-2019 n=8) matched the selection criteria and were included in the systematic review. Of these studies, seven were from Europe (Ignjatovic, Thomas-Gibson et al. 2011, JE, Bastiaansen et al. 2016, Basford, Longcroft-Wheaton et al. 2017, Iacucci, Trovato et al. 2018, Basford, Brown et al. 2019, Smith, Saltzman et al. 2019, Subramaniam, Hayee et al. 2019), six from North America (Rogart, Jain et al. 2008, Raghavendra, Hewett et al. 2010, Coe, Thomas et al. 2012, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, Aihara, Kumar et al. 2018) and three from Asia (Higashi, Uraoka et al. 2010, Sikong, Lin et al. 2016, Bae, Lee et al. 2019). Among included studies, eight reported on didactic training (Rogart, Jain et al. 2008, Higashi, Uraoka et al. 2010, Raghavendra, Hewett et al. 2010, Coe, Thomas et al. 2012, Sikong, Lin et al. 2016, Basford, Longcroft-Wheaton et al. 2017, Iacucci, Trovato et al. 2018, Subramaniam, Hayee et al. 2019), seven on computer-based self-training (Ignjatovic, Thomas-Gibson et al. 2011, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Aihara, Kumar et al. 2018, Bae, Lee et al. 2019, Basford, Brown et al. 2019) and one on computer-based self-training vs. didactic training (Smith, Saltzman et al. 2019). The majority of studies were observational in design (n=14) and there were two randomised trials (Coe, Thomas et al. 2012, Smith, Saltzman et al. 2019). Overall 11 studies were based on NBI system (Rogart, Jain et al. 2008, Higashi, Uraoka et al. 2010, Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Coe, Thomas et al. 2012, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Sikong, Lin et al. 2016, Aihara, Kumar et al. 2018, Bae, Lee et al. 2019), one on iScan (Basford, Brown et al. 2019), one on BLI (Blue Light Imaging) (Subramaniam, Hayee et al. 2019), two on NBI and iScan (Iacucci, Trovato et al. 2018, Smith, Saltzman et al. 2019) and one on High-definition white light (HDWL)/chromoendoscopy (Basford, Longcroft-Wheaton et al. 2017). Sixteen studies reported pre- and post-training values, two studies included trainees (Iacucci, Trovato et al. 2018, Smith, Saltzman et al. 2019), eight included fully qualified/BCSP (Bowel Cancer Screening Programme) endoscopists (Rogart, Jain et al. 2008, Higashi, Uraoka et al. 2010, Coe, Thomas et al. 2012, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Sikong, Lin et al. 2016, Bae, Lee et al. 2019) and six studies included both groups (Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Basford, Longcroft-Wheaton et al. 2017, Aihara, Kumar et al. 2018, Basford, Brown et al. 2019, Subramaniam, Hayee et al. 2019). There were nine studies that only included small/diminutive colorectal polyps (Higashi, Uraoka et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, Iacucci, Trovato et al. 2018, Bae, Lee et al. 2019, Basford, Brown et al. 2019, Smith, Saltzman et al. 2019, Subramaniam, Hayee et al. 2019). Study characteristics are presented in

Table 5-8.

**Table 5-6 PRISM checklist**

|                                    | #  | Checklist item                                                                                                                                                                                                                                                                                              |
|------------------------------------|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title                              | 1  | Identify the report as a systematic review, meta-analysis, or both.                                                                                                                                                                                                                                         |
| Structured summary                 | 2  | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. |
| Rationale                          | 3  | Describe the rationale for the review in the context of what is already known.                                                                                                                                                                                                                              |
| Objectives                         | 4  | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).                                                                                                                                                  |
| Protocol and registration          | 5  | Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.                                                                                                                               |
| Eligibility criteria               | 6  | Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.                                                                                                      |
| Information sources                | 7  | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.                                                                                                                                  |
| Search                             | 8  | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.                                                                                                                                                                               |
| Study selection                    | 9  | State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).                                                                                                                                                   |
| Data collection process            | 10 | Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.                                                                                                                                  |
| Data items                         | 11 | List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.                                                                                                                                                                       |
| Risk of bias in individual studies | 12 | Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.                                                                                      |
| Summary measures                   | 13 | State the principal summary measures (e.g., risk ratio, difference in means).                                                                                                                                                                                                                               |
| Synthesis of results               | 14 | Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.                                                                                                                                                   |
| Risk of bias across studies        | 15 | Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).                                                                                                                                                                |
| Additional analyses                | 16 | Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.                                                                                                                                                            |
| Study selection                    | 17 | Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.                                                                                                                                             |
| Study characteristics              | 18 | For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.                                                                                                                                                                |

|                               |    |                                                                                                                                                                                                          |
|-------------------------------|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Risk of bias within studies   | 19 | Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).                                                                                                |
| Results of individual studies | 20 | For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. |
| Synthesis of results          | 21 | Present results of each meta-analysis done, including confidence intervals and measures of consistency.                                                                                                  |
| Risk of bias across studies   | 22 | Present results of any assessment of risk of bias across studies (see Item 15).                                                                                                                          |
| Additional analysis           | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).                                                                                    |
| Summary of evidence           | 24 | Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).                     |
| Limitations                   | 25 | Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).                                            |
| Conclusions                   | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research.                                                                                  |
| Funding                       | 27 | Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.                                                               |

**Table 5-7 MOOSE Checklist**

| Item No | Recommendation                                                                                                                                                                                                                                                                 |
|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1       | Problem definition                                                                                                                                                                                                                                                             |
| 2       | Hypothesis statement                                                                                                                                                                                                                                                           |
| 3       | Description of study outcome(s)                                                                                                                                                                                                                                                |
| 4       | Type of exposure or intervention used                                                                                                                                                                                                                                          |
| 5       | Type of study designs used                                                                                                                                                                                                                                                     |
| 6       | Study population                                                                                                                                                                                                                                                               |
| 7       | Qualifications of searchers (e.g., librarians and investigators)                                                                                                                                                                                                               |
| 8       | Search strategy, including time period included in the synthesis and key words                                                                                                                                                                                                 |
| 9       | Effort to include all available studies, including contact with authors                                                                                                                                                                                                        |
| 10      | Databases and registries searched                                                                                                                                                                                                                                              |
| 11      | Search software used, name and version, including special features used (e.g., explosion)                                                                                                                                                                                      |
| 12      | Use of hand searching (e.g., reference lists of obtained articles)                                                                                                                                                                                                             |
| 13      | List of citations located and those excluded, including justification                                                                                                                                                                                                          |
| 14      | Method of addressing articles published in languages other than English                                                                                                                                                                                                        |
| 15      | Method of handling abstracts and unpublished studies                                                                                                                                                                                                                           |
| 16      | Description of any contact with authors                                                                                                                                                                                                                                        |
| 17      | Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested                                                                                                                                                                     |
| 18      | Rationale for the selection and coding of data (e.g., sound clinical principles or convenience)                                                                                                                                                                                |
| 19      | Documentation of how data were classified and coded (e.g., multiple raters, blinding and interrater reliability)                                                                                                                                                               |
| 20      | Assessment of confounding (e.g., comparability of cases and controls in studies where appropriate)                                                                                                                                                                             |
| 21      | Assessment of study quality, including blinding of quality assessors, stratification or regression on possible predictors of study results                                                                                                                                     |
| 22      | Assessment of heterogeneity                                                                                                                                                                                                                                                    |
| 23      | Description of statistical methods (e.g., complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated |
| 24      | Provision of appropriate tables and graphics                                                                                                                                                                                                                                   |
| 25      | Graphic summarizing individual study estimates and overall estimate                                                                                                                                                                                                            |
| 26      | Table giving descriptive information for each study included                                                                                                                                                                                                                   |
| 27      | Results of sensitivity testing (eg, subgroup analysis)                                                                                                                                                                                                                         |
| 28      | Indication of statistical uncertainty of findings                                                                                                                                                                                                                              |



**Figure 5-4 PRISMA Flow Chart**

| Author, Year      | Study design* | Type of training            | Type of participant      | Number of participants | Training duration                                            | Training material          | Platform         | Number of polyps post-training |
|-------------------|---------------|-----------------------------|--------------------------|------------------------|--------------------------------------------------------------|----------------------------|------------------|--------------------------------|
| Rogart, 2008      | P, O          | Didactic                    | Experienced              | 4                      | 1 hour                                                       | Ex vivo, Image             | NBI              | 20                             |
| Higashi, 2010     | P, O          | Didactic                    | Experienced              | 4                      | 1 hour                                                       | Ex vivo, image-based       | NBI              | 44                             |
| Raghavendra, 2010 | P, O          | Didactic                    | Experienced and trainee  | 25                     | 20 minutes                                                   | Ex vivo, image-based       | NBI              | 25                             |
| Ignjatovic, 2011  | P, O          | Computer-based              | Experienced and trainee  | 14                     | 15 minutes                                                   | Ex vivo, image-based       | NBI              | 30                             |
| Coe, 2012         | RCT           | Didactic                    | Experienced              | 15                     | 2x 1-hour sessions                                           | Ex vivo, images and videos | NBI              | 774                            |
| Rastogi, 2014     | P, O          | Computer-based              | Experienced              | 10                     | 20 minutes                                                   | Ex vivo, images            | NBI              | 40                             |
| Sinh, 2015        | P, O          | Computer-based              | Experienced              | 15                     | 20 minutes                                                   | Ex vivo, image-based       | NBI              | 40                             |
| IJspeert, 2016    | P, O          | Computer-based              | Experienced              | 10                     | 20 minutes                                                   | Ex vivo, images            | NBI              | 45                             |
| Sikong, 2016      | P, O          | Didactic                    | Experienced              | 10                     | 3 x 1-hour sessions over 3 months                            | Ex vivo, image-based       | NBI              | 130                            |
| Basford, 2017     | P, O          | Didactic                    | Experienced and trainee  | 10                     | 30 minutes                                                   | Ex vivo, image-based       | HD WLE + DCE     | 37                             |
| Aihara, 2018      | P, O          | Computer-based              | Experienced and trainee  | 8                      | 10 minutes                                                   | Ex vivo, image-based       | NBI              | 50                             |
| Iacucci, 2018     | P, O          | Didactic                    | Trainee                  | 6                      | 1 hour                                                       | Ex vivo, images and videos | iScan OE and NBI | 80                             |
| Bae, 2019         | P, O          | Computer-based              | Experienced              | 15                     | 30 minutes, weekly feedback and interim interactive training | Ex vivo, image-based       | NBI              | 80                             |
| Basford, 2019     | P, O          | Computer-based              | Experienced and trainees | 14                     | 20 minutes                                                   | Ex vivo, image-based       | iScan and DCE    | 30                             |
| Subramaniam, 2019 | P, O          | Didactic                    | Experienced and trainees | 10                     | 4 hours                                                      | Ex vivo, image-based       | BLI              | 45                             |
| Smith, 2019       | RCT           | Computer-based and didactic | Trainees                 | 16                     | 1 hour                                                       | Ex vivo, images and videos | iScan OE and NBI | 78                             |

**Table 5-8 Characteristics of studies included in the systematic review and meta-analysis**

\*P=Prospective, O=Observational, RCT=Randomised-controlled trial

#### 5.4.3.2 Overall training efficacy and overall efficacy of training methods:

When pooling together all studies, polyp sizes (16 studies, 179 participants) (Rogart, Jain et al. 2008, Higashi, Uraoka et al. 2010, Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Coe, Thomas et al. 2012, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Sikong, Lin et al. 2016, Basford, Longcroft-Wheaton et al. 2017, Aihara, Kumar et al. 2018, Iacucci, Trovato et al. 2018, Bae, Lee et al. 2019, Basford, Brown et al. 2019, Subramaniam, Hayee et al. 2019), assessing the efficacy of any method of optical diagnosis training, pooled pre-training accuracy was 70.3% (6416/9131 correct diagnoses) vs. post-training accuracy was 81.6% (7416/9213 correct diagnoses) (RR 1.17 95% CI 1.09-1.24  $p < 0.001$ ) (Figure 5-5). The 95% prediction interval was 0.85-1.61. There was considerable heterogeneity in these studies ( $I^2 = 94\%$ ) without significant publication bias (Figure 5-6).

When pooling only studies describing computer-based training (8 studies, 94 participants) (Ignjatovic, Thomas-Gibson et al. 2011, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Aihara, Kumar et al. 2018, Bae, Lee et al. 2019, Basford, Brown et al. 2019, Smith, Saltzman et al. 2019), pooled pre-training accuracy was 69.2% (3125/4514 correct diagnoses) vs. post-training accuracy of 80.0% (3611/4514 correct diagnoses) (RR 1.16 95% CI 1.09-1.23  $p < 0.001$ ) (Figure 5-7). The 95% prediction interval was 0.92-1.46. Substantial heterogeneity was detected in this sub-group ( $I^2 = 84\%$ ) without significant publication bias.

When pooling only studies describing didactic training (nine studies, 85 participants) (Rogart, Jain et al. 2008, Higashi, Uraoka et al. 2010, Raghavendra, Hewett et al. 2010, Coe, Thomas et al. 2012, Sikong, Lin et al. 2016, Iacucci, Trovato et al. 2018, Smith, Saltzman et al. 2019, Subramaniam, Hayee et al. 2019), pooled pre-training accuracy was 71.3% (3291/4617 correct diagnoses) vs. post-training accuracy of 83.1% (3905/7516 correct diagnoses) (RR

1.15 95% CI 1.03-1.29  $p < 0.001$ ) (Figure 5-8). The 95% prediction interval was 0.77-1.71.

Considerable heterogeneity was detected in this sub-group ( $I^2$  96%) without significant publication bias (Figure 5-9).

On meta-regression analysis, there was no significant difference in post-training accuracy between didactic and computer-based delivery methods ( $p = 0.798$ ).

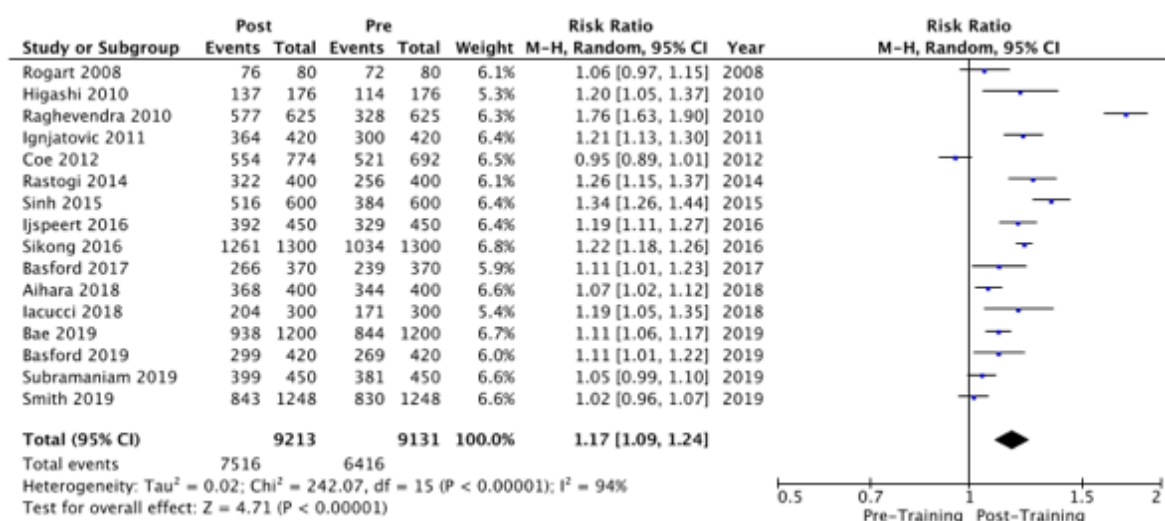


**Figure 5-5 Forest plots for all studies assessing the effect of training on accuracy of optical diagnosis of colorectal polyps**

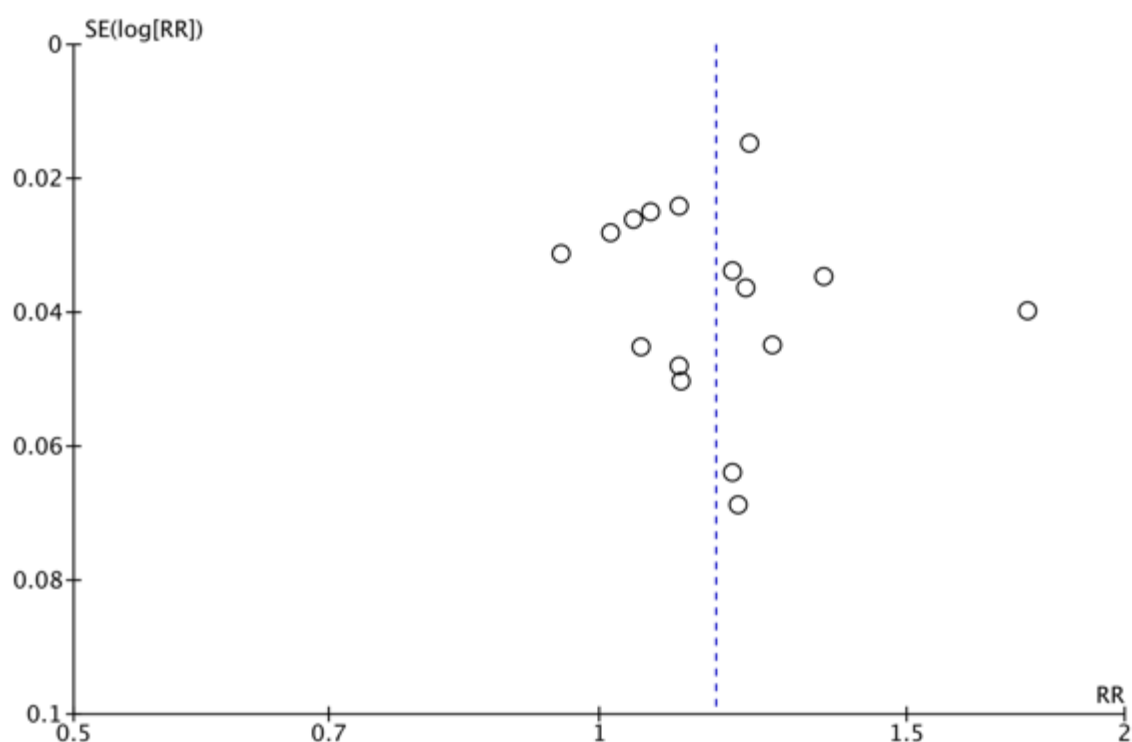


**Figure 5-6 Funnel plots for assessing the impact of training on optical diagnosis accuracy of small/diminutive colorectal polyps in all studies**

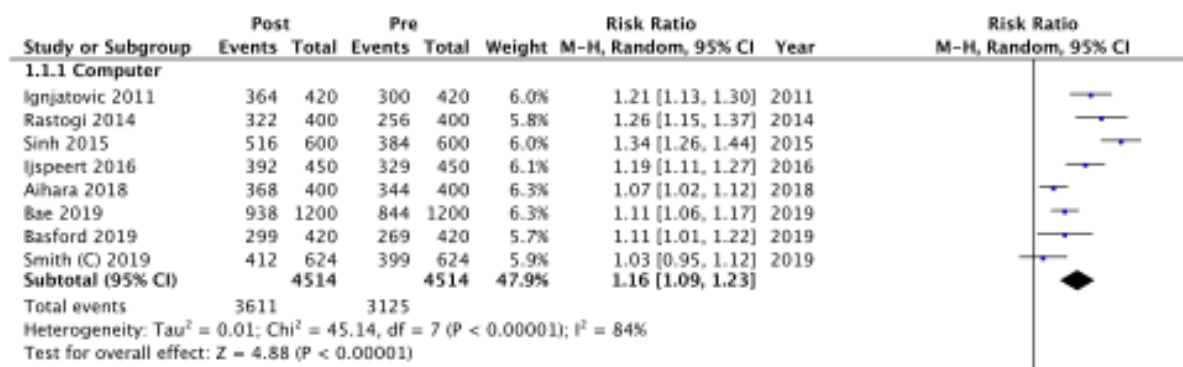


**Figure 5-7 Forest plots for studies assessing the effect of computer-based training on accuracy of optical diagnosis of colorectal polyps**

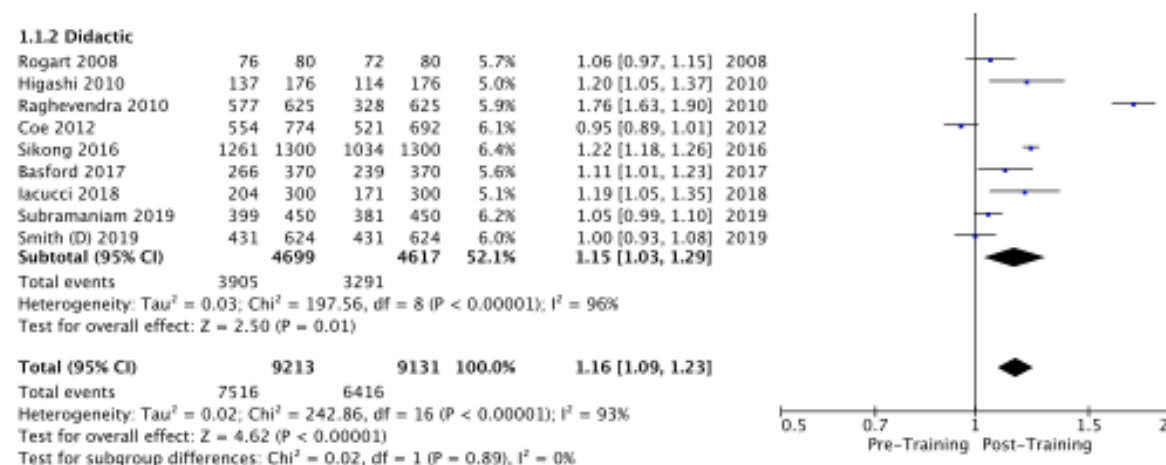


**Figure 5-8 Forest plots for studies assessing the effect didactic training on accuracy of optical diagnosis of colorectal polyps**

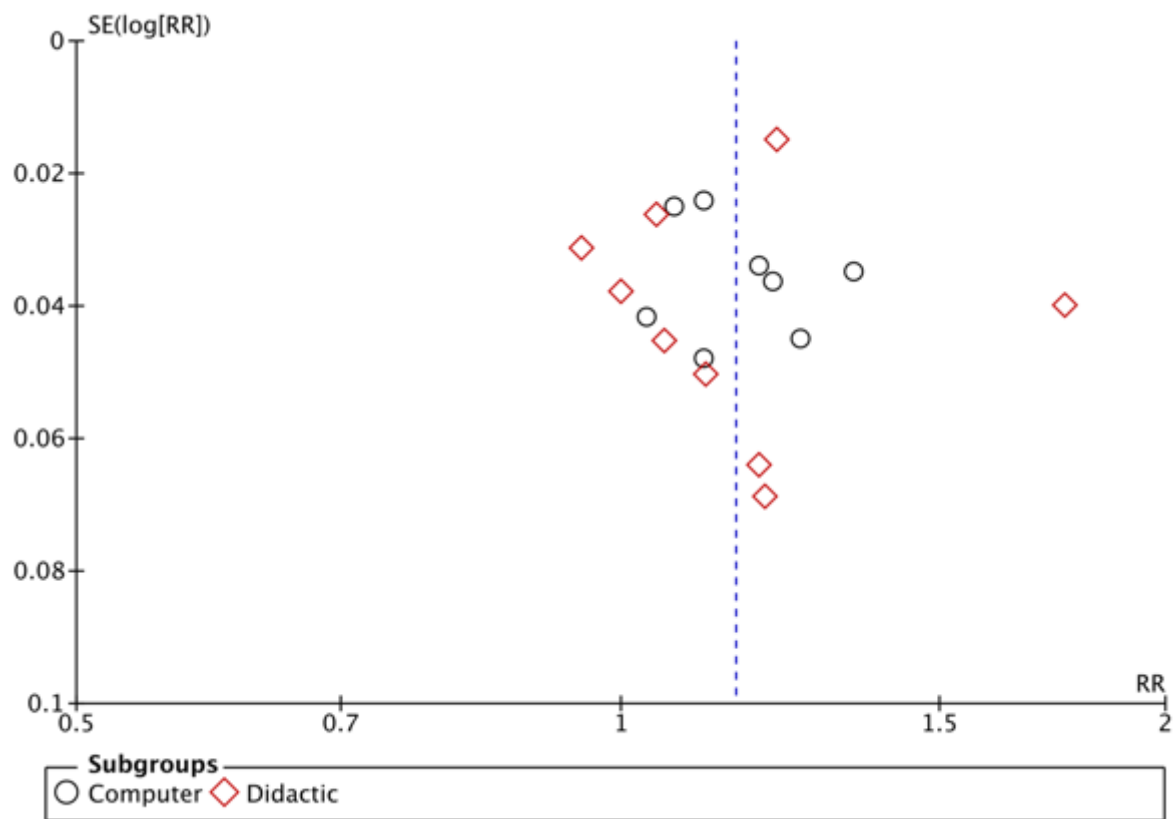


**Figure 5-9 Funnel plots for assessing the impact of training on optical diagnosis accuracy of small/diminutive colorectal polyps in computer-based and didactic training in all endoscopists**

#### 5.4.3.3 Experienced endoscopists:

Studies describing the efficacy of training in experienced endoscopists were then selected.

Method of training for experienced endoscopists comprised of didactic training (seven studies, 49 participants) (Rogart, Jain et al. 2008, Higashi, Uraoka et al. 2010, Raghavendra, Hewett et al. 2010, Coe, Thomas et al. 2012, Sikong, Lin et al. 2016, Basford, Longcroft-Wheaton et al. 2017, Subramaniam, Hayee et al. 2019) and computer-based training (six studies, 64 participants) (Ignjatovic, Thomas-Gibson et al. 2011, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Bae, Lee et al. 2019, Basford, Brown et al. 2019).

After pooling these studies, optical diagnosis training method on optical diagnosis in experienced endoscopists improved accuracy from 69.8% (3771/5403 correct diagnoses) to 82.4% (4521/5485 correct diagnosis) (RR 1.20 95% CI 1.11-1.29  $p < 0.001$ ). The heterogeneity was considerable ( $I^2$  92%).

On subgroup analysis by type of training, didactic training improved optical diagnosis accuracy from 72.2% (1685/2333 correct diagnoses) to 83.9% (2027/2415 correct diagnoses) RR 1.19 (95% CI 1.03-1.36  $p < 0.001$ ), and computer-based training from 67.9% (2086/3070 correct diagnoses) to 81.2% (2494/3070 correct diagnoses) RR 1.21 (95% CI 1.13-1.30  $p < 0.001$ ). The improvement was not significantly different between the two training methods ( $p = 0.792$ ) (Figure 5-10).

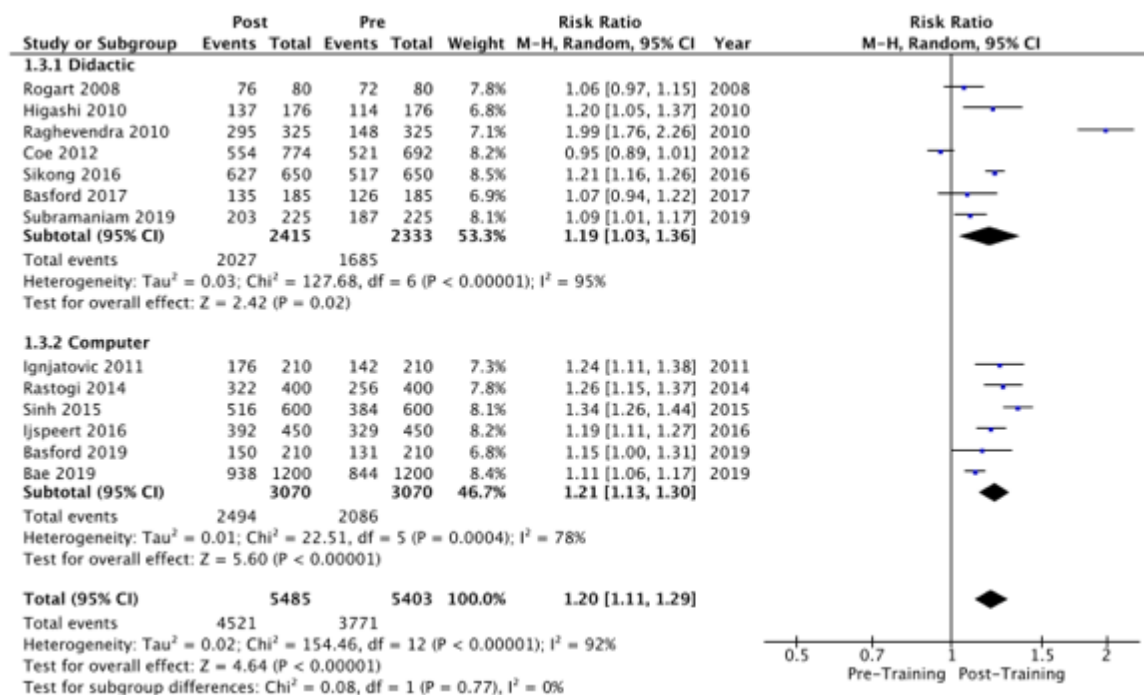


**Figure 5-10 Forest plots for studies assessing the impact of training on the accuracy optical diagnosis of colorectal polyps in experienced endoscopists**

#### 5.4.3.4 Trainee endoscopists:

Studies describing the efficacy of training in trainee endoscopists were then selected. A total of 8 studies (Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Basford, Longcroft-Wheaton et al. 2017, Aihara, Kumar et al. 2018, Iacucci, Trovato et al. 2018, Basford, Brown et al. 2019, Smith, Saltzman et al. 2019, Subramaniam, Hayee et al. 2019) evaluated impact of training among in 58 trainee endoscopists. In detail, 4 studies (Raghavendra, Hewett et al. 2010, Basford, Longcroft-Wheaton et al. 2017, Iacucci, Trovato et al. 2018, Subramaniam, Hayee et al. 2019) with 28 participants used didactic training, and 3 studies (Ignjatovic, Thomas-Gibson et al. 2011, Aihara, Kumar et al. 2018, Basford, Brown et al. 2019) with 22 participants evaluated computer based training. Only one study compared didactic vs computer-based training (Smith, Saltzman et al. 2019).

After aggregating these studies involving trainee participants, optical diagnosis training improved accuracy from 69.6% (2645/3803 correct diagnoses) to 78.8% (2995/3803 correct diagnoses) (RR 1.14; 95% CI 1.06-1.24,  $p < 0.001$ ). There was considerable heterogeneity ( $I^2$  89%). On subgroup analysis didactic training improved accuracy from 68.1% (1606/2359 correct diagnoses) to 79.6% (1878/2359 correct diagnoses) (RR 1.18; 95% CI 1.04-1.34,  $p < 0.001$ ) and computer-based training from 72.0% (1039/1444 correct diagnoses) to 77.4% (1117/1444 correct diagnoses) (RR 1.09; 95% CI 1.03-1.15,  $p < 0.001$ ) (Figure 5-11). The improvement in accuracy did not differ significantly between the two training methods ( $p = 0.312$ ).

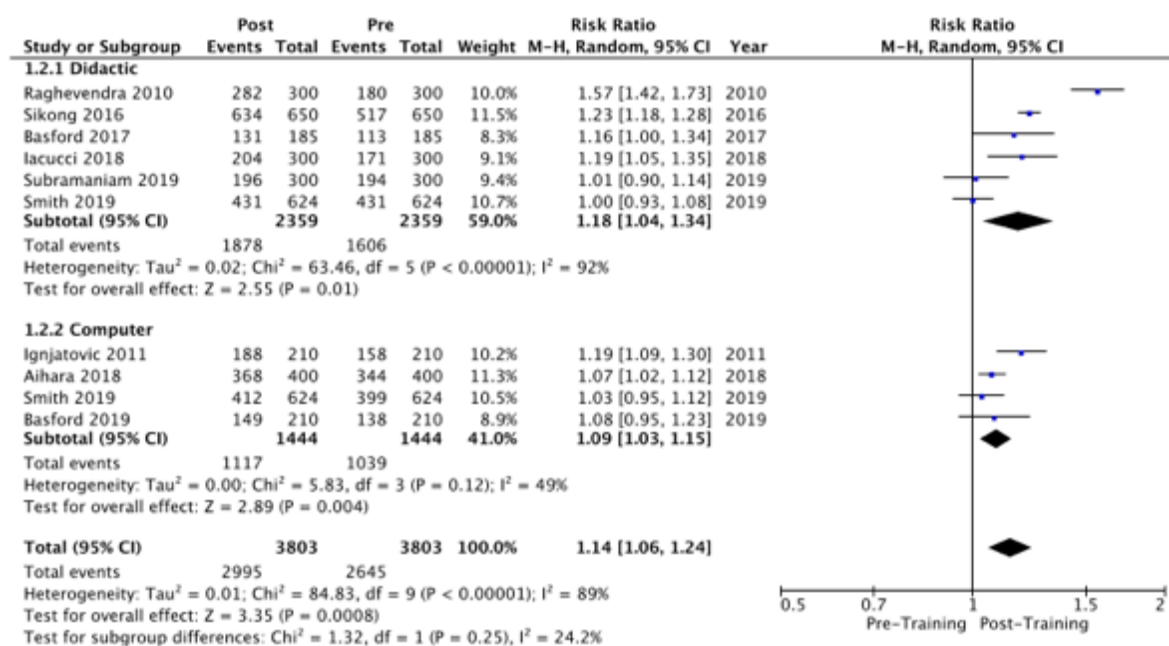


**Figure 5-11 Forest plots for studies assessing the impact of training on the accuracy optical diagnosis of colorectal polyps in trainees**

#### 5.4.3.5 *Small/diminutive colorectal polyps:*

Studies that only included small/diminutive colorectal polyps in the assessment of training in optical diagnosis were selected. A total of nine studies assessed the impact of optical diagnosis training on 104 endoscopists (Higashi, Uraoka et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, Iacucci, Trovato et al. 2018, Bae, Lee et al. 2019, Basford, Brown et al. 2019, Smith, Saltzman et al. 2019, Subramaniam, Hayee et al. 2019). The pooled pre-training accuracy was 68.1% (3549/5214) and post-training accuracy was 77.1% (4022/5214) (RR 1.16 95% CI 1.08-1.24 p <0.001) (Figure 5-12). Substantial heterogeneity in these studies was detected ( $I^2$  87%).

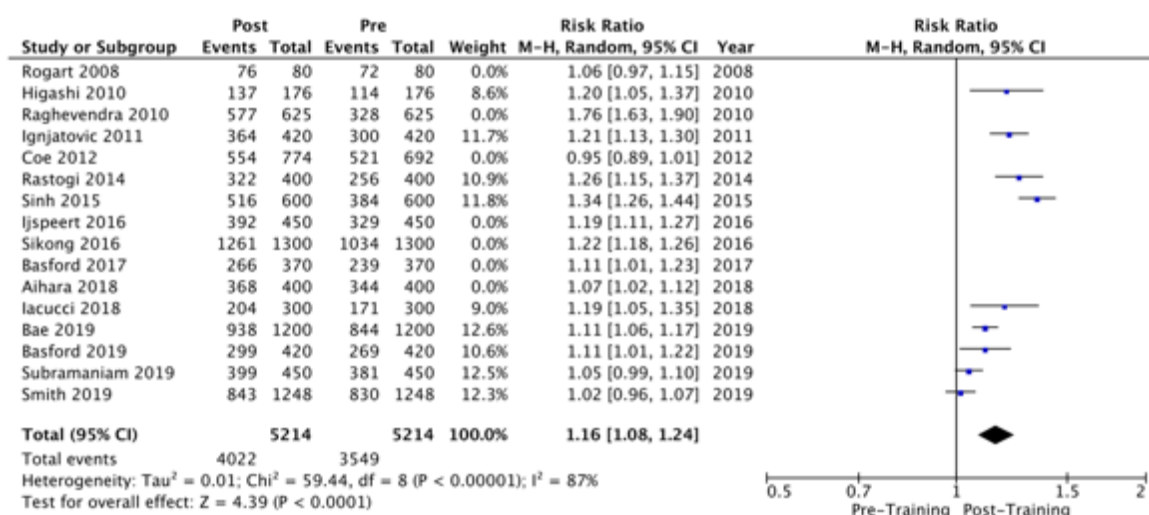


**Figure 5-12 Forest plots for all studies assessing the effect of training on accuracy of optical diagnosis of small/diminutive colorectal polyps only**

#### *5.4.3.6 Study quality and risk of bias:*

The summary of the Cochrane Collaboration's risk of bias tool is presented in Figure 5-13.

Only two randomised trials were included (Coe, Thomas et al. 2012, Smith, Saltzman et al. 2019), both of which report how the randomisation process took place, but both were unable to conceal the allocation to participants. Participants in all studies included were not blinded to the intervention. Participants were blinded to the outcomes of training during the study process for most studies and all studies produced complete datasets without selective reporting.

Using the GRADE approach, the quality of evidence was downgraded by two points to low quality due to risk of bias and inconsistency from the heterogeneity in the studies.

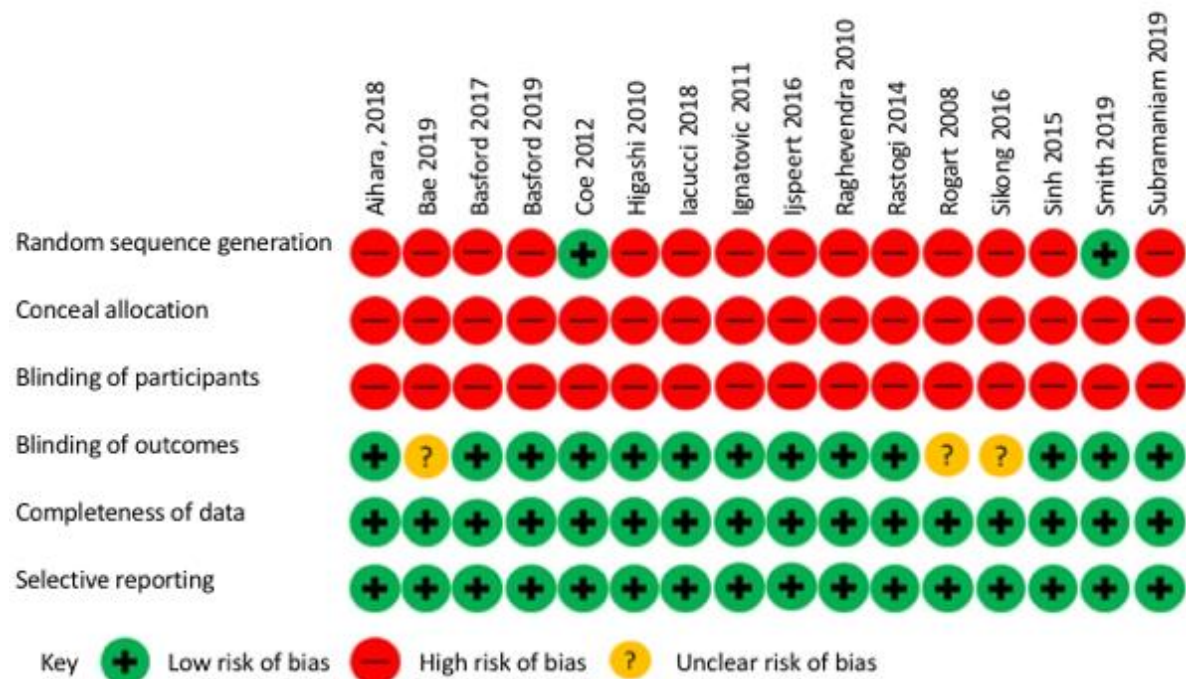


**Figure 5-13 Risk of bias of studies included using Cochrane Collaboration’s risk of bias tool**

#### **5.4.4 Discussion:**

According to this systematic review and meta-analysis, training in the optical diagnosis of colorectal polyps was associated with improved diagnostic accuracy in both experienced and trainee endoscopists. Furthermore, when only considering small/diminutive colorectal polyps there was also a statistically significant improvement in diagnostic accuracy following training. These results are clinically relevant and important for the following considerations. Firstly, the efficacy of training has been proved for both experienced and trainees in endoscopy. Indeed, trainees showed a similar post-training accuracy level as compared to experienced endoscopists, confirming the positive impact of training on optical diagnosis. Secondly, irrespective of the pre-training accuracy level, training programs resulted in uniformly elevated accuracy levels. Thirdly, there was no statistically significant difference seen in the improvement in accuracy between computer-based and didactic training.

Endoscopy practice is going through major changes which has accelerated in response to the Covid-19 pandemic. To minimise the risk of exposure to Covid-19, units are taking steps to lessen the footfall in procedural rooms, (Gralnek, Hassan et al. 2020). This will inevitably have a serious consequence on endoscopy training, particularly given that restrictions may be in place until 2022 (Gralnek, Hassan et al. 2020, Kissler, Tedijanto et al. 2020). It is recognised that during these unprecedented times that trainees seek training in alternative means, perhaps focusing on cognitive skills (Gralnek, Hassan et al. 2020). This increased importance in alternative educational resources such as computer-based and simulation based training in optical diagnosis would address the unmet need during this period and likely in the longer term too (Siau, Iacucci et al. 2020).

This study confirms computer-based training is effective in improving optical diagnosis accuracy and furthermore, there is no statistical difference between computer-based and didactic training which is further confirmed by the randomised trial presented earlier (Smith, Saltzman et al. 2019). The  $I^2$  values indicate considerable heterogeneity in most analyses, therefore there needs to be a focus on validation, standardisation and quality assurance of training and long-term studies to assess the retention of optical diagnosis skills after a training intervention.

Traditionally the method of training endoscopists was an apprenticeship model with competence determined at a local level by trainers, however there is now more focus on mechanical skills (Siau, Hodson et al. 2019) using key performance thresholds such as caecal intubation and polyp detection rate. Inexperienced trainees may not be receptive to learning about pathology *in vivo* whilst focusing on achieving key performance statistics. In the UK the mean number of colonoscopies completed upon conclusion of training is 529 with less than 100 polypectomies completed (Malik, Siau et al. 2019). Optical diagnosis accuracies improve with experience with a learning curve of approximately 89-110 polyps (Neumann, Vieth et al. 2013). Therefore, at completion of training it is unlikely that trainees will have the expertise to accurately perform optical diagnosis.

Didactic training remains the most resource intensive and expensive form of training and does not allow participants to learn at an individual pace with one-off training sessions unlikely to sustain performance over time. Computer-based training is attractive as it allows individuals to train at their own pace at a time that suits, is resource light, inexpensive and could augment a “just for me” personalised training for continuity of learning. Trainees develop motor skills

at different speeds (Bretthauer and Brandrud 2014, Ward, Mohammed et al. 2014) and there is no reason to suggest the same does not apply to cognitive skills, which includes optical diagnosis. However, it is likely that this will not be effective in isolation and further studies need to be conducted in finding the optimum method of self-learning applicable to all endoscopic platforms and levels of experience. It also needs to be combined with regular audits of performance in accordance with guidelines (Bisschops, East et al. 2019). Continuous feedback on optical diagnosis predictions is also likely to be effective at improving performance (Bae, Lee et al. 2019), particularly when coupled with the endoscopic images thereby allowing endoscopists to learn from mistakes and as well as good practice, however this will require major infrastructural changes and investment. The most effective training is likely a combination of methods, whereby expert-led didactic training is followed by a standardised validated computer-based training and monitoring of in vivo performance together with regular auditing.

Most studies investigating the effectiveness of training methods on the optical diagnosis of colorectal polyps were small observational studies with only two randomised trials. The studies included a baseline level of performance prior to training, which enabled the effectiveness of training to be clearer. Including trainee and experienced endoscopists, the target of intervention was relevant to the clinical practice of endoscopy. The substantial heterogeneity of studies may be explained by the following differences:

1. Training intervention (number of sessions and duration). The duration of training provided was as little as 15-20 minutes (Ignjatovic, Thomas-Gibson et al. 2011, Rastogi, Rao et al. 2014, JE, Bastiaansen et al. 2016, Basford, Brown et al. 2019) through to more intensive monthly hour long training sessions (Sikong, Lin et al. 2016).

2. Assessment methods (type of media used, endoscopic platform and whether material was repeated pre- and post-training). Still images (Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Basford, Longcroft-Wheaton et al. 2017, Aihara, Kumar et al. 2018, Subramaniam, Hayee et al. 2019) and videos of polyps have been used with some studies using a combination, however the difference in effectiveness of training may be attributable to type and quality of media included. The assessment process pre- and post-training varies between studies. Studies have included the same set of media (image or video) in the pre- and post-training assessment phase, which ensures continuity of image quality but does raise the possibility recall bias (Ignjatovic, Thomas-Gibson et al. 2011, Basford, Longcroft-Wheaton et al. 2017, Aihara, Kumar et al. 2018, Iacucci, Trovato et al. 2018). There was not an overall consistency in the media used in training nor a consistent assessment process across studies. Effective training also needs to consider multiple endoscopic platforms, particularly when aimed at trainees who may be exposed to different endoscopic platforms through training, thereby studies including one platform are limited (Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Rastogi, Rao et al. 2014, Sinh, Gupta et al. 2015, JE, Bastiaansen et al. 2016, Sikong, Lin et al. 2016, Aihara, Kumar et al. 2018, Bae, Lee et al. 2019, Subramaniam, Hayee et al. 2019), whilst studies using different endoscopic platforms adds to the heterogeneity of the studies.
3. Type of lesions included. For optical diagnosis to be effective, all lesions (adenoma, hyperplastic and sessile serrated lesions) should be included, however many studies only differentiate between hyperplastic and adenoma (Raghavendra, Hewett et al. 2010, Ignjatovic, Thomas-Gibson et al. 2011, Subramaniam, Hayee et al. 2019).

4. Washout period between pre-training assessment and training, and participant characteristics (pre-training accuracy levels were as low as 45% in one study amongst experienced endoscopists (Raghavendra, Hewett et al. 2010) and as high as 88% in some trainees (Subramaniam, Hayee et al. 2019)).
5. Baseline level of performance. Some pre-training accuracy levels were as low as 45% in one study amongst experienced endoscopists (Raghavendra, Hewett et al. 2010) and as high as 88% in some trainees (Subramaniam, Hayee et al. 2019). It is expected that the motivation of participants will vary, therefore studies including highly experienced bowel cancer screening endoscopists (JE, Bastiaansen et al. 2016), who are likely to be highly motivated, may not reflect the true impact of training in everyday endoscopists.
6. Variability in trainers and training materials. The effectiveness of didactic training will vary between trainers, and therefore we cannot assume the intervention is consistent across studies. The same applies for computer-based learning with variable quality in educational material, and some participants may be less likely to respond to this type of training method.

This systematic review and meta-analysis have limitations. Firstly, all included studies were unblinded due to the nature of the intervention (i.e., training methods). This however is a common bias of many evaluations of technological improvements in gastrointestinal endoscopy. Secondly, participants were aware that their performance in assessments were being monitored and analysed, which may introduce further bias as there may have been a subconscious increased effort during the study period. One possible solution to this would be tracking *in vivo* optical diagnosis accuracy for a prolonged period of time before and after training. However,

this would be resource intensive and may still incur bias. Third, not all studies provided data to calculate sensitivity, specificity or negative predictive values (against the recommended ASGE PIVI threshold of >90%), nor was it possible to compare receiver operating characteristics curves pre- and post-training. Given the large degree of heterogeneity, the inconsistency of results, the lack of randomized controlled trials and inherent bias the quality of evidence of the studies is considered to be low.

This systematic review and meta-analysis has uncovered important gaps in evidence, which will need to be addressed in future studies. Further study on optical diagnosis training is required to establish the optimum method. There is a lack of robust randomised trials comparing training modes with only one comparing didactic with computer-based training, which was presented earlier in this chapter (Smith, Saltzman et al. 2019). Another important consideration is the effectiveness of training across endoscopic platforms, while initial attempts at this suggest that training is effective across NBI and iScan (Iacucci, Trovato et al. 2018, Smith, Saltzman et al. 2019), however there is yet to be a study that includes all of NBI, iScan and BLI. The issue of retention of optical diagnosis over time is one that has yet to be thoroughly explored. The majority of studies included in this systematic review and meta-analysis involve participants scoring media on the same day as training, however none have demonstrated that optical diagnosis skills can be retained over a prolonged period of time. This is essential for optical diagnosis to be incorporated in everyday practice. Future studies should re-assess participants after a prolonged period of time such as 3-6 months or even longer and should include an *in vivo* assessment component in order to formally investigate whether *ex vivo* training can translate to *in vivo* performance over time.

A future training strategy may include a combination of methods with perhaps didactic training from experts to set a foundation of knowledge, further supported by computer-based training or artificial intelligence (AI) systems. Computer based training could be periodically reinforced relatively easily. The ESGE suggest the most likely scenario of artificial intelligence, being introduced in endoscopy, will be as a “second reader” as opposed to a “stand alone” system (East, Vleugels et al. 2016). Therefore, for AI to be utilised in everyday practice there is a need for endoscopists to up skill in optical diagnosis to be able to safely interpret and act upon the readings from AI systems.

In conclusion, training in optical diagnosis improves the accuracy of histology predictions of colorectal polyps, including small/diminutive polyps. The optimal method of training may include a combination of training methods augmented with continuous *in vivo* training, which may be provided by trainers or AI systems. Future studies need to focus on standardizing and validating training modules to enhance cognitive skills of endoscopists.

The section above is taken from the following publication:

**Smith SCL**, Siau K, Cannatelli R, Antonelli G, Shivaji UN, Ghosh S, Saltzman JR, Hassan C, Iacucci M. Training methods in optical diagnosis and characterisation of colorectal polyps: a systematic review and meta-analysis. *Endosc Int Open* 2021; 9: E716-E726

## 5.5 Training in advanced endoscopy in IBD

### 5.5.1 Background

Patients with Inflammatory Bowel Disease (IBD) are at an increased risk of colorectal cancer (CRC) (Ekbom, Helmick et al. 1990, Ekbom, Helmick et al. 1990, Bergeron, Vienne et al. 2010, Dyson and Rutter 2012, Lutgens, van Oijen et al. 2013). The risk of CRC increases with duration of disease with a 1% risk after 10 years increasing to 14% after 30 years (Rutter, Saunders et al. 2006, Herrinton, Liu et al. 2012). Regular and high-quality surveillance colonoscopy every 1-5 years, depending on risk profile and region of practice, is recommended for early detection and treatment of colonic lesions (Cairns, Scholefield et al. 2010, Magro, Gionchetti et al. 2017). The aim is to detect lesions early and at a stage whereby endoscopic removal can occur before advanced histology or cancer is present.

In the bowel cancer screening population (non-IBD) there are clear quality indicators that need to be met to perform screening colonoscopy; for example Adenoma Detection Rate (ADR), withdrawal time and caecal intubation rate (CIR) (Lee, Rutter et al. 2012). SCENIC (Surveillance for Colorectal Endoscopic Neoplasia Detection and Management in Inflammatory Bowel Disease Patients: International Consensus Recommendations) and ECCO (European Crohn's Colitis Organization) have guidance on how to perform IBD surveillance, for example recommending dye-based chromoendoscopy (DCE) but do not set out clear key quality indicators (Laine, Kaltenbach et al. 2015, Magro, Gionchetti et al. 2017). This is likely to cause variation in practice between individual endoscopists and between endoscopy units-due to a lack of expertise, adequate training and the need for different equipment (Magro, Gionchetti et al. 2017). The American Society of Gastrointestinal Endoscopy (ASGE) organized an educational event to optimise quality in endoscopy in IBD

involving specialists in IBD and endoscopy. After this meeting over half of the participants (51.7%) adopted DCE in most surveillance colonoscopies, reinforcing the need for appropriate education (Kaltenbach, Soetikno et al. 2017).

Unlike bowel cancer screening programmes, clear guidance on quality indicators is lacking in IBD surveillance and is currently an unmet need. Key quality performance indicators should be introduced and must be auditable, measurable and reflect good practice. They may also have a major impact on lesion detection but also characterization and management. New strategies should be developed introducing new roadmaps for future training and implementation of surveillance colonoscopy in IBD. Some of the elements from the recommendations mentioned above may be used as a preliminary guide to quality performance (Laine, Kaltenbach et al. 2015, Magro, Gionchetti et al. 2017).

### **5.5.2 Aims**

The aim of this study was to identify performance in key quality indicators in IBD surveillance, before and after service changes, which included training in optical diagnosis of IBD lesions, were made in IBD surveillance colonoscopy.

### **5.5.3 Methods**

#### *5.5.3.1 Study design*

A retrospective review of all IBD surveillance colonoscopies performed January 2014-October 2018 (58 months) in a tertiary referral unit (University Hospital Birmingham NHS

Trust, UK) was completed. The audit was registered at the University Hospital Birmingham, UK (CARMS-14727).

Using hospital coding, a list of patients having an IBD surveillance procedure was obtained from electronic medical records and clinical and endoscopic notes were reviewed.

Colonoscopic assessment of disease activity and patients undergoing flexible sigmoidoscopy were excluded. In the absence of specific IBD surveillance recommendations, the quality measure domains developed by the European Society of Gastrointestinal Endoscopy (ESGE) including completeness of colonoscopy, lesion detection and management (Rembacken, Hassan et al. 2012) were included. All colonoscopies included in the study were completed using Olympus (Tokyo, Japan) 260/290 series high-definition endoscopes.

#### *5.5.3.2 Intervention*

To improve quality of IBD surveillance changes were implemented to the endoscopy service in August 2016. These included (a) Education in a structured training module (this will be discussed in the next paragraph); (b) DCE (as panchromoendoscopy) as a standard of practice; (c) introduction of foot pedal operated pump-jets for cleaning and application of dye spray; (d) allocation of 45-minute procedure slot; (e) targeted biopsies (except high risk patients); (f) formal scoring of endoscopic disease activity; (g) detection of lesions (neoplastic lesions: dysplasia, cancer, sessile serrated adenoma/polyp (SSA/P), adenomas and non-neoplastic lesions: pseudopolyps and hyperplastic polyps); (h) characterisation of lesion morphology. In addition to these changes all IBD surveillance colonoscopies were allocated to a small team of DCE trained endoscopists (n=4). Data collected included patient demographics (type of IBD, co-existence of Primary Sclerosing Cholangitis (PSC), duration and distribution of disease), procedural details (endoscopist, caecal intubation, sedation used,

bowel preparation quality), type of surveillance used (DCE, standard white light, targeted or random biopsy surveillance), documented reasons DCE not used, endoscopic activity of disease score-such as MES (Schroeder, Tremaine et al. 1987), UCEIS (Travis, Schnell et al. 2012) and SES-CD (Daperno, D'Haens et al. 2004), and lesion documentation (size, Paris classification (Endoscopic Classification Review 2005), Kudo pit pattern (Kudo, Hirota et al. 1994), borders, presence of ulceration and dysplasia detection), which is based on the SCENIC recommendations (Laine, Kaltenbach et al. 2015). Withdrawal time data for procedures were not collected as these were not routinely recorded as a standard during the time of the study.

#### *5.5.3.3 Training and education module*

The training module delivered was an interactive Microsoft PowerPoint (Microsoft Inc., Redmond, Washington, USA). The content of the module included SCENIC recommendations, the Paris modified Kudo pit pattern (Laine, Kaltenbach et al. 2015), Hazewinkel criteria for SSA/P (Hazewinkel, Lopez-Ceron et al. 2013), and the Frankfurt Advanced Chromoendoscopic IBD Lesions (FACILE) classification (Iacucci, McQuaid et al. 2018). Representative interactive still images (n=50) and videos (n=10) of IBD-lesions including adenomas, SSA/Ps, cancer and inflammatory polyps were displayed. The training module described the process of performing a high quality IBD surveillance procedure, which is summarised by the roadmap in Figure 5-14. Training was delivered to the DCE-trained endoscopists.

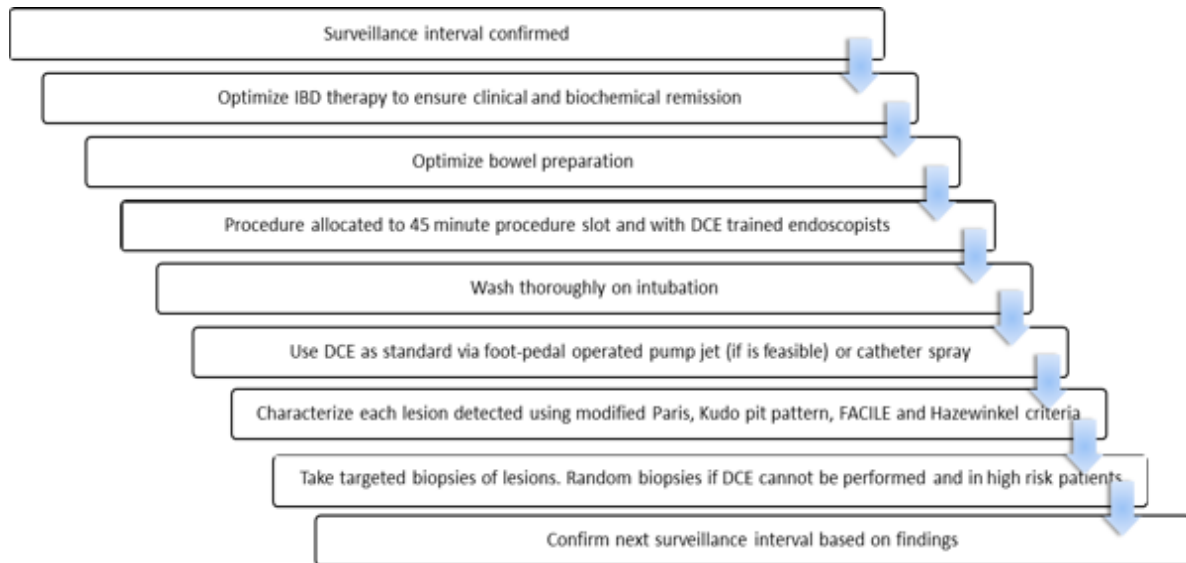


**Figure 5-14 Recommended roadmap for high quality IBD surveillance**

#### 5.5.3.4 Outcomes

The primary outcome was the change in proportion of IBD surveillance colonoscopies using DCE (panchromoendoscopy). Secondary outcomes included the change in lesion/dysplasia detection rate, random biopsy surveillance, formal endoscopic scoring of inflammatory activity and formal descriptors of lesion morphology documented.

#### 5.5.3.5 Statistical analysis

Statistical analysis was performed using STATA Version 15.1 (StataCorp, Texas, USA). A p value of <0.05 was considered statistically significant. Quantitative variables were expressed as medians and ranges and/or means  $\pm$  SDs. Categorical variables were expressed as total number and frequencies. Performance in quality measures before and after changes to local IBD surveillance practice (August 2016) were compared using chi-square analysis for categorical data and Mann-Whitney U test for continuous data.

## **5.5.4 Results**

### *5.5.4.1 Demographic characteristics*

A total of 780 procedures were coded as IBD surveillance between January 2014 and October 2018, with 182 procedures excluded as procedures were a flexible sigmoidoscopy or an assessment of IBD disease activity as the primary indication. There was no statistical difference in patient demographics between the pre-August 2016 and post-August 2016 groups (Table 5-9). There was a high proportion of surveillance colonoscopies performed in PSC patients (39.4% pre-August 2016 vs. 41.7% post-August 2016  $p=0.552$ ) which reflects the tertiary referral nature of the unit. There was no significant difference between the two groups in terms of poor bowel preparation using the Boston Bowel Preparation Score ( $<2$ ) (Calderwood and Jacobson 2010) (14.4% vs. 17.4%  $p=0.32$ ). There was no difference in type of sedation used (conscious sedation 66.8% vs. 67.0%  $p=0.96$ ) with the remainder using Entonox only. It was standard practice that colonoscopies were completed using CO<sub>2</sub> insufflation therefore in this study all procedures were carried out in this manner.

**Table 5-9 Patient demographics of the IBD surveillance study**

|                                              | Pre-August 2016<br>(n=277) | Post-August 2016<br>(n=321) | <i>P value</i> |
|----------------------------------------------|----------------------------|-----------------------------|----------------|
| Gender (n procedures, %)                     |                            |                             |                |
| Male                                         | 158 (57.0%)                | 180 (56.1%)                 | 0.81           |
| Female                                       | 119 (43.0%)                | 131 (43.9%)                 |                |
| Age (mean ± SD) years                        | 48.8 ±16.6                 | 46.9 ± 16.5                 | 0.16           |
| IBD diagnosis (n procedures, %)              |                            |                             |                |
| Ulcerative colitis                           | 262 (94.6%)                | 303 (94.4%)                 | 0.31           |
| Crohn’s colitis                              | 10 (3.6%)                  | 8 (2.5%)                    |                |
| Indeterminate colitis                        | 4 (1.4%)                   | 10 (3.1%)                   |                |
| Unknown                                      | 1 (0.4%)                   | 0 (0.0%)                    |                |
| Diagnosis of PSC (n procedures, %)           | 109 (39.4%)                | 134 (41.7%)                 | 0.55           |
| Duration of the disease (mean ± SD)<br>years | 16.0 ± 10.3                | 16.1 ± 9.9                  | 0.77           |
| Distribution of disease (n procedures, %)    |                            |                             |                |
| Pan-colitis                                  | 201 (72.6%)                | 229 (71.3%)                 | 0.42           |
| Sub-total colitis                            | 11 (4.0%)                  | 16 (5.0%)                   |                |
| Left-sided colitis                           | 58 (20.9%)                 | 73 (22.7%)                  |                |
| Unknown                                      | 7 (2.5%)                   | 3 (0.9%)                    |                |
| Sedation used                                |                            |                             |                |
| Conscious sedation                           | 185 (66.8%)                | 215 (67.0%)                 | 0.96           |
| Boston Bowel Preparation Scale <2            | 40 (14.4%)                 | 56 (17.4%)                  | 0.32           |

#### 5.5.4.2 *Comparison of key quality indicators before and after changing practice*

Comparisons of performance in key quality indicators in IBD surveillance between the two groups (pre- vs. post-Aug 2016) are highlighted in Table 5-10. There are significant differences between the groups (pre-August 2016 vs. all endoscopists post-August 2016/4 and DCE trained endoscopists post-August 2016) with a higher proportion of procedures completed to the terminal ileum (TI) (20.9% vs. 45.8/49.4%,  $p<0.0005$ ). The primary outcome of the proportion of procedures using DCE increased (54.2% vs. 76.0/78.7%  $p<0.0005$ ). One of the changes implemented included the use of DCE as a standard with Table 5-11 highlighting the reasons for not using DCE. Insufficient bowel preparation and persistent disease activity were the most common reasons for DCE not being carried out. There was a decrease in procedures not performed with DCE due to bowel preparation from 90.6% to 45.9% ( $p=0.018$ ) between pre- and post-August 2016, however when combining with procedures not performed due to both disease activity and bowel prep, there was no significant difference in bowel preparation (90.6% vs. 64.9%  $p=0.199$ ). There was a significant increase in the proportion of procedures performed without DCE due to disease activity between pre-August 2016 and post-August 2016 (6.3% vs. 27.0%  $p=0.037$ ). The number of colonoscopies that had no formal surveillance technique (DCE/random biopsy collection) nor documented reasons why DCE could not be used was 61 (22.0%) pre-August 2016 and 30 (9.3%) post-August 2016 ( $p<0.005$ ).

Procedures using random biopsy surveillance reduced (12.3% vs. 3.1/2.3%  $p<0.0005$ ). Formal description of disease activity (2.5% vs. 46.4/49.8%  $p<0.0005$ ), size of lesion documentation (65.2% vs. 79.0/80.5%  $p<0.05$ ), use of the Paris classification (26.1% vs. 57.0/62.1%  $p<0.0005$ ), use of Kudo Pit Pattern (21.7% vs. 59.0/66.7%  $p<0.0005$ ) and descriptions of lesion borders (0.0% vs. 18.0/20.7%  $p<0.0005$ ) all improved following the

implementation of the changes previously described. Compared with all endoscopists pre-August 2016, the four DCE trained endoscopists had an improved lesion detection rate after training 24.9% vs. 33.1%  $p<0.05$  (Table 5-10). Although there was an increase in the lesion detection rate, there was not a statistically significant increase in dysplasia detection rate between pre-August 2016 and post-August 2016 (Table 5-10). Of the 20 dysplastic lesions detected in the pre-August 2016 group, 15 were located within a colitic area (75%) whereas 21 of the 32 dysplastic lesions detected in the post-August 2016 group were located within a colitic area (75% vs. 66%  $p=0.563$ ). The breakdown of lesions detected is shown in Table 5-12.

Overall, 3 patients were diagnosed with adenocarcinoma in the present study. All were diagnosed after August 2016 and 2 had a previous colonoscopy  $<3$  years prior to the cancer diagnosis (one-no data on previous colonoscopy). This gives an interval colorectal cancer rate of 0.7% in the pre-August 2016 cohort, however as we included colonoscopies performed until October 2018, we are unable to report the interval cancer rate in the post-August 2016 cohort.

#### *5.5.4.3 Quality measures performance of non-experienced vs. DCE trained endoscopists post-August 2016*

The performances of all endoscopists versus the four DCE trained endoscopists post-August 2016 are shown in Table 5-13. Improvements in the use of DCE (63.8% vs. 78.7%  $p<0.05$ ), formal description of endoscopic activity (29.3% vs. 50.2%  $p<0.01$ ), use of the Paris classification (23.1% vs. 62.1%  $p<0.01$ ) and the Kudo Pit Pattern (15.4% vs. 66.5%  $p<0.01$ ) were seen in the DCE trained group.

**Table 5-10 Summary of performance in IBD quality indicators pre- and post-August 2016 in all endoscopists and the 4 DCE trained endoscopists**

| Quality measure                                                                        | Pre-August 2016<br>N=277 | All colonoscopists<br>post-August 2016<br>N=321 | 4 DCE trained<br>endoscopists post-<br>August 2016<br>N=263 |
|----------------------------------------------------------------------------------------|--------------------------|-------------------------------------------------|-------------------------------------------------------------|
| EXTENT OF COLONOSCOPY n (%)                                                            |                          |                                                 |                                                             |
| Caecal intubation                                                                      | 265 (95.7)               | 314 (97.8) NS                                   | 257 (97.7)                                                  |
| Terminal ileum intubation                                                              | 58 (20.9)                | 147 (45.8) ***                                  | 130 (49.4) ***                                              |
| COLONOSCOPY PROCEDURE n (%)                                                            |                          |                                                 |                                                             |
| Dye spray chromoendoscopy                                                              | 150 (54.2)               | 244 (76.0) ***                                  | 207 (78.7) ***                                              |
| Random biopsy surveillance                                                             | 34 (12.3)                | 10 (3.1) ***                                    | 6 (2.3) ***                                                 |
| ASSESSMENTS n (%)                                                                      |                          |                                                 |                                                             |
| Endoscopic disease activity score                                                      | 7 (2.5)                  | 149 (46.4) ***                                  | 132 (50.2) ***                                              |
| Size of lesion documentation                                                           | 45 (65.2)                | 79 (79.0) *                                     | 70 (80.5) *                                                 |
| Paris classification                                                                   | 18 (26.1)                | 57 (57.0) ***                                   | 54 (62.1) ***                                               |
| Kudo pit pattern                                                                       | 15 (21.7)                | 59 (59.0) ***                                   | 57 (65.5) ***                                               |
| Border of lesions                                                                      | 0 (0.0)                  | 18 (18.0) ***                                   | 18 (20.7) ***                                               |
| Ulcerated surface of lesions                                                           | 0 (0.0)                  | 3 (3.0) NS                                      | 3 (3.4) NS                                                  |
| DETECTION n (%)                                                                        |                          |                                                 |                                                             |
| Lesion detection                                                                       | 69 (24.9)                | 100 (31.2) NS                                   | 87 (33.1) *                                                 |
| Dysplasia detection                                                                    | 20 (7.2)                 | 32 (10.0) NS                                    | 30 (11.4)                                                   |
| P value: * $p < 0.05$ , ** $p < 0.01$ . *** $p < 0.0005$ compared with pre-August 2016 |                          |                                                 |                                                             |

**Table 5-11 A summary of the reasons DCE could not be used for procedures between January 2014 and October 2018**

| Reason                                                      | Pre-august 2016<br>(n=32) | Post-August 2016<br>(n=37) | P value |
|-------------------------------------------------------------|---------------------------|----------------------------|---------|
| Insufficient bowel preparation                              | 29 (90.6%)                | 17 (45.9%)                 | 0.018   |
| Disease activity                                            | 2 (6.3%)                  | 10 (27.0%)                 | 0.037   |
| Both disease activity and<br>insufficient bowel preparation | 0 (0.0%)                  | 7 (18.9%)                  |         |
| Colonic stricture                                           | 0 (0.0%)                  | 3 (8.1%)                   |         |
| Insufficient procedural time<br>allocated                   | 1 (3.1%)                  | 0                          |         |

**Table 5-12 Breakdown of detected lesions**

|              | Pre-August 2016 | Post-August 2016 | P value |
|--------------|-----------------|------------------|---------|
| LGD          | 20              | 26               | 0.687   |
| HGD/Cancer   | 0               | 3                |         |
| Hyperplastic | 6               | 23               | 0.005   |
| Pseudopolyps | 22              | 17               | 0.191   |
| SSA          | 0               | 3                |         |

**Table 5-13 Performance of DCE trained endoscopists vs. all other endoscopists post-August 2016**

| Quality measure                                                                        | Non-DCE trained<br>endoscopists<br>N=58 | DCE trained endoscopists<br>post-August 2016<br>N=263 |
|----------------------------------------------------------------------------------------|-----------------------------------------|-------------------------------------------------------|
| EXTENT OF COLONOSCOPY n (%)                                                            |                                         |                                                       |
| Caecal intubation                                                                      | 57 (98.3)                               | 257 (97.7)                                            |
| Terminal ileum Intubation                                                              | 17 (29.3)                               | 130 (49.4) **                                         |
| Co-existence of PSC n (%)                                                              | 26 (44.8)                               | 108 (41.1)                                            |
| COLONOSCOPY PROCEDURE n (%)                                                            |                                         |                                                       |
| Dye spray chromoendoscopy                                                              | 37 (63.8)                               | 207 (78.7) *                                          |
| Random biopsy surveillance                                                             | 2 (3.4)                                 | 6 (2.3)                                               |
| Endoscopic disease activity score                                                      | 17 (29.3)                               | 132 (50.2) **                                         |
| Size of lesion documentation                                                           | 9 (69.2)                                | 70 (80.5)                                             |
| Paris classification                                                                   | 3 (23.1)                                | 54 (62.1) **                                          |
| Kudo pit pattern                                                                       | 2 (15.4)                                | 57 (66.5) **                                          |
| Border of lesions                                                                      | 0 (0.0)                                 | 18 (20.7)                                             |
| Ulcerated surface of lesions                                                           | 0 (0.0)                                 | 3 (3.4)                                               |
| Lesion detection                                                                       | 13 (22.4)                               | 87 (33.1)                                             |
| Dysplasia detection                                                                    | 2 (3.4)                                 | 30 (11.4)                                             |
| P value: * $p < 0.05$ , ** $p < 0.01$ . *** $p < 0.0005$ compared with pre-August 2016 |                                         |                                                       |

### 5.5.5 Discussion

Surveillance colonoscopy in IBD is crucial for detecting and managing dysplasia/cancer, much like CRC screening programmes in the general population. The challenges faced are the background mucosal abnormalities and often subtle, flat lesions. Whilst there are guidelines on when and how surveillance should be performed, there is a lack of guidance regarding how to optimise surveillance and which key quality indicators should be considered in IBD patients (Cairns, Scholefield et al. 2010, Laine, Kaltenbach et al. 2015, Magro, Gionchetti et al. 2017). The introduction of quality indicators in the CRC screening population has led to an improvement in performance (Lee, Rutter et al. 2012), and has significantly impacted patients' clinical outcomes and quality of life. Quality measures should also be proposed and identified as a priority for surveillance colonoscopy in IBD. DCE is recommended as standard and should be considered as a key quality indicator in line with these recommendations (Laine, Kaltenbach et al. 2015, Magro, Gionchetti et al. 2017). However, multiple barriers and practical issues are related with a lack of wide adoption of DCE which may be due to lack of adequate training, increased time and costs (Kaltenbach, Soetikno et al. 2017).

We have shown that after focussing on DCE as standard of care with dedicated lists and by restricting IBD surveillance to DCE trained endoscopists, introduction of foot pedal operated pump-jets and appropriate training led to a significant improvement in its use and therefore improved quality of IBD surveillance colonoscopy. Indeed, the proportion of procedures using DCE in our study significantly increased and random biopsy surveillance reduced in favour of targeted biopsies from pre-August 2016 to post-August 2016 in hands of DCE trained endoscopists after this implementation.

The most common reason for not performing DCE was inadequate bowel preparation and the presence of disease activity, which is consistent with other studies (Megna, Weiss et al. 2018). There was overall no difference in the proportion of colonoscopies limited by poor bowel preparation pre- and post- August 2016, however the proportion of procedures not using DCE due to disease activity increased, which likely reflects the increase in DCE-trained endoscopists reporting, who may feel more confident report colitis activity. Bowel preparation remains one of the main limitations to adopting DCE in >90% of IBD surveillance procedures. Endoscopy units may consider organising bowel preparation for IBD surveillance comparable to Bowel Cancer Screening Programmes to optimise procedural efficiency which will be cost effective and impact patients' quality of life.

Major emphasis should be focused on training, especially on the SCENIC recommendations on how to characterise IBD lesions with the use of Paris classification, Kudo pit pattern, FACILE classification, description of lesion margins and the presence of an ulcerated surface (Kudo, Hirota et al. 1994, Endoscopic Classification Review 2005, Laine, Kaltenbach et al. 2015, Iacucci, McQuaid et al. 2018). An educational training module in IBD lesions led to a significant improvement in key quality indicators, in particular lesions detected with an increased use of morphology descriptors (table 5.7). Educational classification systems, such as FACILE (Frankfurt Advanced Chromoendoscopic Ibd Lesions), which has been shown to improve diagnostic performance in all non-expert groups, will improve detection rate and accurate characterisation of IBD lesions (Iacucci, McQuaid et al. 2018). Training and education can play a key role in improving the ability to detect, characterise IBD neoplasia and aid endoscopists to decide on the right therapeutic approach (Sanduleanu, Kaltenbach et al. 2016). Moreover, training can disseminate information on best practices and should be

considered in endoscopy units to optimize quality of surveillance colonoscopy in IBD patients.

In this study despite the service changes and training provided, a small minority (9.3%) of procedures did not incorporate a formal surveillance technique nor documented reasons for not using DCE. Despite best efforts and given clinical service demands, some procedures were allocated to endoscopists whom are not formally trained in DCE, which may explain the 9.3% of procedures with sub-optimal surveillance practice. This highlights the need for well-structured strategies to ensure well trained specialists in IBD perform IBD surveillance colonoscopy by using DCE. International organizations should develop a curriculum on how surveillance should be performed and how detected colonic lesions are managed that will ensure uniform high-quality performance.

Our results are in line with other studies demonstrating the impact of an educational event increasing the use of DCE (Kaltenbach, Soetikno et al. 2017). A study conducted at three Mayo Clinic sites involving six endoscopists without any experience in chromoendoscopy had shown that after training with images, the interobserver agreement for determination of polypoid abnormalities in patients with long-standing UC was excellent (Picco, Pasha et al. 2013).

There is a significant focus on ADR in the bowel cancer screening population (Lee, Rutter et al. 2012, Rex, Schoenfeld et al. 2015) and lesion detection should be considered a quality measure in IBD surveillance with defined performance metrics. In this study there was an improvement in lesion detection between the pre- and post-August 2016 groups which may lead to an improved dysplasia detection rate over time.

There are limitations to this study. Firstly, being a single tertiary referral centre may make generalising the results to other units challenging. Secondly, the retrospective nature does bring potential bias, particularly as the audit is reliant on the data entered at the time of the procedure. Withdrawal times were not routinely documented during the period of the study and are not included in this dataset. The procedures performed in our centre used high-definition endoscopes (Olympus 260/290 series), without the use of magnifying endoscopy. Magnifying endoscopy is a promising modality in the optical diagnosis of IBD lesions (Kawasaki, Nakamura et al. 2019, Oka, Uraoka et al. 2019), however we are unable to comment on the role of this technology in IBD surveillance. However, given that magnifying endoscopy is not in widespread use in the West our results are likely to reflect “real-life” practice. The new generation of Olympus (Tokyo, Japan) 290 series endoscopes with near focus allows magnified images and is easier to use which may promote the adoption of magnification in the West. Given the retrospective design, it was not possible to comment on the influence of operator experience. All endoscopists were considered experienced however detailed statistics on lifetime colonoscopy numbers and number of procedures per annum was not possible.

In conclusion, by implementing specific changes, performance in quality indicators can be improved. It highlights how education, training and dedicated DCE trained endoscopists performing IBD surveillance colonoscopies can significantly change practice. Areas to improve are ensuring all procedures are allocated to DCE-trained endoscopists, ensuring bowel preparation is optimum, colitis activity is recorded and is optimised for surveillance, lesion morphology and characterization descriptors as mandatory. These changes need to be implemented at a unit organisation level but also at the individual endoscopist level through effective educational modules. Further audit, training and education may improve adoption of

quality indicators. More formal guidance on quality measures in IBD surveillance has been published (Iacucci, Cannatelli et al. 2019), however international consensus on IBD surveillance measures will also ensure uniform high quality of colonoscopy.

The section above is taken from the following publication:

**Smith SCL**, Cannatelli R, Bazarova A, Sharma N, McCulloch A, Mak J, Shivaji UN, Iqbal T, Kane K, Ghosh S, Cooney R, Iacucci M. Performance measures in IBD surveillance colonoscopy: Implementing changes to practice improves performance. *Dig Endoscopy* 2020; 32: 592-599

## 6 CONCLUSIONS AND FUTURE WORK

### 6.1 Overview

In this thesis I have discussed the role of endoscopy in IBD management and the current recommended treatment paradigm. I have presented advanced endoscopic technologies, in particular virtual electronic chromoendoscopy (VEC), and how they can be applied in assessing inflammatory activity associated with inflammatory bowel disease and colorectal polyps. In addition to this I provided an overview of a novel technology, namely Raman Spectroscopy, which may provide additional data on the molecular make up of tissue. This may allow for more precise assessments of mucosal healing in IBD in the future. I have presented the findings of four studies examining the clinical use of VEC in assessing inflammatory activity in UC, a novel application of Raman Spectroscopy in differentiating MH from active inflammation in IBD and the implementation of training in using advanced endoscopic technology in the IBD and non-IBD population. In this chapter I summarise the results and discuss the clinical implications and future research in the field.

### 6.2 The real-life prospective validation study of the PICaSSO score.

In the second chapter of this thesis, I provide an overview of IBD and the associated morbidity. To prevent complications associated with the disease, the “treat to target” strategy that has been proposed for the management of patients with IBD (Peyrin-Biroulet, Sandborn et al. 2015, Turner, Ricciuto et al. 2021). This means clinicians are increasingly aiming to achieve MH and I presented the literature that confirms the difficulty with current definitions of MH.

When looking at UC, current definitions of mucosal healing using MES, leave a lot to be desired with an unacceptable high relapse rate, and include definitions that do not truly reflect histological remission. As mentioned earlier, novel technologies can obtain imaging that can increasingly reveal the histology of tissue. For this reason, the PICaSSO score was developed, and I presented the first real-life validation study of this score in Chapter 3. Based on the previous work by Iacucci and colleagues, PICaSSO was developed and validated in an ex vivo setting using endoscopic videos (Iacucci, Daperno et al. 2017). In Chapter 3, following the recruitment of over 300 patients with UC, I demonstrated that the PICaSSO score correlates strongly with histology, using 5 existing histological scores. Previous attempts at correlating endoscopic scores with histological activity are often limited by being single centre and single endoscopist, however the study presented represents the largest validation study conducted in UC and by including multiple endoscopists/pathologists and histology scores makes the results more generalisable.

I presented that the PICaSSO score can accurately predict histological remission, using a score of  $\leq 3$ . This describes endoscopic findings in keeping with quiescent UC, such as the presence of elongated crypts and sparse deep vessels, which are findings that are visible using VEC. The accurate prediction of MH could prove to revolutionise the assessment of inflammatory activity in UC. As previously mentioned, knowing when to escalate or de-escalate therapy can be challenging. If features of MH are present using the PICaSSO score this will give more confidence to clinicians to either maintain current therapy, or consider de-escalation, a so-called drug holiday. However, the latter is more difficult since there is no widely accepted “exit strategy” and currently the PICaSSO score cannot be used in this manner without further studies.

The assessment of MH is a field of great interest for the reasons mentioned in Chapter 2 and is currently an area that is inadequately covered by existing endoscopic scores. At the severe end of the inflammatory spectrum, it is unlikely that VEC will contribute, since when ulcerations and/or luminal bleeding are present these are often visible under WLE. However, it is at the mild/quiescent end of the spectrum is where there is added value of VEC over white light. It is here where subtle inflammatory features can be detected, and also features in keeping with healing as suggested by *Iacucci et al* (Iacucci, Fort Gasia et al. 2015).

I have discussed the clinical implication and use of the PICaSSO going forward, in terms of guiding therapy decision making, and it could also be utilised in clinical trials. The most commonly used endoscopic end point in UC used in clinical trials is the MES. The PICaSSO score has superior correlation with histology and prediction of histological remission and therefore could be utilised in trials, which is seen as a key end point. Given the excellent inter-rater agreement, the PICaSSO score could also be considered for central read-out for clinical trials.

It is very probable that in the near future artificial intelligence will become a part of everyday endoscopic practice. There is extensive evidence of its use in the non-IBD population, most notably in the detection and characterisation of colorectal polyps (Hassan, Spadaccini et al. 2021). There is increasing interest in the use of artificial intelligence in IBD and in particular predicting inflammatory activity. A study by *Stidham et al*, aimed to assess whether deep learning algorithms can assess severity of UC when applied to endoscopic videos (Stidham, Liu et al. 2019). The authors included a large set of images of colonic mucosa of UC patients (n=16514) for constructing a convolutional neural network to train a deep learning model. The MES was used to classify images into remission (MES 0 and 1) or moderate-severe

(MES 2 and 3). The model was able to distinguish between the two groups with an AUROC of 0.966 (95% CI 0.967-0.972) and kappa agreement of 0.84 (95% CI 0.83-0.86) in predicting exact mayo scores, which was similar to that of experienced human reviewers (0.86 95% CI 0.85-0.87) (Stidham, Liu et al. 2019). When applied to endoscopic videos, a similar accuracy in predicting moderate-severe disease was found (Stidham, Liu et al. 2019). This study demonstrate that AI models could have a role in IBD, however it does have several limitations. The most relevant limitation is largely through the use of the MES, and categorising patients into either remission or moderate-severe. It can be argued that AI models are not required in differentiating this degree of inflammation, but a more clinically relevant question is whether AI models are more superior in differentiating quiescent from mild disease (MES 0 from MES1). For this reason, the PICaSSO score, with its increased bandwidth at the quiescent end of the spectrum, could be adopted for the development and training of artificial neural networks, for example using PICaSSO  $\leq 3$  as quiescent disease and 4-8 for mild disease.

A further study by *Ozawa et al* also developed a deep learning model in the form of convolutional neural network, which was trained using 26,304 images from 841 UC patients (Ozawa, Ishihara et al. 2019). The images captured were classified by at least two of the authors using the MES, with images discarded when a consensus was not met. Histological inflammation was graded using the Matts grade. Following training, the model was tested using 3981 images from 114 UC patients, and 73% of MES 0, 70% of MES 1 and 63% of MES 2-3 were correctly classified, giving an AUROC of 0.86 when differentiating MES 0 from MES 1-3 and 0.98 from MES 0-1 from 2-3 (Ozawa, Ishihara et al. 2019). This retrospective study was limited by a comparatively low proportion of moderate-severe UC.

Given the limitations of the MES, training a model using the PICaSSO score may equate to even further improvements in diagnostic performance.

One of the main limitations of the studies examining the role of AI in assessing inflammatory activity is the tool used to train the neural networks. The neural networks will only perform as well as the data used to train them will allow. At present it is reliant on human interpretation and scoring, in the studies mentioned, using the MES. It is widely appreciated that the MES has a limited interobserver agreement (Fernandes, Pinto et al. 2018). To improve the training data, the PICaSSO score could be used, given the superior prediction of MH, correlation with histology and inter-observer agreement.

A novel technique has been, described by *Bossuyt et al*, whereby a computer algorithm was developed using pixel data from the red channel, so-called red density (Bossuyt, Nakase et al. 2020). Using the data from 29 UC patients and 6 non-IBD for training, the red density correlated well with RHI, MES and UCEIS when tested on patients with UC patients requiring treatment escalation (Bossuyt, Nakase et al. 2020). This technology has the added advantage of being truly objective and does not involve human assessment which may infer bias, particularly when aware of the clinical background. However, by testing on patients with active disease requiring treatment escalation, means that it remains untested in the assessment of MH, which remains a bigger problem than assessing moderate-severe disease, for which existing technologies are likely to be sufficient. It will be interesting to see how this technology fairs in real-world studies.

The exit strategy of treatment in IBD, in particular the biologic agents, remains a challenging topic. Treatment escalation strategies are well established in the presence of inflammation, however when therapies can be stopped, to minimise treatment risks to patients and healthcare costs, are less clear. Patients who are in clinical, biochemical and endoscopic remission are more likely to stay in remission (Doherty, Katsanos et al. 2018). Definitions of endoscopic remission can be debated, as covered in Chapter 2, however it could be argued that the PICaSSO score is in a stronger position to define remission. A randomised trial comparing existing definitions of MH (i.e., MES 0/1) vs. PICaSSO ( $\leq 3$ ), and stopping therapy when this is achieved. Patients could be followed up over 12-24 months and clinical and endoscopic relapse could be compared.

### 6.3 Raman Spectroscopy

In the fourth Chapter of this thesis, I presented an overview of Raman Spectroscopy, and how it can provide biomolecular information on the make-up of tissue. The main appeal of Raman Spectroscopy is its non-destructive nature, and therefore can be considered for *in vivo* use. This together with short acquisition time means there is increasing interest in its clinical application. I have presented the studies that have completed to date in the area of IBD, and its use in differentiating CD from UC and in grading levels of inflammatory activity in UC. The diagnostic performances of these studies are mixed which is likely explained by the heterogeneity of the number of samples collected, the grading of inflammatory activity and the limitations of using PCA in spectral analysis.

I have discussed the evolving ability to predict histology using novel endoscopic technologies, in particular through the use of VEC given its widespread availability and easy

access for everyday endoscopic use. Despite these developments, there are still limitations. In Chapter 3 I presented the findings of the validation study of the PICaSSO score and despite advanced endoscopic features of MH, which has been shown to be highly accurate, there is still a proportion of patients who suffer clinical relapse within 12 months.

As previously mentioned, it is possible that there are molecular changes which have not manifested at the endoscopic nor histological level. Raman Spectroscopy is a technique that can reveal the molecular “make-up” of biological tissue. I have presented for the first time, that Raman Spectroscopy can differentiate active inflammation from MH, and I found spectral features in keeping with MH. Of note, the increase at  $1304\text{cm}^{-1}$  when mucosal healing was present in both UC and CD, and seen between in the post-biologic spectra, may become clinically relevant. A scenario whereby Raman Spectroscopy may provide a useful adjunct, is when endoscopists feel there are endoscopic signs of mucosal healing- if Raman Spectroscopy was performed and showed increases at the  $1304\text{ cm}^{-1}$  then this may reflect more robust healing data and treatment decisions such as maintaining or de-escalating current therapy as opposed to increasing dosages or switching class may take place. However, if increases at this Raman band were not seen, then clinicians may opt to keep a patient under close monitoring or increase treatment to minimise the risk of relapse. This is currently theoretical and further evidence needs to be collected before this can become reality in everyday clinical practice, in particular *in vivo* studies.

The study I presented on Raman Spectroscopy is the first utilising the newly developed SKiNET in the field of IBD. As previously mentioned, SKiNET describes a machine learning technique that can present the complexity of Raman Spectroscopy data in a simple easy to understand format. This is in the form of SOM, which is a 2D representation of hyper-

dimensional data. It then presents the areas of the Raman spectrum that create the separation between two tissue types seen in the SOM, and finally a model can be trained and diagnostic performances in differentiating the two can be tested.

Going forward a potential area of interest would be in utilising models, that have been built in the *ex vivo* setting to *in vivo* use. An example would be if endoscopic features of mucosal healing were seen, Raman Spectra could be acquired and the spectra on the trained model could be observed, for example does it appear within the blue areas (mucosal healing) or red areas (active inflammation). This would then perhaps guide future treatment decisions. To ascertain whether Raman Spectroscopy can detect biomolecular changes that may predict early relapse in IBD, a study recruiting patients with MH at endoscopy with Raman Spectroscopy analysed followed by observation of patients over 12-24 months. This would then allow a comparison of spectra of patients who suffered early relapse with those who remained in remission. An artificial intelligence model could be created which would be useful in clinical practice and provide additional information to clinicians in deciding which patients require a step-up in therapy or close monitoring.

A further area for future research could be in the prediction of IBD lesion histology. Polyps that are removed from IBD patients could be analysed using Raman Spectroscopy initially in *ex vivo* setting and make direct comparisons to background quiescent and active IBD. It would be interesting to see if Raman Spectroscopy can differentiate IBD lesions from background mucosa, since this will assist endoscopists in predicting the borders of IBD lesions, which are often challenging to delineate through endoscopic appearances alone. This will then allow more complete endoscopic resections of lesions.

## 6.4 Training in advanced endoscopy

The advances in endoscopic technology allows more precise predictions of histology, both in terms of inflammatory activity and polyp type. This can translate into enhanced patient management and care if endoscopists are trained and utilise it in everyday clinical practice. Given the clear benefits of optical diagnosis in practicing the “Resect and discard” strategy, there will be healthcare cost savings and reduced risks to patients. The recognition of advanced histology is an important topic and improved optical diagnosis skills will mean advanced histology is recognised and allow limited targeted biopsies of suspicious areas. This minimises fibrosis to the mucosa and allow safer and easier endoscopic removal if appropriate. Targeted biopsies also mean the correct diagnosis is more likely, as it is possible for an area of low-grade dysplasia to be sampled within a lesion of advanced histology, such as carcinoma. This often means the patient is committed to having a further colonoscopy for polyp removal before correct recognition of carcinoma, therefore delaying the correct management plan for the patient.

In Chapter 5 I presented a systematic review of the existing evidence of training methods in optical diagnosis of small/diminutive colorectal polyps. It is clear that without training, non-expert endoscopists do not have the necessary skills to meet the thresholds for the “Resect and discard” strategy. The evidence from the studies I presented, including the randomised controlled trial, shows that optical diagnosis training is effective in improving accuracy of histology prediction. The body of evidence behind the available training methods is low, however, given the heterogeneity and bias.

The field of learning and training is a complex one with many variables, not least individual learners own cognitive ability and learning style, therefore it is unlikely that a one-size fits all

training approach will be optimal. Effective training may involve multiple methods. Didactic training from an expert in the field is appealing to introduce optical diagnosis and pass on tips based on previous experience and individual feedback. Given the limitations of this way of training, as previously discussed in Chapter 5, further training through computer-based training, which allows users to learn at their own pace and builds on the knowledge acquired through didactic training. As mentioned, this may not be sufficient and *in vivo* training from trainers for those endoscopists in training, however this does rely on the trainer's ability in optical diagnosis. A method that may prove effective, albeit challenging to implement, might be feedback on optical diagnosis predictions together with images and histology of the polyp. Whilst this is possible currently, it requires effort and dedication on behalf of the busy endoscopist. This could be used to disseminate images and polyp histology more widely allowing others to learn. Whichever method is utilised, regular auditing of practice is essential to ensure safe practice and highlight areas for further learning. A more realistic adjunct, to currently available training methods, is the training possible from artificial intelligence. This approach may lend itself well to initial baseline training from computer-based training. This is an area that requires further study.

I also presented the findings of an observational study, comparing performance in quality indicators of IBD surveillance before and after service changes and training in IBD surveillance and lesions. Bowel cancer screening in the non-IBD population has been a success story in early detection of bowel lesions that can be managed endoscopically or limited surgery, thereby preventing advanced malignancy and improving outcomes for patients. This was done by means of a series of measures, including rigorous assessments for endoscopists and strict measurement of performance indicators, for example caecal intubation

and adenoma detection rate. At present, no such standards exist for IBD surveillance despite the goals are similar: early detection of colorectal lesions.

I demonstrated that through relatively simple measures in service provision and a training session, the performance in key quality indicators can be improved. By treating IBD surveillance in a similar fashion to bowel cancer screening, patient care can be enhanced. Restricting the procedure to endoscopists who are dedicated and trained in DCE is likely an independent factor in improving quality and represents a simple way to improve outcomes. IBD lesions are often subtle and can be challenging to detect and characterise, therefore having dedicated endoscopists receive training on methods to characterise lesions such as the FACILE classification (Iacucci, McQuaid et al. 2018) can lead to improvements in quality. Indeed, in the development of FACILE, training was provided to endoscopists and trainees, which led to an improvement in accuracy of optical diagnosis of IBD lesions (Iacucci, McQuaid et al. 2018). Rather disappointingly that despite a one-off didactic training session the interobserver agreement amongst trainees was moderate (Iacucci, McQuaid et al. 2018), which further raises the question as to what the optimum training method is. Much like in the non-IBD population, optical diagnosis is crucial in optimising management of patients, not least in accurate prediction of inflammatory activity but also in recognising lesion borders and correct prediction of advanced histology. The optical diagnosis of IBD lesions is a much less investigated field compared with non-IBD polyps, and development of training materials and conducting randomised trials comparing didactic with computer-based training is a potential future interest.

There have been and continue to be significant developments in artificial intelligence systems in polyp detection and characterisation. Several studies demonstrate the effectiveness of

computer-aided diagnosis utilising advanced endoscopic technologies such as magnification endoscopy (Takemura, Yoshida et al. 2012, Kominami, Yoshida et al. 2016) and endocytoscopy (Mori, Kudo et al. 2015, Mori, Kudo et al. 2016), however given the majority of centres do not have the ability to use these technologies these advancements may be limited to few academic units. VEC technologies such as NBI, BLI and iScan-OE represent more widely available and acceptable technologies to use in conjunction with artificial intelligence (Mesejo, Pizarro et al. 2016). *Hassan et al* assessed the ability of AI to detect colorectal polyps by training an AI system using 2684 videos of polyps under HD WLE. This system was able to detect polyps with a sensitivity of 99.7% and was able to react in less time when compared with 5 expert endoscopists (Hassan, Wallace et al. 2019). A recent meta-analysis confirms the ability of AI in accurately predicting histology, with a pooled accuracy of 0.96 (95% CI 0.95-0.98), sensitivity 92.35 (95% CI 88.8-94.9) and specificity 89.8% (95% CI 85.3-93.0) (Lui, Guo et al. 2020). The performance of AI was superior to non-expert endoscopists, achieving a diagnostic NPV of >90% and achieved an area under the receiving operator curve of 0.90 (95% CI 0.67-1.00) for polyp detection (Lui, Guo et al. 2020). As technology evolves and rapid development and interest in artificial intelligence models, we may be in a better position to employ a “Resect and Discard” strategy.

It is very likely that artificial intelligence will be in everyday use during endoscopy in the near future and can be seen as an aid to training and development of optical diagnosis skills. The ESGE suggest the most likely scenario will be as a “second reader” as opposed to a “stand alone” system (East, Vleugels et al. 2016). This will mean that an AI model will alert the endoscopist to a lesion and give a prediction of the histology with a degree of confidence. It will then be up to the endoscopist as to how to manage the lesion. This means that training in *in vivo* optical diagnosis may be provided in part by AI models, which would be an easier

transition if endoscopists are already accustomed to computer-based training. Irrespective of how AI will be incorporated, there is an increasing need for endoscopists to up-skill in optical diagnosis.

## 6.5 Conclusion

In conclusion the work in this thesis demonstrates the advancements in endoscopic technology and can provide endoscopists with the tools to accurately predict inflammatory activity, and therefore help predict the clinical course of patients. This will ultimately assist clinicians to more accurately determine which patients have achieved MH and who may benefit changes in therapy due to continuing inflammatory activity to minimise the risk of adverse outcomes. This could change practice in the real world since the technology employed is currently readily available to endoscopists.

Despite these advancements and strong correlations with histology, there remains patients who relapse despite strong evidence of being in remission. Therefore, there is an additional need for further molecular assessments to fully risk stratify patients in terms of relapse risk and response to treatment. One such technology may include Raman Spectroscopy, which can provide insightful data on the molecular make-up of biomolecular tissue. I have demonstrated that Raman Spectroscopy can differentiate patients with MH from those with active inflammation in IBD. This has the potential to be taken *in vivo* and when combined with endoscopic imaging a more robust assessment of MH can be obtained, ensuring patients management is optimised.

Endoscopic assessments of inflammatory activity have limitations previously discussed. Molecular characterisation could provide pivotal, with Raman Spectroscopy being an example. Further to this I have involved in work, though not part of this thesis, identifying gene expression markers, such as macrophage inflammatory proteins and soluble vascular cell adhesion molecule 1, that predicted histological remission (Iacucci, Jeffery et al. 2021).

It is clear that artificial intelligence will become part and parcel of everyday endoscopy. Its use is becoming more common in non-IBD polyp detection and characterisation. The future may consist of artificial intelligence using a combination of technologies such as VEC and molecular assessments, such as Raman Spectroscopy, to provide thorough data on patient's disease status. However, using a similar analogy in the aviation industry; auto-pilots and human pilots co-exist and enhance safety but there is still a need to train human pilots. The same applies in endoscopy. The work in this thesis shows that training and service changes can improve optical diagnosis and quality of endoscopy. Training will be essential and is likely to be adjunctive to that provided by artificial intelligence systems.

## 7 List of References

- Abu Dayyeh, B. K., N. Thosani, V. Konda, M. B. Wallace, D. K. Rex, S. S. Chauhan, J. H. Hwang, S. Komanduri, M. Manfredi, J. T. Maple, F. M. Murad, U. D. Siddiqui and S. Banerjee (2015). "ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps." *Gastrointest Endosc* **81**(3): 502.e501-502.e516.
- Addis, J., N. Mohammed, O. Rotimi, D. Magee, A. Jha and V. Subramanian (2016). "Raman spectroscopy of endoscopic colonic biopsies from patients with ulcerative colitis to identify mucosal inflammation and healing." *Biomed Opt Express* **7**(5): 2022-2035.
- Aihara, H., N. Kumar and C. C. Thompson (2018). "A Web-Based Education Program for Colorectal Lesion Diagnosis with Narrow Band Imaging Classification." *Digestion* **98**(1): 11-18.
- Almond, L. M., J. Hutchings, G. Lloyd, H. Barr, N. Shepherd, J. Day, O. Stevens, S. Sanders, M. Wadley, N. Stone and C. Kendall (2014). "Endoscopic Raman spectroscopy enables objective diagnosis of dysplasia in Barrett's esophagus." *Gastrointest Endosc* **79**(1): 37-45.
- Annese, V., M. Daperno, M. D. Rutter, A. Amiot, P. Bossuyt, J. East, M. Ferrante, M. Gotz, K. H. Katsanos, R. Kiesslich, I. Ordas, A. Repici, B. Rosa, S. Sebastian, T. Kucharzik, R. Eliakim, C. s. European and O. Colitis (2013). "European evidence based consensus for endoscopy in inflammatory bowel disease." *J Crohns Colitis* **7**(12): 982-1018.
- Arai, M., M. Naganuma, S. Sugimoto, H. Kiyohara, K. Ono, K. Mori, K. Saigusa, K. Nanki, M. Mutaguchi, S. Mizuno, R. Bessho, Y. Nakazato, N. Hosoe, K. Matsuoka, N. Inoue, H. Ogata, Y. Iwao and T. Kanai (2016). "The Ulcerative Colitis Endoscopic Index of Severity is Useful to Predict Medium- to Long-Term Prognosis in Ulcerative Colitis Patients with Clinical Remission." *J Crohns Colitis* **10**(11): 1303-1309.
- Baars, J. E., V. J. Nuij, B. Oldenburg, E. J. Kuipers and C. J. van der Woude (2012). "Majority of patients with inflammatory bowel disease in clinical remission have mucosal inflammation." *Inflamm Bowel Dis* **18**(9): 1634-1640.
- Bae, J. H., C. Lee, H. Y. Kang, M. S. Kwak, E. Y. Doo, J. Y. Seo, J. H. Song, S. Y. Yang, J. I. Yang, S. H. Lim, J. Y. Yim, J. H. Lim, G. E. Chung, S. J. Chung, E. H. Jin, B. Park and J. S. Kim (2019). "Improved Real-Time Optical Diagnosis of Colorectal Polyps Following a Comprehensive Training Program." *Clin Gastroenterol Hepatol* **17**(12): 2479-2488 e2474.
- Bae, J. H., C. Lee, H. Y. Kang, M. S. Kwak, E. Y. Doo, J. Y. Seo, J. H. Song, S. Y. Yang, J. I. Yang, S. H. Lim, J. Y. Yim, J. H. Lim, G. E. Chung, S. J. Chung, E. H. Jin, B. Park and J. S. Kim (2019). "Improved Real-Time Optical Diagnosis of Colorectal Polyps Following a Comprehensive Training Program." *Clin Gastroenterol Hepatol*.
- Banbury, C., R. Mason, I. Styles, N. Eisenstein, M. Clancy, A. Belli, A. Logan and P. Goldberg Oppenheimer (2019). "Development of the Self Optimising Kohonen Index Network (SKiNET) for Raman Spectroscopy Based Detection of Anatomical Eye Tissue." *Sci Rep* **9**(1): 10812.
- Baron, J. H., A. M. Connell and J. E. Lennard-Jones (1964). "Variation between Observers in Describing Mucosal Appearances in Proctocolitis." *Br Med J* **1**(5375): 89-92.

- Barreiro-de Acosta, M., N. Vallejo, D. de la Iglesia, L. Uribarri, I. Baston, R. Ferreiro-Iglesias, A. Lorenzo and J. E. Dominguez-Munoz (2016). "Evaluation of the Risk of Relapse in Ulcerative Colitis According to the Degree of Mucosal Healing (Mayo 0 vs 1): A Longitudinal Cohort Study." J Crohns Colitis **10**(1): 13-19.
- Basford, P., J. Brown, S. Cooper and P. Bhandari (2019). "Endoscopic characterization of small colonic polyps: baseline performance of experienced endoscopists is no different to that of medical students." Endosc Int Open **7**(4): E403-E411.
- Basford, P., G. Longcroft-Wheaton, R. Higashi, T. Uraoka and P. Bhandari (2017). "Colonic lesion characterisation skills among UK endoscopists and the impact of a brief training intervention." Frontline Gastroenterol **8**(1): 2-7.
- Beaugerie, L., P. Seksik, I. Nion-Larmurier, J. P. Gendre and J. Cosnes (2006). "Predictors of Crohn's disease." Gastroenterology **130**(3): 650-656.
- Bergeron, V., A. Vienne, H. Sokol, P. Seksik, I. Nion-Larmurier, A. Ruskone-Fourmestreaux, M. Svrcek, L. Beaugerie and J. Cosnes (2010). "Risk factors for neoplasia in inflammatory bowel disease patients with pancolitis." Am J Gastroenterol **105**(11): 2405-2411.
- Bessho, R., T. Kanai, N. Hosoe, T. Kobayashi, T. Takayama, N. Inoue, M. Mukai, H. Ogata and T. Hibi (2011). "Correlation between endocytoscopy and conventional histopathology in microstructural features of ulcerative colitis." J Gastroenterol **46**(10): 1197-1202.
- Bessissow, T., B. Lemmens, M. Ferrante, R. Bisschops, K. Van Steen, K. Geboes, G. Van Assche, S. Vermeire, P. Rutgeerts and G. De Hertogh (2012). "Prognostic value of serologic and histologic markers on clinical relapse in ulcerative colitis patients with mucosal healing." Am J Gastroenterol **107**(11): 1684-1692.
- Bhandari, P., S. Thayalasekaran, R. Keisslich, R. Bisschops, A. Hoffmann, R. Haidry, J. Esteban, B. Lopez Viedma, E. Godzhello, M. Almadi, H. Neumann and S. Sanduleanu (2018). "Detection and characterization of colorectal polyps using high-definition white light and i-Scan: Evidence-based consensus recommendations using a modified Delphi process." United European Gastroenterol J **6**(5): 748-754.
- Bi, X., A. Walsh, A. Mahadevan-Jansen and A. Herline (2011). "Development of spectral markers for the discrimination of ulcerative colitis and Crohn's disease using Raman spectroscopy." Dis Colon Rectum **54**(1): 48-53.
- Bielecki, C., T. W. Bocklitz, M. Schmitt, C. Krafft, C. Marquardt, A. Gharbi, T. Knosel, A. Stallmach and J. Popp (2012). "Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells." J Biomed Opt **17**(7): 076030.
- Bisschops, R., E. Dekker, J. E. East, G. Johnson, P. Pimentel-Nunes, D. S. Sanders, M. Dinis-Ribeiro and T. Ponchon (2019). "European Society of Gastrointestinal Endoscopy (ESGE) curricula development for postgraduate training in advanced endoscopic procedures: rationale and methodology." Endoscopy **51**(10): 976-979.
- Bisschops, R., J. E. East, C. Hassan, Y. Hazewinkel, M. F. Kaminski, H. Neumann, M. Pellise, G. Antonelli, M. Bustamante Balen, E. Coron, G. Cortas, M. Iacucci, M. Yuichi, G. Longcroft-Wheaton, S. Mouzyka, N. Pilonis, I. Puig, J. E. van Hooft and E. Dekker (2019). "Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019." Endoscopy **51**(12): 1155-1179.

- Bisschops, R., C. Hassan, P. Bhandari, E. Coron, H. Neumann, O. Pech, L. Correale and A. Repici (2018). "BASIC (BLI Adenoma Serrated International Classification) classification for colorectal polyp characterization with blue light imaging." Endoscopy **50**(3): 211-220.
- Bitton, A., M. A. Peppercorn, D. A. Antonioli, J. L. Niles, S. Shah, A. Bousvaros, B. Ransil, G. Wild, A. Cohen, M. D. Edwardes and A. C. Stevens (2001). "Clinical, biological, and histologic parameters as predictors of relapse in ulcerative colitis." Gastroenterology **120**(1): 13-20.
- Boal Carvalho, P., F. Dias de Castro, B. Rosa, M. J. Moreira and J. Cotter (2016). "Mucosal Healing in Ulcerative Colitis--When Zero is Better." J Crohns Colitis **10**(1): 20-25.
- Bocklitz, T. W., S. X. Guo, O. Ryabchykov, N. Vogler and J. Popp (2016). "Raman Based Molecular Imaging and Analytics: A Magic Bullet for Biomedical Applications!?" Analytical Chemistry **88**(1): 133-151.
- Bossuyt, P., F. Baert, F. D'Heygere, A. Nakad, C. Reenaers, F. Fontaine, D. Franchimont, O. Dewit, P. Van Hootegem, S. Vanden Branden, G. Lambrecht and M. Ferrante (2019). "Early Mucosal Healing Predicts Favorable Outcomes in Patients With Moderate to Severe Ulcerative Colitis Treated With Golimumab: Data From the Real-life BE-SMART Cohort." Inflamm Bowel Dis **25**(1): 156-162.
- Bossuyt, P., H. Nakase, S. Vermeire, G. de Hertogh, T. Eelbode, M. Ferrante, T. Hasegawa, H. Willekens, Y. Ikemoto, T. Makino and R. Bisschops (2020). "Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density." Gut.
- Bouwens, M. W., R. de Ridder, A. A. Masclee, A. Driessen, R. G. Riedl, B. Winkens and S. Sanduleanu (2013). "Optical diagnosis of colorectal polyps using high-definition i-scan: an educational experience." World J Gastroenterol **19**(27): 4334-4343.
- Brahmania, M. and C. N. Bernstein (2014). "Physician global assessments or blood tests do not predict mucosal healing in ulcerative colitis." Can J Gastroenterol Hepatol **28**(6): 325-329.
- Brereton, R. G. (2012). "Self organising maps for visualising and modelling." Chem Cent J **6 Suppl 2**: S1.
- Brethauer, M. and A. S. Brandrud (2014). "Endoscopy training: time to stop counting procedures?" Gut **63**(11): 1686-1687.
- Bryant, R. V., D. C. Burger, J. Delo, A. J. Walsh, S. Thomas, A. von Herbay, O. C. Buchel, L. White, O. Brain, S. Keshav, B. F. Warren and S. P. Travis (2016). "Beyond endoscopic mucosal healing in UC: histological remission better predicts corticosteroid use and hospitalisation over 6 years of follow-up." Gut **65**(3): 408-414.
- Bryant, R. V., S. Winer, S. P. Travis and R. H. Riddell (2014). "Systematic review: histological remission in inflammatory bowel disease. Is 'complete' remission the new treatment paradigm? An IOIBD initiative." J Crohns Colitis **8**(12): 1582-1597.
- Buda, A., G. Hatem, H. Neumann, R. D'Inca, C. Mescoli, P. Piselli, J. Jackson, M. Bruno and G. C. Sturniolo (2014). "Confocal laser endomicroscopy for prediction of disease relapse in ulcerative colitis: a pilot study." J Crohns Colitis **8**(4): 304-311.
- Butler, H. J., L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone and F. L.

Martin (2016). "Using Raman spectroscopy to characterize biological materials." Nat Protoc **11**(4): 664-687.

Cairns, S. R., J. H. Scholefield, R. J. Steele, M. G. Dunlop, H. J. Thomas, G. D. Evans, J. A. Eaden, M. D. Rutter, W. P. Atkin, B. P. Saunders, A. Lucassen, P. Jenkins, P. D. Fairclough, C. R. Woodhouse, G. British Society of, B. Association of Coloproctology for Great and Ireland (2010). "Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002)." Gut **59**(5): 666-689.

Calafat, M., T. Lobaton, A. Hernandez-Gallego, M. Manosa, P. Torres, F. Canete, E. Cabre, I. Ojanguren and E. Domenech (2017). "Acute histological inflammatory activity is associated with clinical relapse in patients with ulcerative colitis in clinical and endoscopic remission." Dig Liver Dis **49**(12): 1327-1331.

Calderwood, A. H. and B. C. Jacobson (2010). "Comprehensive validation of the Boston Bowel Preparation Scale." Gastrointest Endosc **72**(4): 686-692.

Cannatelli, R., A. Bazarova, D. Zardo, O. M. Nardone, U. Shivaji, S. C. L. Smith, G. Gkoutos, C. Ricci, X. S. Gui, S. Ghosh and M. Iacucci (2020). "Fecal Calprotectin Thresholds to Predict Endoscopic Remission Using Advanced Optical Enhancement Techniques and Histological Remission in IBD Patients." Inflammatory Bowel Diseases.

Cheriyadat, A. and L. M. Bruce (2003). "Why principal component analysis is not an appropriate feature extraction method for hyperspectral data." Igarss 2003: Ieee International Geoscience and Remote Sensing Symposium, Vols I - Vii, Proceedings: 3420-3422.

Chiu, H. M., C. Y. Chang, C. C. Chen, Y. C. Lee, M. S. Wu, J. T. Lin, C. T. Shun and H. P. Wang (2007). "A prospective comparative study of narrow-band imaging, chromoendoscopy, and conventional colonoscopy in the diagnosis of colorectal neoplasia." Gut **56**(3): 373-379.

Christensen, B., S. B. Hanauer, J. Erlich, O. Kassim, P. R. Gibson, J. R. Turner, J. Hart and D. T. Rubin (2017). "Histologic Normalization Occurs in Ulcerative Colitis and Is Associated With Improved Clinical Outcomes." Clin Gastroenterol Hepatol **15**(10): 1557-1564.e1551.

Coe, S. G., C. Thomas, J. Crook, V. Ussui, N. Diehl and M. B. Wallace (2012). "Colorectal surveillance interval assignment based on in vivo prediction of polyp histology: impact of endoscopic quality improvement program." Gastrointest Endosc **76**(1): 118-125 e111.

Colombel, J. F., G. D'Haens, W. J. Lee, J. Petersson and R. Panaccione (2019). "Outcomes and Strategies to Support a Treat-to-target Approach in Inflammatory Bowel Disease: A Systematic Review." J Crohns Colitis.

Colombel, J. F., M. E. Keir, A. Scherl, R. Zhao, G. de Hertogh, W. A. Faubion and T. T. Lu (2017). "Discrepancies between patient-reported outcomes, and endoscopic and histological appearance in UC." Gut **66**(12): 2063-2068.

Colombel, J. F., P. Rutgeerts, W. Reinisch, D. Esser, Y. Wang, Y. Lang, C. W. Marano, R. Strauss, B. J. Oddens, B. G. Feagan, S. B. Hanauer, G. R. Lichtenstein, D. Present, B. E. Sands and W. J. Sandborn (2011). "Early mucosal healing with infliximab is associated with improved long-term clinical outcomes in ulcerative colitis." Gastroenterology **141**(4): 1194-1201.

Committee, A. T. (2013). "Enhanced imaging in the GI tract: spectroscopy and optical coherence tomography." Gastrointest Endosc **78**(4): 568-573.

- Committee, A. T. (2014). "Confocal laser endomicroscopy." *Gastrointest Endosc* **80**(6): 928-938.
- Cui, L., H. J. Butler, P. L. Martin-Hirsch and F. L. Martin (2016). "Aluminium foil as a potential substrate for ATR-FTIR, transfection FTIR or Raman spectrochemical analysis of biological specimens." *Analytical methods* **8**(3): 481-487.
- D'Haens, G., W. J. Sandborn, B. G. Feagan, K. Geboes, S. B. Hanauer, E. J. Irvine, M. Lemann, P. Marteau, P. Rutgeerts, J. Scholmerich and L. R. Sutherland (2007). "A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis." *Gastroenterology* **132**(2): 763-786.
- D'Haens, G., S. Van Deventer, R. Van Hogezaand, D. Chalmers, C. Kothe, F. Baert, T. Braakman, T. Schaible, K. Geboes and P. Rutgeerts (1999). "Endoscopic and histological healing with infliximab anti-tumor necrosis factor antibodies in Crohn's disease: A European multicenter trial." *Gastroenterology* **116**(5): 1029-1034.
- Danese, S., G. Fiorino, E. Angelucci, S. Vetrano, N. Pagano, G. Rando, A. Spinelli, A. Malesci and A. Repici (2010). "Narrow-band imaging endoscopy to assess mucosal angiogenesis in inflammatory bowel disease: a pilot study." *World J Gastroenterol* **16**(19): 2396-2400.
- Daperno, M., G. D'Haens, G. Van Assche, F. Baert, P. Bulois, V. Maunoury, R. Sostegni, R. Rocca, A. Pera, A. Gevers, J. Y. Mary, J. F. Colombel and P. Rutgeerts (2004). "Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD." *Gastrointest Endosc* **60**(4): 505-512.
- Deeks, J. J., Higgins, J.P.T, Altman, D.G. (2020). "Chapter 10: Analysing data and undertaking meta-analyses." Retrieved 18.11.2020, from <https://training.cochrane.org/handbook/current/chapter-10>.
- Diedenhofen, B. and J. Musch (2015). "cocor: a comprehensive solution for the statistical comparison of correlations." *PLoS One* **10**(3): e0121945.
- Ding, H., A. W. Dupont, S. Singhal, L. D. Scott, S. Guha, M. Younes and X. Bi (2017). "In vivo analysis of mucosal lipids reveals histological disease activity in ulcerative colitis using endoscope-coupled Raman spectroscopy." *Biomed Opt Express* **8**(7): 3426-3439.
- Doherty, G., K. H. Katsanos, J. Burisch, M. Allez, K. Papamichael, A. Stallmach, R. Mao, I. P. Berset, J. P. Gisbert, S. Sebastian, J. Kierkus, L. Lopetuso, E. Szymanska and E. Louis (2018). "European Crohn's and Colitis Organisation Topical Review on Treatment Withdrawal ['Exit Strategies'] in Inflammatory Bowel Disease." *J Crohns Colitis* **12**(1): 17-31.
- Doherty, G., K. H. Katsanos, J. Burisch, M. Allez, K. Papamichael, A. Stallmach, R. Mao, I. P. Berset, J. P. Gisbert, S. Sebastian, J. Kierkuś, L. Lopetuso, E. Szymanska and E. Louis (2017). "European Crohn's and Colitis Organisation Topical Review on Treatment Withdrawal ['Exit Strategies'] in Inflammatory Bowel Disease." *Journal of Crohn's and Colitis* **12**(1): 17-31.
- Dyson, J. K. and M. D. Rutter (2012). "Colorectal cancer in inflammatory bowel disease: what is the real magnitude of the risk?" *World J Gastroenterol* **18**(29): 3839-3848.
- East, J. E., J. L. Vleugels, P. Roelandt, P. Bhandari, R. Bisschops, E. Dekker, C. Hassan, G. Horgan, R. Kiesslich, G. Longcroft-Wheaton, A. Wilson and J. M. Dumonceau (2016).

"Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) Technology Review." Endoscopy **48**(11): 1029-1045.

Ekbom, A., C. Helmick, M. Zack and H. O. Adami (1990). "Increased risk of large-bowel cancer in Crohn's disease with colonic involvement." Lancet **336**(8711): 357-359.

Ekbom, A., C. Helmick, M. Zack and H. O. Adami (1990). "Ulcerative colitis and colorectal cancer. A population-based study." N Engl J Med **323**(18): 1228-1233.

Endoscopic Classification Review, G. (2005). "Update on the paris classification of superficial neoplastic lesions in the digestive tract." Endoscopy **37**(6): 570-578.

Falvey, J. D., T. Hoskin, B. Meijer, A. Ashcroft, R. Walmsley, A. S. Day and R. B. Gearry (2015). "Disease activity assessment in IBD: clinical indices and biomarkers fail to predict endoscopic remission." Inflamm Bowel Dis **21**(4): 824-831.

Feagan, B. G., W. J. Sandborn, G. D'Haens, S. Pola, J. W. D. McDonald, P. Rutgeerts, P. Munkholm, U. Mittmann, D. King, C. J. Wong, G. Zou, A. Donner, L. M. Shackelton, D. Gilgen, S. Nelson, M. K. Vandervoort, M. Fahmy, E. V. Loftus, Jr., R. Panaccione, S. P. Travis, G. A. Van Assche, S. Vermeire and B. G. Levesque (2013). "The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis." Gastroenterology **145**(1): 149-157 e142.

Fernandes, S. R., J. Pinto, P. Marques da Costa, L. Correia and Gedii (2018). "Disagreement Among Gastroenterologists Using the Mayo and Rutgeerts Endoscopic Scores." Inflamm Bowel Dis **24**(2): 254-260.

Ferrante, M., S. Vermeire, H. Fidder, F. Schnitzler, M. Noman, G. Van Assche, G. De Hertogh, I. Hoffman, A. D'Hoore, K. Van Steen, K. Geboes, F. Penninckx and P. Rutgeerts (2008). "Long-term outcome after infliximab for refractory ulcerative colitis." J Crohns Colitis **2**(3): 219-225.

Flores, B. M., A. O'Connor and A. C. Moss (2017). "Impact of mucosal inflammation on risk of colorectal neoplasia in patients with ulcerative colitis: a systematic review and meta-analysis." Gastrointest Endosc **86**(6): 1006-1011.e1008.

Frosli, K. F., J. Jahnsen, B. A. Moum, M. H. Vatn and I. Group (2007). "Mucosal healing in inflammatory bowel disease: results from a Norwegian population-based cohort." Gastroenterology **133**(2): 412-422.

Fukuda, H., Y. Miura, Y. Hayashi, T. Takezawa, Y. Ino, M. Okada, H. Osawa, A. K. Lefor and H. Yamamoto (2015). "Linked color imaging technology facilitates early detection of flat gastric cancers." Clin J Gastroenterol **8**(6): 385-389.

Fumery, M., S. Singh, P. S. Dulai, C. Gower-Rousseau, L. Peyrin-Biroulet and W. J. Sandborn (2018). "Natural History of Adult Ulcerative Colitis in Population-based Cohorts: A Systematic Review." Clin Gastroenterol Hepatol **16**(3): 343-356.e343.

Geboes, K., R. Riddell, A. Ost, B. Jensfelt, T. Persson and R. Lofberg (2000). "A reproducible grading scale for histological assessment of inflammation in ulcerative colitis." Gut **47**(3): 404-409.

Gisbert, J. P., A. C. Marin and M. Chaparro (2016). "The Risk of Relapse after Anti-TNF Discontinuation in Inflammatory Bowel Disease: Systematic Review and Meta-Analysis." Am J Gastroenterol **111**(5): 632-647.

Gono, K., K. Yamazaki, N. Doguchi, T. Nonami, T. Obi, M. Yamaguchi, N. Ohyama, H. Machida, Y. Sano, S. Yoshida, Y. Hamamoto and T. Endo (2003). "Endoscopic observation of tissue by narrowband illumination." Optical Review **10**(4): 211-215.

Graff, L. A., N. Vincent, J. R. Walker, I. Clara, R. Carr, J. Ediger, N. Miller, L. Rogala, P. Rawsthorne, L. Lix and C. N. Bernstein (2011). "A population-based study of fatigue and sleep difficulties in inflammatory bowel disease." Inflamm Bowel Dis **17**(9): 1882-1889.

Gralnek, I. M., C. Hassan, U. Beilenhoff, G. Antonelli, A. Ebigbo, M. Pellise, M. Arvanitakis, P. Bhandari, R. Bisschops, J. E. Van Hooft, M. F. Kaminski, K. Triantafyllou, G. Webster, H. Pohl, I. Dunkley, B. Fehrke, M. Gazic, T. Gjergjek, S. Maasen, W. Waagenes, M. de Pater, T. Ponchon, P. D. Siersema, H. Messmann and M. Dinis-Ribeiro (2020). "ESGE and ESGENA Position Statement on gastrointestinal endoscopy and the COVID-19 pandemic." Endoscopy.

Gralnek, I. M., C. Hassan and M. Dinis-Ribeiro (2020). "COVID-19 and endoscopy: implications for healthcare and digestive cancer screening." Nat Rev Gastroenterol Hepatol: 1-3.

Hassan, C., M. Spadaccini, A. Iannone, R. Maselli, M. Jovani, V. T. Chandrasekar, G. Antonelli, H. Yu, M. Areia, M. Dinis-Ribeiro, P. Bhandari, P. Sharma, D. K. Rex, T. Rösch, M. Wallace and A. Repici (2021). "Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis." Gastrointest Endosc **93**(1): 77-85.e76.

Hassan, C., M. B. Wallace, P. Sharma, R. Maselli, V. Craviotto, M. Spadaccini and A. Repici (2019). "New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection." Gut.

Hazewinkel, Y., M. Lopez-Ceron, J. E. East, A. Rastogi, M. Pellise, T. Nakajima, S. van Eeden, K. M. Tytgat, P. Fockens and E. Dekker (2013). "Endoscopic features of sessile serrated adenomas: validation by international experts using high-resolution white-light endoscopy and narrow-band imaging." Gastrointest Endosc **77**(6): 916-924.

Herrinton, L. J., L. Liu, T. R. Levin, J. E. Allison, J. D. Lewis and F. Velayos (2012). "Incidence and mortality of colorectal adenocarcinoma in persons with inflammatory bowel disease from 1998 to 2010." Gastroenterology **143**(2): 382-389.

Hewett, D. G., T. Kaltenbach, Y. Sano, S. Tanaka, B. P. Saunders, T. Ponchon, R. Soetikno and D. K. Rex (2012). "Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging." Gastroenterology **143**(3): 599-607 e591.

Higashi, R., T. Uraoka, J. Kato, K. Kuwaki, S. Ishikawa, Y. Saito, T. Matsuda, H. Ikematsu, Y. Sano, S. Suzuki, Y. Murakami and K. Yamamoto (2010). "Diagnostic accuracy of narrow-band imaging and pit pattern analysis significantly improved for less-experienced endoscopists after an expanded training program." Gastrointest Endosc **72**(1): 127-135.

Higgins, J. P., D. G. Altman, P. C. Gotzsche, P. Juni, D. Moher, A. D. Oxman, J. Savovic, K. F. Schulz, L. Weeks, J. A. Sterne, G. Cochrane Bias Methods and G. Cochrane Statistical Methods (2011). "The Cochrane Collaboration's tool for assessing risk of bias in randomised trials." BMJ **343**: d5928.

- Hoivik, M. L., B. Moum, I. C. Solberg, M. Henriksen, M. Cvancarova and T. Bernklev (2013). "Work disability in inflammatory bowel disease patients 10 years after disease onset: results from the IBSEN Study." Gut **62**(3): 368-375.
- Iacucci, M., R. Cannatelli, X. S. Gui, D. Zardo, A. Bazarova, G. Gkoutos, B. C. Lethebe, G. G. Kaplan, R. Panaccione, R. Kiesslich and S. Ghosh (2020). "Assessment of endoscopic healing by using advanced technologies reflects histologic healing in ulcerative colitis." J Crohns Colitis.
- Iacucci, M., R. Cannatelli, G. E. Tontini, R. Panaccione, S. Danese, G. Fiorino, T. Matsumoto, G. S. Kochhar, B. Shen, R. Kiesslich and S. Ghosh (2019). "Improving the quality of surveillance colonoscopy in inflammatory bowel disease." Lancet Gastroenterol Hepatol **4**(12): 971-983.
- Iacucci, M., M. Daperno, M. Lazarev, R. Arsenescu, G. E. Tontini, O. Akinola, X. S. Gui, V. Villanacci, M. Goetz, M. Lowerison, B. C. Lethebe, M. Vecchi, H. Neumann, S. Ghosh, R. Bisschops and R. Kiesslich (2017). "Development and reliability of the new endoscopic virtual chromoendoscopy score: the PICaSSO (Paddington International Virtual ChromoendoScopy ScOre) in ulcerative colitis." Gastrointest Endosc **86**(6): 1118-1127 e1115.
- Iacucci, M., M. Fort Gasia, C. Hassan, R. Panaccione, G. G. Kaplan, S. Ghosh and X. Gui (2015). "Complete mucosal healing defined by endoscopic Mayo subscore still demonstrates abnormalities by novel high definition colonoscopy and refined histological gradings." Endoscopy **47**(8): 726-734.
- Iacucci, M., F. Furfaro, T. Matsumoto, T. Uraoka, S. Smith, S. Ghosh and R. Kiesslich (2018). "Advanced endoscopic techniques in the assessment of inflammatory bowel disease: new technology, new era." Gut.
- Iacucci, M., F. Furfaro, T. Matsumoto, T. Uraoka, S. Smith, S. Ghosh and R. Kiesslich (2019). "Advanced endoscopic techniques in the assessment of inflammatory bowel disease: New technology, new era." Gut **68**(3): 562-572.
- Iacucci, M., L. Jeffery, A. Acharjee, O. M. Nardone, D. Zardo, S. C. L. Smith, A. Bazarova, R. Cannatelli, U. N. Shivaji, J. Williams, G. Gkoutos and S. Ghosh (2021). "Ultra-high Magnification Endocytoscopy and Molecular Markers for Defining Endoscopic and Histologic Remission in Ulcerative Colitis-An Exploratory Study to Define Deep Remission." Inflamm Bowel Dis.
- Iacucci, M., R. Kiesslich, X. Gui, R. Panaccione, J. Heatherington, O. Akinola and S. Ghosh (2017). "Beyond white light: optical enhancement in conjunction with magnification colonoscopy for the assessment of mucosal healing in ulcerative colitis." Endoscopy **49**(6): 553-559.
- Iacucci, M., K. McQuaid, X. S. Gui, Y. Iwao, B. C. Lethebe, M. Lowerison, T. Matsumoto, U. N. Shivaji, S. C. L. Smith, V. Subramanian, T. Uraoka, S. Sanduleanu, S. Ghosh and R. Kiesslich (2018). "A multimodal (FACILE) classification for optical diagnosis of inflammatory bowel disease associated neoplasia." Endoscopy.
- Iacucci, M., K. McQuaid, X. S. Gui, Y. Iwao, B. C. Lethebe, M. Lowerison, T. Matsumoto, U. N. Shivaji, S. C. L. Smith, V. Subramanian, T. Uraoka, S. Sanduleanu, S. Ghosh and R. Kiesslich (2019). "A multimodal (FACILE) classification for optical diagnosis of inflammatory bowel disease associated neoplasia." Endoscopy **51**(2): 133-141.

Iacucci, M., S. C. L. Smith, A. Bazarova, U. N. Shivaji, P. Bhandari, R. Cannatelli, M. Daperno, J. Ferraz, M. Goetz, X. Gui, B. Hayee, G. De Hertogh, M. Lazarev, J. Li, O. M. Nardone, A. Parra-Blanco, L. Pastorelli, R. Panaccione, V. Occhipinti, T. Rath, G. E. Tontini, M. Vieth, V. Villanacci, D. Zardo, R. Bisschops, R. Kiesslich and S. Ghosh (2021). "An International Multicenter Real-Life Prospective Study of Electronic Chromoendoscopy Score PICaSSO in Ulcerative Colitis." Gastroenterology.

Iacucci, M., C. Trovato, M. Daperno, O. Akinola, D. Greenwald, S. A. Gross, A. Hoffman, J. Lee, B. C. Lethebe, M. Lowerison, J. Naylor, H. Neumann, T. Rath, S. Sanduleanu, P. Sharma, R. Kiesslich, S. Ghosh, J. R. Saltzman and S. c. i. t. and the (2018). "Development and validation of the SIMPLE endoscopic classification of diminutive and small colorectal polyps." Endoscopy **50**(8): 779-789.

Ignjatovic, A., S. Thomas-Gibson, J. E. East, A. Haycock, P. Bassett, P. Bhandari, R. Man, N. Suzuki and B. P. Saunders (2011). "Development and validation of a training module on the use of narrow-band imaging in differentiation of small adenomas from hyperplastic colorectal polyps." Gastrointest Endosc **73**(1): 128-133.

Irani, N. R., L. M. Wang, G. S. Collins, S. Keshav and S. P. L. Travis (2018). "Correlation Between Endoscopic and Histological Activity in Ulcerative Colitis Using Validated Indices." J Crohns Colitis **12**(10): 1151-1157.

JE, I. J., B. A. Bastiaansen, M. E. van Leerdam, G. A. Meijer, S. van Eeden, S. Sanduleanu, E. J. Schoon, T. M. Bisseling, M. C. Spaander, N. van Lelyveld, M. Bargeman, J. Wang, E. Dekker, S. Dutch Workgroup serrated polyp and Polyposis (2016). "Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps." Gut **65**(6): 963-970.

Jharap, B., W. J. Sandborn, W. Reinisch, G. D'Haens, A. M. Robinson, W. Wang, B. Huang, A. Lazar, R. B. Thakkar and J. F. Colombel (2015). "Randomised clinical study: discrepancies between patient-reported outcomes and endoscopic appearance in moderate to severe ulcerative colitis." Aliment Pharmacol Ther **42**(9): 1082-1092.

Kaltenbach, T. R., R. M. Soetikno, R. DeVivo, L. A. Laine, A. Barkun, K. R. McQuaid and I. B. D. G. Optimizing Quality of Endoscopy in (2017). "Optimizing the quality of endoscopy in inflammatory bowel disease: focus on surveillance and management of colorectal dysplasia using interactive image- and video-based teaching." Gastrointest Endosc **86**(6): 1107-1117 e1101.

Kanazawa, M., F. Takahashi, K. Tominaga, K. Abe, N. Izawa, K. Fukushi, K. Nagashima, A. Kanamori, K. Takenaka, T. Sugaya, M. Iijima, A. Takada, Y. Imai, H. Hiraishi and A. Irisawa (2019). "Relationship between endoscopic mucosal healing and histologic inflammation during remission maintenance phase in ulcerative colitis: a retrospective study." Endosc Int Open **7**(4): E568-e575.

Kawasaki, K., S. Nakamura, M. Esaki, K. Kurahara, M. Eizuka, Y. Nuki, S. Kochi, M. Fujiwara, Y. Oshiro, T. Sugai and T. Matsumoto (2019). "Clinical usefulness of magnifying colonoscopy for the diagnosis of ulcerative colitis-associated neoplasia." Dig Endosc **31 Suppl 1**: 36-42.

Kendall, C., N. Stone, N. Shepherd, K. Geboes, B. Warren, R. Bennett and H. Barr (2003). "Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus." J Pathol **200**(5): 602-609.

- Keswani, R. N., A. Sethi, A. Repici, H. Messmann and P. W. Chiu (2020). "How To Maximize Trainee Education During the Coronavirus Disease-2019 Pandemic: Perspectives From Around the World." Gastroenterology **159**(1): 26-29.
- Kevans, D., R. Kirsch, C. Dargavel, B. Kabakchiev, R. Riddell and M. S. Silverberg (2019). "Histological Markers of Clinical Relapse in Endoscopically Quiescent Ulcerative Colitis." Inflamm Bowel Dis.
- Khan, T., B. Cinnor, N. Gupta, L. Hosford, A. Bansal, M. S. Olyae, S. Wani and A. Rastogi (2017). "Didactic training vs. computer-based self-learning in the prediction of diminutive colon polyp histology by trainees: a randomized controlled study." Endoscopy **49**(12): 1243-1250.
- Khanna, R., G. Zou, L. Stitt, B. G. Feagan, W. J. Sandborn, P. Rutgeerts, J. W. D. McDonald, E. Dubcenco, R. Fogel, R. Panaccione, V. Jairath, S. Nelson, L. M. Shackelton, B. Huang, Q. Zhou, A. M. Robinson, B. G. Levesque and G. D'Haens (2017). "Responsiveness of Endoscopic Indices of Disease Activity for Crohn's Disease." Am J Gastroenterol **112**(10): 1584-1592.
- Kiesslich, R., J. Fritsch, M. Holtmann, H. H. Koehler, M. Stolte, S. Kanzler, B. Nafe, M. Jung, P. R. Galle and M. F. Neurath (2003). "Methylene blue-aided chromoendoscopy for the detection of intraepithelial neoplasia and colon cancer in ulcerative colitis." Gastroenterology **124**(4): 880-888.
- Kim, J. H., J. H. Cheon, Y. Park, H. J. Lee, S. J. Park, T. I. Kim and W. H. Kim (2016). "Effect of mucosal healing (Mayo 0) on clinical relapse in patients with ulcerative colitis in clinical remission." Scand J Gastroenterol **51**(9): 1069-1074.
- Kissler, S. M., C. Tedijanto, E. Goldstein, Y. H. Grad and M. Lipsitch (2020). "Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period." Science.
- Kodashima, S. and M. Fujishiro (2010). "Novel image-enhanced endoscopy with i-scan technology." World J Gastroenterol **16**(9): 1043-1049.
- Kohonen, T. (1982). "Self-Organized Formation of Topologically Correct Feature Maps." Biological Cybernetics **43**(1): 59-69.
- Kominami, Y., S. Yoshida, S. Tanaka, Y. Sanomura, T. Hirakawa, B. Raytchev, T. Tamaki, T. Koide, K. Kaneda and K. Chayama (2016). "Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy." Gastrointest Endosc **83**(3): 643-649.
- Krafft, C., T. Knetschke, A. Siegner, R. H. W. Funk and R. Salzer (2003). "Mapping of single cells by near infrared Raman microspectroscopy." Vibrational Spectroscopy **32**(1): 75-83.
- Kudo, S., S. Hirota, T. Nakajima, S. Hosobe, H. Kusaka, T. Kobayashi, M. Himori and A. Yagyuu (1994). "Colorectal tumours and pit pattern." J Clin Pathol **47**(10): 880-885.
- Kudo, S., C. A. Rubio, C. R. Teixeira, H. Kashida and E. Kogure (2001). "Pit pattern in colorectal neoplasia: endoscopic magnifying view." Endoscopy **33**(4): 367-373.
- Kudo, T., T. Matsumoto, M. Esaki, T. Yao and M. Iida (2009). "Mucosal vascular pattern in ulcerative colitis: observations using narrow band imaging colonoscopy with special reference to histologic inflammation." Int J Colorectal Dis **24**(5): 495-501.

- Ladabaum, U., A. Fioritto, A. Mitani, M. Desai, J. P. Kim, D. K. Rex, T. Imperiale and N. Gunaratnam (2013). "Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions." Gastroenterology **144**(1): 81-91.
- Laharie, D., J. Filippi, X. Roblin, S. Nancey, J. B. Chevaux, X. Hebuterne, B. Flourie, M. Capdepon and L. Peyrin-Biroulet (2013). "Impact of mucosal healing on long-term outcomes in ulcerative colitis treated with infliximab: a multicenter experience." Aliment Pharmacol Ther **37**(10): 998-1004.
- Laine, L., T. Kaltenbach, A. Barkun, K. R. McQuaid, V. Subramanian, R. Soetikno and S. G. D. Panel (2015). "SCENIC international consensus statement on surveillance and management of dysplasia in inflammatory bowel disease." Gastrointest Endosc **81**(3): 489-501 e426.
- Lee, T. J., M. D. Rutter, R. G. Blanks, S. M. Moss, A. F. Goddard, A. Chilton, C. Nickerson, R. J. McNally, J. Patnick and C. J. Rees (2012). "Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme." Gut **61**(7): 1050-1057.
- Lemmens, B., I. Arijs, G. Van Assche, X. Sagaert, K. Geboes, M. Ferrante, P. Rutgeerts, S. Vermeire and G. De Hertogh (2013). "Correlation between the endoscopic and histologic score in assessing the activity of ulcerative colitis." Inflamm Bowel Dis **19**(6): 1194-1201.
- Levesque, B. G., W. J. Sandborn, J. Ruel, B. G. Feagan, B. E. Sands and J. F. Colombel (2015). "Converging goals of treatment of inflammatory bowel disease from clinical trials and practice." Gastroenterology **148**(1): 37-51 e31.
- Lieberman, D., M. Moravec, J. Holub, L. Michaels and G. Eisen (2008). "Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography." Gastroenterology **135**(4): 1100-1105.
- Limdi, J. K., M. Picco and F. A. Farraye (2020). "A review of endoscopic scoring systems and their importance in a treat-to-target approach in inflammatory bowel disease (with videos)." Gastrointest Endosc **91**(4): 733-745.
- Lui, T. K., C. G. Guo and W. K. Leung (2020). "ACCURACY OF ARTIFICIAL INTELLIGENCE ON HISTOLOGY PREDICTION AND DETECTION OF COLORECTAL POLYPS: A SYSTEMATIC REVIEW AND META-ANALYSIS." Gastrointest Endosc.
- Lutgens, M. W., M. G. van Oijen, G. J. van der Heijden, F. P. Vleggaar, P. D. Siersema and B. Oldenburg (2013). "Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies." Inflamm Bowel Dis **19**(4): 789-799.
- Magro, F., P. Gionchetti, R. Eliakim, S. Ardizzone, A. Armuzzi, M. Barreiro-de Acosta, J. Burisch, K. B. Gecse, A. L. Hart, P. Hindryckx, C. Langner, J. K. Limdi, G. Pellino, E. Zagorowicz, T. Raine, M. Harbord and F. Rieder (2017). "Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders." J Crohns Colitis **11**(6): 649-670.
- Magro, F., P. Gionchetti, R. Eliakim, S. Ardizzone, A. Armuzzi, M. Barreiro-de Acosta, J. Burisch, K. B. Gecse, A. L. Hart, P. Hindryckx, C. Langner, J. K. Limdi, G. Pellino, E. Zagorowicz, T. Raine, M. Harbord, F. Rieder, C. s. European and O. Colitis (2017). "Third

European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders." *J Crohns Colitis* **11**(6): 649-670.

Malik, K. I., K. Siau, P. Dunckley and S. T. Ward (2019). "Colorectal trainees in the UK struggle to meet JAG certification standards in colonoscopy by the end of their training." *Colorectal Dis* **21**(6): 715-722.

Marchal Bressenot, A., R. H. Riddell, C. Boulagnon-Rombi, W. Reinisch, S. Danese, S. Schreiber and L. Peyrin-Biroulet (2015). "Review article: the histological assessment of disease activity in ulcerative colitis." *Aliment Pharmacol Ther* **42**(8): 957-967.

Marchal-Bressenot, A., J. Salleron, C. Boulagnon-Rombi, C. Bastien, V. Cahn, G. Cadiot, M. D. Diebold, S. Danese, W. Reinisch, S. Schreiber, S. Travis and L. Peyrin-Biroulet (2017). "Development and validation of the Nancy histological index for UC." *Gut* **66**(1): 43-49.

Matsumoto, T., F. Kuroki, M. Mizuno, S. Nakamura and M. Iida (1997). "Application of magnifying chromoscopy for the assessment of severity in patients with mild to moderate ulcerative colitis." *Gastrointest Endosc* **46**(5): 400-405.

Matts, S. G. (1961). "The value of rectal biopsy in the diagnosis of ulcerative colitis." *Q J Med* **30**: 393-407.

McGill, S. K., E. Evangelou, J. P. Ioannidis, R. M. Soetikno and T. Kaltenbach (2013). "Narrow band imaging to differentiate neoplastic and non-neoplastic colorectal polyps in real time: a meta-analysis of diagnostic operating characteristics." *Gut* **62**(12): 1704-1713.

Megna, B., J. Weiss, D. Ley, S. Saha, P. Pfau, I. Grimes, Z. Li and F. Caldera (2018). "Clear liquid diet before bowel preparation predicts successful chromoendoscopy in patients with inflammatory bowel disease." *Gastrointest Endosc*.

Mesejo, P., D. Pizarro, A. Abergel, O. Rouquette, S. Beorchia, L. Poincloux and A. Bartoli (2016). "Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy." *IEEE Trans Med Imaging* **35**(9): 2051-2063.

Meucci, G., R. Fasoli, S. Saibeni, D. Valpiani, R. Gullotta, E. Colombo, R. D'Inca, M. Terpin and G. Lombardi (2012). "Prognostic significance of endoscopic remission in patients with active ulcerative colitis treated with oral and topical mesalazine: a prospective, multicenter study." *Inflamm Bowel Dis* **18**(6): 1006-1010.

Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman and P. Group (2009). "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *PLoS Med* **6**(7): e1000097.

Morasso, C., M. Truffi, R. Vanna, S. Albasini, S. Mazzucchelli, F. Colombo, L. Sorrentino, G. Sampietro, S. Ardizzone and F. Corsi (2020). "Raman analysis reveals biochemical differences in plasma of Crohn's Disease patients." *J Crohns Colitis*.

Mori, Y., S. E. Kudo, P. W. Chiu, R. Singh, M. Misawa, K. Wakamura, T. Kudo, T. Hayashi, A. Katagiri, H. Miyachi, F. Ishida, Y. Maeda, H. Inoue, Y. Nimura, M. Oda and K. Mori (2016). "Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study." *Endoscopy* **48**(12): 1110-1118.

Mori, Y., S. E. Kudo, K. Wakamura, M. Misawa, Y. Ogawa, M. Kutsukawa, T. Kudo, T. Hayashi, H. Miyachi, F. Ishida and H. Inoue (2015). "Novel computer-aided diagnostic

system for colorectal lesions by using endocytoscopy (with videos)." Gastrointest Endosc **81**(3): 621-629.

Mosli, M. H., B. G. Feagan, G. Zou, W. J. Sandborn, G. D'Haens, R. Khanna, C. Behling, K. Kaplan, D. K. Driman, L. M. Shackelton, K. A. Baker, J. K. MacDonald, M. K. Vandervoort, M. A. Samaan, K. Geboes, M. A. Valasek, R. Pai, C. Langner, R. Riddell, N. Harpaz, M. Sewitch, M. Peterson, L. W. Stitt and B. G. Levesque (2015). "Reproducibility of histological assessments of disease activity in UC." Gut **64**(11): 1765-1773.

Mosli, M. H., B. G. Feagan, G. Zou, W. J. Sandborn, G. D'Haens, R. Khanna, L. M. Shackelton, C. W. Walker, S. Nelson, M. K. Vandervoort, V. Frisbie, M. A. Samaan, V. Jairath, D. K. Driman, K. Geboes, M. A. Valasek, R. K. Pai, G. Y. Lauwers, R. Riddell, L. W. Stitt and B. G. Levesque (2017). "Development and validation of a histological index for UC." Gut **66**(1): 50-58.

Murakami, T., N. Sakamoto and A. Nagahara (2019). "Clinicopathological features, diagnosis, and treatment of sessile serrated adenoma/polyp with dysplasia/carcinoma." J Gastroenterol Hepatol **34**(10): 1685-1695.

Murr, C., T. B. Grammer, A. Meinitzer, M. E. Kleber, W. Marz and D. Fuchs (2014). "Immune activation and inflammation in patients with cardiovascular disease are associated with higher phenylalanine to tyrosine ratios: the ludwigshafen risk and cardiovascular health study." J Amino Acids **2014**: 783730.

Naganuma, M., F. Hirai, K. Kobayashi, K. Watanabe, K. Takeuchi, N. Aoyama, H. Nozawa, S. Motoya, T. Ohmori, A. Harada, Y. Nagai, T. Abe, Y. Yamada, K. Inagaki, N. Shimizu, T. Kanai and M. Watanabe (2019). "Middle-term prognosis in patients with ulcerative colitis who achieved clinical and endoscopic remission by budesonide rectal foam." PLoS One **14**(8): e0220413.

Narang, V., R. Kaur, B. Garg, R. Mahajan, V. Midha, N. Sood and A. Sood (2018). "Association of endoscopic and histological remission with clinical course in patients of ulcerative colitis." Intest Res **16**(1): 55-61.

Neumann, H., F. S. Fuchs, M. Vieth, R. Atreya, J. Siebler, R. Kiesslich and M. F. Neurath (2011). "Review article: in vivo imaging by endocytoscopy." Aliment Pharmacol Ther **33**(11): 1183-1193.

Neumann, H., M. Vieth, R. Atreya, M. Grauer, J. Siebler, T. Bernatik, M. F. Neurath and J. Mudter (2012). "Assessment of Crohn's disease activity by confocal laser endomicroscopy." Inflamm Bowel Dis **18**(12): 2261-2269.

Neumann, H., M. Vieth, L. C. Fry, C. Gunther, R. Atreya, M. F. Neurath and K. Monkemuller (2013). "Learning curve of virtual chromoendoscopy for the prediction of hyperplastic and adenomatous colorectal lesions: a prospective 2-center study." Gastrointest Endosc **78**(1): 115-120.

Neumann, H., M. Vieth, C. Gunther, C. Neufert, R. Kiesslich, M. Grauer, R. Atreya and M. F. Neurath (2013). "Virtual chromoendoscopy for prediction of severity and disease extent in patients with inflammatory bowel disease: a randomized controlled study." Inflamm Bowel Dis **19**(9): 1935-1942.

- Neumann, H., M. Vieth, M. F. Neurath and R. Atreya (2013). "Endocytoscopy allows accurate in vivo differentiation of mucosal inflammatory cells in IBD: a pilot study." Inflamm Bowel Dis **19**(2): 356-362.
- Ng, S. C., H. Y. Shi, N. Hamidi, F. E. Underwood, W. Tang, E. I. Benchimol, R. Panaccione, S. Ghosh, J. C. Y. Wu, F. K. L. Chan, J. J. Y. Sung and G. G. Kaplan (2018). "Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies." Lancet **390**(10114): 2769-2778.
- Niewiadomski, O., C. Studd, C. Hair, J. Wilson, N. S. Ding, N. Heerasing, A. Ting, J. McNeill, R. Knight, J. Santamaria, E. Prewett, P. Dabkowski, D. Dowling, S. Alexander, B. Allen, B. Popp, W. Connell, P. Desmond and S. Bell (2015). "Prospective population-based cohort of inflammatory bowel disease in the biologics era: Disease course and predictors of severity." J Gastroenterol Hepatol **30**(9): 1346-1353.
- Oka, S., T. Uraoka, K. Watanabe, K. Hata, K. Kawasaki, K. Mizuno, M. Misawa, N. Hosoe, T. Moriyama and H. Kawachi (2019). "Endoscopic diagnosis and treatment of ulcerative colitis-associated neoplasia." Dig Endosc **31 Suppl 1**: 26-30.
- Olaion, G., R. Sjodahl and C. Tagesson (1990). "Glucocorticoid treatment in ileal Crohn's disease: relief of symptoms but not of endoscopically viewed inflammation." Gut **31**(3): 325-328.
- Omberg, K. M., J. C. Osborn, S. L. L. Zhang, J. P. Freyer, J. R. Maurant and J. R. Schoonover (2002). "Raman spectroscopy and factor analysis of tumorigenic and non-tumorigenic cells." Applied Spectroscopy **56**(7): 813-819.
- Ozawa, T., S. Ishihara, M. Fujishiro, H. Saito, Y. Kumagai, S. Shichijo, K. Aoyama and T. Tada (2019). "Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis." Gastrointest Endosc **89**(2): 416-421 e411.
- Pai, R. K., R. Khanna, G. R. D'Haens, W. J. Sandborn, J. Jeyarajah, B. G. Feagan, N. Vande Castele and V. Jairath (2019). "Definitions of response and remission for the Robarts Histopathology Index." Gut **68**(11): 2101-2102.
- Park, S., T. Abdi, M. Gentry and L. Laine (2016). "Histological Disease Activity as a Predictor of Clinical Relapse Among Patients With Ulcerative Colitis: Systematic Review and Meta-Analysis." Am J Gastroenterol **111**(12): 1692-1701.
- Patel, S. G., A. Rastogi, G. Austin, M. Hall, B. A. Siller, K. Berman, R. Yen, A. Bansal, D. J. Ahnen and S. Wani (2013). "Gastroenterology trainees can easily learn histologic characterization of diminutive colorectal polyps with narrow band imaging." Clin Gastroenterol Hepatol **11**(8): 997-1003 e1001.
- Pence, I. and A. Mahadevan-Jansen (2016). "Clinical instrumentation and applications of Raman spectroscopy." Chem Soc Rev **45**(7): 1958-1979.
- Pence, I. J., D. B. Beaulieu, S. N. Horst, X. Bi, A. J. Herline, D. A. Schwartz and A. Mahadevan-Jansen (2017). "Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy." Biomed Opt Express **8**(2): 524-535.
- Peyrin-Biroulet, L., W. Sandborn, B. E. Sands, W. Reinisch, W. Bemelman, R. V. Bryant, G. D'Haens, I. Dotan, M. Dubinsky, B. Feagan, G. Fiorino, R. Gearry, S. Krishnareddy, P. L. Lakatos, E. V. Loftus, Jr., P. Marteau, P. Munkholm, T. B. Murdoch, I. Ordas, R. Panaccione, R. H. Riddell, J. Ruel, D. T. Rubin, M. Samaan, C. A. Siegel, M. S. Silverberg, J. Stoker, S.

- Schreiber, S. Travis, G. Van Assche, S. Danese, J. Panes, G. Bouguen, S. O'Donnell, B. Pariente, S. Winer, S. Hanauer and J. F. Colombel (2015). "Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): Determining Therapeutic Goals for Treat-to-Target." Am J Gastroenterol **110**(9): 1324-1338.
- Picco, M. F., S. Pasha, J. A. Leighton, D. Bruining, E. V. Loftus, Jr., C. S. Thomas, J. E. Crook, M. Krishna and M. Wallace (2013). "Procedure time and the determination of polypoid abnormalities with experience: implementation of a chromoendoscopy program for surveillance colonoscopy for ulcerative colitis." Inflamm Bowel Dis **19**(9): 1913-1920.
- Ponte, A., R. Pinho, S. Fernandes, A. Rodrigues, L. Alberto, J. C. Silva, J. Silva, J. Rodrigues, M. Sousa, A. P. Silva, L. Proenca, T. Freitas, S. Leite and J. Carvalho (2017). "Impact of Histological and Endoscopic Remissions on Clinical Recurrence and Recurrence-free Time in Ulcerative Colitis." Inflamm Bowel Dis **23**(12): 2238-2244.
- Raghavendra, M., D. G. Hewett and D. K. Rex (2010). "Differentiating adenomas from hyperplastic colorectal polyps: narrow-band imaging can be learned in 20 minutes." Gastrointest Endosc **72**(3): 572-576.
- Raman, C. V. (1928). "A change of wave-length in light scattering." Nature **121**: 619-619.
- Rasmussen, D. N., J. G. Karstensen, L. B. Riis, J. Brynskov and P. Vilmann (2015). "Confocal Laser Endomicroscopy in Inflammatory Bowel Disease--A Systematic Review." J Crohns Colitis **9**(12): 1152-1159.
- Rastogi, A., D. S. Rao, N. Gupta, S. W. Grisolan, D. C. Buckles, E. Sidorenko, J. Bonino, T. Matsuda, E. Dekker, T. Kaltenbach, R. Singh, S. Wani, P. Sharma, M. S. Olyae, A. Bansal and J. E. East (2014). "Impact of a computer-based teaching module on characterization of diminutive colon polyps by using narrow-band imaging by non-experts in academic and community practice: a video-based study." Gastrointest Endosc **79**(3): 390-398.
- Rath, T., G. E. Tontini, M. F. Neurath and H. Neumann (2015). "From the surface to the single cell: Novel endoscopic approaches in inflammatory bowel disease." World J Gastroenterol **21**(40): 11260-11272.
- Rees, C. J., P. T. Rajasekhar, A. Wilson, H. Close, M. D. Rutter, B. P. Saunders, J. E. East, R. Maier, M. Moorghen, U. Muhammad, H. Hancock, A. Jayaprakash, C. MacDonald, A. Ramadas, A. Dhar and J. M. Mason (2017). "Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study." Gut **66**(5): 887-895.
- Rembacken, B., C. Hassan, J. F. Riemann, A. Chilton, M. Rutter, J. M. Dumonceau, M. Omar and T. Ponchon (2012). "Quality in screening colonoscopy: position statement of the European Society of Gastrointestinal Endoscopy (ESGE)." Endoscopy **44**(10): 957-968.
- Rembacken, B. J., T. Fujii, A. Cairns, M. F. Dixon, S. Yoshida, D. M. Chalmers and A. T. Axon (2000). "Flat and depressed colonic neoplasms: a prospective study of 1000 colonoscopies in the UK." Lancet **355**(9211): 1211-1214.
- Rex, D. K. (2009). "Narrow-band imaging without optical magnification for histologic analysis of colorectal polyps." Gastroenterology **136**(4): 1174-1181.
- Rex, D. K., C. Kahi, M. O'Brien, T. R. Levin, H. Pohl, A. Rastogi, L. Burgart, T. Imperiale, U. Ladabaum, J. Cohen and D. A. Lieberman (2011). "The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic

Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps." Gastrointest Endosc **73**(3): 419-422.

Rex, D. K., A. J. Overhiser, S. C. Chen, O. W. Cummings and T. M. Ulbright (2009). "Estimation of impact of American College of Radiology recommendations on CT colonography reporting for resection of high-risk adenoma findings." Am J Gastroenterol **104**(1): 149-153.

Rex, D. K., P. S. Schoenfeld, J. Cohen, I. M. Pike, D. G. Adler, M. B. Fennerty, J. G. Lieb, 2nd, W. G. Park, M. K. Rizk, M. S. Sawhney, N. J. Shaheen, S. Wani and D. S. Weinberg (2015). "Quality indicators for colonoscopy." Gastrointest Endosc **81**(1): 31-53.

Riley, R. D., J. P. Higgins and J. J. Deeks (2011). "Interpretation of random effects meta-analyses." BMJ **342**: d549.

Riley, S. A., V. Mani, M. J. Goodman, S. Dutt and M. E. Herd (1991). "Microscopic activity in ulcerative colitis: what does it mean?" Gut **32**(2): 174-178.

Riley, S. A., V. Mani, M. J. Goodman, M. E. Herd, S. Dutt and L. A. Turnberg (1988). "Comparison of delayed release 5 aminosalicylic acid (mesalazine) and sulphasalazine in the treatment of mild to moderate ulcerative colitis relapse." Gut **29**(5): 669-674.

Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez and M. Müller (2011). "pROC: an open-source package for R and S+ to analyze and compare ROC curves." BMC Bioinformatics **12**(1): 77.

Robles, L. Y., S. Singh and P. M. Fisichella (2015). "Emerging enhanced imaging technologies of the esophagus: spectroscopy, confocal laser endomicroscopy, and optical coherence tomography." J Surg Res **195**(2): 502-514.

Rogart, J. N., D. Jain, U. D. Siddiqui, T. Oren, J. Lim, P. Jamidar and H. Aslanian (2008). "Narrow-band imaging without high magnification to differentiate polyps during real-time colonoscopy: improvement with experience." Gastrointest Endosc **68**(6): 1136-1145.

Rutgeerts, P., W. J. Sandborn, B. G. Feagan, W. Reinisch, A. Olson, J. Johannis, S. Travers, D. Rachmilewitz, S. B. Hanauer, G. R. Lichtenstein, W. J. de Villiers, D. Present, B. E. Sands and J. F. Colombel (2005). "Infliximab for induction and maintenance therapy for ulcerative colitis." N Engl J Med **353**(23): 2462-2476.

Rutter, M., B. Saunders, K. Wilkinson, S. Rumbles, G. Schofield, M. Kamm, C. Williams, A. Price, I. Talbot and A. Forbes (2004). "Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis." Gastroenterology **126**(2): 451-459.

Rutter, M. D., B. P. Saunders, G. Schofield, A. Forbes, A. B. Price and I. C. Talbot (2004). "Pancolonial indigo carmine dye spraying for the detection of dysplasia in ulcerative colitis." Gut **53**(2): 256-260.

Rutter, M. D., B. P. Saunders, K. H. Wilkinson, S. Rumbles, G. Schofield, M. A. Kamm, C. B. Williams, A. B. Price, I. C. Talbot and A. Forbes (2006). "Thirty-year analysis of a colonoscopic surveillance program for neoplasia in ulcerative colitis." Gastroenterology **130**(4): 1030-1038.

Sakamoto, T., T. Matsuda, T. Aoki, T. Nakajima and Y. Saito (2012). "Time saving with narrow-band imaging for distinguishing between neoplastic and non-neoplastic small colorectal lesions." J Gastroenterol Hepatol **27**(2): 351-355.

- Samaan, M. A., M. H. Mosli, W. J. Sandborn, B. G. Feagan, G. R. D'Haens, E. Dubcenco, K. A. Baker and B. G. Levesque (2014). "A systematic review of the measurement of endoscopic healing in ulcerative colitis clinical trials: recommendations and implications for future research." Inflamm Bowel Dis **20**(8): 1465-1471.
- Sanduleanu, S., T. Kaltenbach, A. Barkun, R. P. McCabe, F. Velayos, M. F. Picco, L. Laine, R. Soetikno and K. R. McQuaid (2016). "A roadmap to the implementation of chromoendoscopy in inflammatory bowel disease colonoscopy surveillance practice." Gastrointest Endosc **83**(1): 213-222.
- Sano, Y., S. Tanaka, S. E. Kudo, S. Saito, T. Matsuda, Y. Wada, T. Fujii, H. Ikematsu, T. Uraoka, N. Kobayashi, H. Nakamura, K. Hotta, T. Horimatsu, N. Sakamoto, K. I. Fu, O. Tsuruta, H. Kawano, H. Kashida, Y. Takeuchi, H. Machida, T. Kusaka, N. Yoshida, I. Hirata, T. Terai, H. O. Yamano, K. Kaneko, T. Nakajima, T. Sakamoto, Y. Yamaguchi, N. Tamai, N. Nakano, N. Hayashi, S. Oka, M. Iwatate, H. Ishikawa, Y. Murakami, S. Yoshida and Y. Saito (2016). "Narrow-band imaging (NBI) magnifying endoscopic classification of colorectal tumors proposed by the Japan NBI Expert Team." Dig Endosc **28**(5): 526-533.
- Sasanuma, S., K. Ohtsuka, S. E. Kudo, N. Ogata, Y. Maeda, M. Misawa, Y. Mori, T. Kudo, T. Hisayuki, K. Wakamura, T. Hayashi, A. Katagiri, H. Miyachi, T. Baba and F. Ishida (2018). "Narrow band imaging efficiency in evaluation of mucosal healing/relapse of ulcerative colitis." Endosc Int Open **6**(5): E518-e523.
- Schroeder, K. W., W. J. Tremaine and D. M. Ilstrup (1987). "Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study." N Engl J Med **317**(26): 1625-1629.
- Shah, S. C., J. F. Colombel, B. E. Sands and N. Narula (2016). "Systematic review with meta-analysis: mucosal healing is associated with improved long-term outcomes in Crohn's disease." Aliment Pharmacol Ther **43**(3): 317-333.
- Shi, H., S. Y. Chen and K. Lin (2016). "Raman spectroscopy for early real-time endoscopic optical diagnosis based on biochemical changes during the carcinogenesis of Barrett's esophagus." World J Gastrointest Endosc **8**(5): 273-275.
- Siau, K., J. Hodson, S. Ravindran, M. D. Rutter, M. Iacucci and P. Dunckley (2019). "Variability in cecal intubation rate by calculation method: a call for standardization of key performance indicators in endoscopy." Gastrointest Endosc **89**(5): 1026-1036.e1022.
- Siau, K., M. Iacucci, P. Dunckley and I. Penman (2020). "The Impact of COVID-19 on Gastrointestinal Endoscopy Training in the United Kingdom." Gastroenterology.
- Sikong, Y., X. Lin, K. Liu, J. Wu, W. Lin, N. Wei, G. Jiang, W. Tai, H. Su, H. Liu and M. Meng (2016). "Effectiveness of systematic training in the application of narrow-band imaging international colorectal endoscopic (NICE) classification for optical diagnosis of colorectal polyps: Experience from a single center in China." Dig Endosc **28**(5): 583-591.
- Sinh, P., N. Gupta, D. S. Rao, S. Wani, P. Sharma, A. Bansal, J. Allen, S. R. Ketover and A. Rastogi (2015). "Community gastroenterologists can learn diminutive colon polyp histology characterization with narrow band imaging by a computer-based teaching module." Dig Endosc **27**(3): 374-380.
- Smith, S. C. L., J. Saltzman, U. N. Shivaji, B. C. Lethebe, R. Cannatelli, G. Birmingham Colonic Polyp Characterisation, S. Ghosh and M. Iacucci (2019). "Randomized controlled

study of the prediction of diminutive/small colorectal polyp histology using didactic versus computer-based self-learning module in gastroenterology trainees." Dig Endosc **31**(5): 535-543.

Solon, C., R. Klausnitzer, D. Blissett and Z. Ihara (2016). "Economic value of narrow band imaging versus white light endoscopy for the characterization of diminutive polyps in the colon: systematic literature review and cost-consequence model." J Med Econ **19**(11): 1040-1048.

Stidham, R. W., W. Liu, S. Bishu, M. D. Rice, P. D. R. Higgins, J. Zhu, B. K. Nallamotheu and A. K. Waljee (2019). "Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis." JAMA Netw Open **2**(5): e193963.

Stroup, D. F., J. A. Berlin, S. C. Morton, I. Olkin, G. D. Williamson, D. Rennie, D. Moher, B. J. Becker, T. A. Sipe and S. B. Thacker (2000). "Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group." JAMA **283**(15): 2008-2012.

Subramaniam, S., B. Hayee, P. Aepli, E. Schoon, M. Stefanovic, K. Kandiah, S. Thayalasekaran, A. Alkandari, P. Bassett, E. Coron, O. Pech, C. Hassan, H. Neumann, R. Bisschops, A. Repici and P. Bhandari (2019). "Optical diagnosis of colorectal polyps with Blue Light Imaging using a new international classification." United European Gastroenterol J **7**(2): 316-325.

Sumimoto, K., S. Tanaka, K. Shigita, D. Hirano, Y. Tamaru, Y. Ninomiya, N. Asayama, N. Hayashi, S. Oka, K. Arihiro, M. Yoshihara and K. Chayama (2017). "Clinical impact and characteristics of the narrow-band imaging magnifying endoscopic classification of colorectal tumors proposed by the Japan NBI Expert Team." Gastrointest Endosc **85**(4): 816-821.

Takemura, Y., S. Yoshida, S. Tanaka, R. Kawase, K. Onji, S. Oka, T. Tamaki, B. Raytchev, K. Kaneda, M. Yoshihara and K. Chayama (2012). "Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video)." Gastrointest Endosc **75**(1): 179-185.

Talari, A. C. S., Z. Movasaghi, S. Rehman and I. U. Rehman (2015). "Raman spectroscopy of biological tissues." Applied spectroscopy reviews **50**(1): 46-111.

Tontini, G. E., T. Rath and H. Neumann (2016). "Advanced gastrointestinal endoscopic imaging for inflammatory bowel diseases." World J Gastroenterol **22**(3): 1246-1259.

Travis, S. P., D. Schnell, P. Krzeski, M. T. Abreu, D. G. Altman, J. F. Colombel, B. G. Feagan, S. B. Hanauer, M. Lemann, G. R. Lichtenstein, P. R. Marteau, W. Reinisch, B. E. Sands, B. R. Yacyshyn, C. A. Bernhardt, J. Y. Mary and W. J. Sandborn (2012). "Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS)." Gut **61**(4): 535-542.

Travis, S. P., D. Schnell, P. Krzeski, M. T. Abreu, D. G. Altman, J. F. Colombel, B. G. Feagan, S. B. Hanauer, G. R. Lichtenstein, P. R. Marteau, W. Reinisch, B. E. Sands, B. R. Yacyshyn, P. Schnell, C. A. Bernhardt, J. Y. Mary and W. J. Sandborn (2013). "Reliability and initial validation of the ulcerative colitis endoscopic index of severity." Gastroenterology **145**(5): 987-995.

- Trivedi, P. J., R. Kiesslich, J. Hodson, N. Bhala, R. A. Boulton, R. Cooney, X. Gui, T. Iqbal, K. K. Li, S. Mumtaz, S. Pathmakanthan, M. N. Quraishi, V. M. Sagar, A. Shah, N. Sharma, K. Siau, S. Smith, S. Ward, M. M. Widlak, R. Bisschops, S. Ghosh and M. Iacucci (2018). "The Paddington International Virtual Chromoendoscopy Score in ulcerative colitis exhibits very good inter-rater agreement after computerized module training: a multicenter study across academic and community practice (with video)." Gastrointest Endosc **88**(1): 95-106 e102.
- Truelove, S. C. and W. C. Richards (1956). "Biopsy studies in ulcerative colitis." Br Med J **1**(4979): 1315-1318.
- Truelove, S. C. and L. J. Witts (1959). "Cortisone and corticotrophin in ulcerative colitis." Br Med J **1**(5119): 387-394.
- Turner, D., A. Ricciuto, A. Lewis, F. D'Amico, J. Dhaliwal, A. M. Griffiths, D. Bettenworth, W. J. Sandborn, B. E. Sands, W. Reinisch, J. Scholmerich, W. Bemelman, S. Danese, J. Y. Mary, D. Rubin, J. F. Colombel, L. Peyrin-Biroulet, I. Dotan, M. T. Abreu, A. Dignass and I. B. D. International Organization for the Study of (2021). "STRIDE-II: An Update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) Initiative of the International Organization for the Study of IBD (IOIBD): Determining Therapeutic Goals for Treat-to-Target strategies in IBD." Gastroenterology.
- Uchiyama, K., T. Takagi, S. Kashiwagi, Y. Toyokawa, M. Tanaka, Y. Hotta, O. Dohi, T. Okayama, N. Yoshida, K. Katada, K. Kamada, T. Ishikawa, O. Handa, H. Konishi, M. Kishimoto, N. Yagi, Y. Naito and Y. Itoh (2017). "Assessment of Endoscopic Mucosal Healing of Ulcerative Colitis Using Linked Colour Imaging, a Novel Endoscopic Enhancement System." J Crohns Colitis **11**(8): 963-969.
- Veenstra, M. A., O. Palyvoda, H. Alahwal, M. Jovanovski, L. A. Reisner, B. King, J. Poulik and M. D. Klein (2015). "Raman spectroscopy in the diagnosis of ulcerative colitis." Eur J Pediatr Surg **25**(1): 56-59.
- Venables, W. N., B. D. Ripley and W. N. M. a. s. w. S. P. Venables (2002). Modern applied statistics with S. New York, Springer.
- Villanacci, V., E. Antonelli, F. Lanzarotto, A. Bozzola, M. Cadei and G. Bassotti (2017). "Usefulness of Different Pathological Scores to Assess Healing of the Mucosa in Inflammatory Bowel Diseases: A Real Life Study." Sci Rep **7**(1): 6839.
- Villanacci, V., E. Antonelli, G. Reboldi, M. Salemm, G. Casella and G. Bassotti (2014). "Endoscopic biopsy samples of naive "colitides" patients: role of basal plasmacytosis." J Crohns Colitis **8**(11): 1438-1443.
- Villanacci, V., E. Antonelli, M. Salemm and G. Bassotti (2015). "Assessing mucosal healing in ulcerative colitis: the simpler, the better." Endoscopy **47**(8): 759.
- Villanacci, V., G. Bassotti and C. Langner (2015). "Histological remission in inflammatory bowel disease: where are we, and where are we going?" J Crohns Colitis **9**(5): 428.
- Villanacci, V., L. Reggiani-Bonetti, F. Caprioli, L. Saragoni, T. Salviato, C. Mescoli, G. Canavese, S. Manenti, E. Spada, L. Baron, G. Leoncini, M. Cadei, S. Battista and A. Armuzzi (2019). "Histopathology of inflammatory bowel disease - Position statement of the Pathologists of the Italian Group for the Study of Inflammatory Bowel Disease (IG-IBD) and Italian Group of Gastrointestinal Pathologists (GIPAD-SIAPEC)." Dig Liver Dis.

- Walmsley, R. S., R. C. Ayres, R. E. Pounder and R. N. Allan (1998). "A simple clinical colitis activity index." Gut **43**(1): 29-32.
- Ward, S. T., M. A. Mohammed, R. Walt, R. Valori, T. Ismail and P. Dunckley (2014). "An analysis of the learning curve to achieve competency at colonoscopy using the JETS database." Gut **63**(11): 1746-1754.
- Watson, A. J. (2016). "Confocal Laser Endomicroscopy in the Management of IBD: Clearly Superior?" Dig Dis Sci **61**(7): 1785-1787.
- Wood, J. J., C. Kendall, J. Hutchings, G. R. Lloyd, N. Stone, N. Shepherd, J. Day and T. A. Cook (2014). "Evaluation of a confocal Raman probe for pathological diagnosis during colonoscopy." Colorectal Dis **16**(9): 732-738.
- Wu, L., Y. Li, Z. Li, Y. Cao and F. Gao (2013). "Diagnostic accuracy of narrow-band imaging for the differentiation of neoplastic from non-neoplastic colorectal polyps: a meta-analysis." Colorectal Dis **15**(1): 3-11.
- Youden, W. J. (1950). "Index for rating diagnostic tests." Cancer **3**(1): 32-35.
- Zenlea, T., E. U. Yee, L. Rosenberg, M. Boyle, K. S. Nanda, J. L. Wolf, K. R. Falchuk, A. S. Cheifetz, J. D. Goldsmith and A. C. Moss (2016). "Histology Grade Is Independently Associated With Relapse Risk in Patients With Ulcerative Colitis in Clinical Remission: A Prospective Study." Am J Gastroenterol **111**(5): 685-690.