What makes an expert Barrett's pathologist?

Concordance and pathologist expertise within a digital review panel

van der Wel, M.J.

GENERAL DISCUSSION AND FUTURE PERSPECTIVES

M. J. van der Wel, L. C. Duits

AIM OF THIS THESIS AND RELEVANCE OF PROBLEM

Our aim was to set up a national, digital, histopathological review panel for patients with biopsies containing dysplastic BE. Such a panel, consisting of a collaboration of expert BE pathologists, would lead to a more accurate histological diagnosis, allowing for a more accurate risk stratification according to progression risk, and subsequent efficient allocation of endoscopic surveillance and treatment.

BE is the precursor lesion for EAC, a deadly cancer with a strongly increased incidence in the Western world over the last 30 years, to 0.7-2.8 per 100,000 person-years in 2012. [1 2] Patients often present with advanced stage disease and concurrent dismal prognosis, with a 5-year survival rate of approximately 15%. In the precursor lesion BE, the normal stratified squamous lining of the esophagus is replaced by columnar epithelium containing intestinal metaplasia. [3] Patients with BE have varying risks of progression to HGD or EAC, from 0.1-0.5% [4-6] per patient-year for NDBE, to 9-13% per patient-year for LGD. [7 8] Therefore, all patients with BE receive regular endoscopic surveillance, dysplastic BE at a higher frequency than NDBE, with 4-quadrant biopsies taken of the entire Barrett segment according to the Seattle protocol. [9] These biopsies are then histologically evaluated by a pathologist for the presence of dysplasia.

However, current surveillance strategies are hampered by general pathologist observer variation and lack of training, leading to over- and underdiagnosis in this patient group. Earlier studies have shown that patients with community diagnosed LGD could often be downstaged to NDBE by expert BE pathologists, and consecutively exhibited a concurrent low progression risk. [7 8] Patients with confirmed LGD by expert BE pathologists had a significantly higher risk to develop EAC. Therefore, in the Netherlands, a workflow was constructed whereby patients with community diagnosed dysplastic BE were referred to a so-called 'Barrett expert center' and, if the dysplasia was confirmed, would receive regular surveillance there. Requirements for a Barrett expert center are described in the European Society of Gastrointestinal Endoscopy Position Statement. [10] There are eight of these Barrett expert centers in the Netherlands consisting of a collaboration of endoscopists and also, since finalizing this thesis, a collaboration of pathologists. These endoscopist-pathologist teams are jointly responsible for the diagnosis and treatment of all patients with dysplastic BE and/or complex BE segments. The workflow refers BE patients with community diagnosed LGD to a Barrett expert center. Oftentimes, they are subsequently

downstaged to NDBE by an expert BE pathologist. The rationale behind this thesis is, that it would be much more efficient if all BE patients with community diagnosed LGD were first histopathologically confirmed by expert BE pathologists, before referring them to a Barrett expert center. In this way, less endoscopic and histopathological oversurveillance would occur, which would improve quality of life and reduce health care costs by reducing the number of unnecessary surveillance endoscopies.

SUMMARY OF RESULTS

During the course of this thesis, we have set up a collaboration among the Dutch pathologists responsible for the diagnostic work-up of dysplastic BE biopsies in the Netherlands. These pathologists were all working at one of the eight Barrett expert centers in the Netherlands, but had not been working together on a national scale before the start-up of this collaboration and digital review panel. We chose the structure of a panel on a national level to ensure uniform review of the BE biopsy material. We chose digital review over conventional glass slide review in order to enable multiple assessors to view the slides at once, to prevent breakage or loss of slides during transportation, to allow for easy remote group discussions of difficult cases and to generate annotated histopathological material for teaching purposes. After five years of preliminary studies, all pathologists of the panel jointly had assessed 31,500 slides, yielding 6,000 individual case diagnoses. As from January 2017, the national digital review panel 'LANS' (Landelijk Adviesorgaan Neoplasie Slokdarm) has officially been in operation, has reviewed 400 cases to date and currently consists of 16 expert BE pathologists and expert BE pathologists in training. To our knowledge, this is the first histopathologist review panel worldwide that has quantified its complete set-up, and in such a transparent, step-wise manner. General recommendations for panel set-up, learnt from this process, are described below.

Throughout the process of setting up the national digital review panel for dysplastic BE biopsies, we quantified the performance of the (aspiring) panel members. Quantification of performance was carried out by scoring variables related to 1) their consistency over multiple assessment rounds, 2) their contribution to panel diagnoses, 3) their agreement with a gold standard, and 4) their number of significant misdiagnoses. The gold standard with which their scores were compared were generated by five expert BE pathologists considered the 'core' of our panel. These pathologists were defined as such, because of their extensive experience, and

recognition as experts by their peers. Moreover, they had successfully collaborated on previous BE intervention studies where patient outcome had been evaluated prospectively, [7 8 11-17] as well as on the Amsterdam Barrett's Advisory Committee. [11]

Scoring the variables explained above and comparing them to a consensus gold standard diagnosis across multiple study sets ensured optimal training and eventually inclusion of these aspiring panel member-pathologists into the panel. Moreover, assessing all BE biopsy cases by this pathologist group has yielded a large amount of histopathological annotations that can be used for teaching purposes in the future. The panel is now assessing approximately three BE review cases per week. Improvement and maintenance of pathologist assessment skills is maintained by continuous structural education in the form of peer feedback: monthly group teleconferences in which review cases without a majority diagnosis are reviewed and jointly discussed in a protected, online environment. Future maintenance also includes periodic retesting of the expert BE pathologists, using the same benchmark quality criteria within new study sets.

Quantification of expertise of this group of Dutch pathologists led to two additional lines of research, the first results of which are described in this thesis as well. The first additional research line investigates pathologist assessment concordance worldwide. This first study calculated the concordance of a cohort of 51 GI pathologists with varying levels of expertise, by comparing their histopathological assessments of BE biopsies to the consensus diagnoses of four Dutch expert BE pathologists. Subsequently, in multivariate analysis, their concordance was compared to their demographic background, i.e. factors like working situation and BE assessment experience, in order to distil demographic histopathologist factors predictive of high concordance. The second additional research line investigates pathologist assessment concordance in ER specimens. This first study is a descriptive study, related to the concordance of the pathologists concerning specific risk factors that increase the risk of lymph node metastases and can be present in ER specimens. The ultimate goal of these two additional lines of research is to be able to expand the review panel to a European level, and to BE-related endoscopic resections in the future.

PANEL SET-UP: PRE-REQUISITES AND ASSOCIATED RECOMMENDATIONS

From our experience of setting up our national digital review panel for dysplastic BE biopsies, we feel that we have been able to distill three general prerequisites that can aid the general set-up of a histopathologist review panel. These are described below, together with recommendations per prerequisite.

The first prerequisite is a digital set-up, with an adequate platform for slide viewing plus panel administration. Advantages of a digital workflow are practical, such as no more breakage or loss of slides; disseminative, such as teaching purposes or remote consultations; and innovative, such as the potential application of artificial intelligence. Worldwide, Dutch pathology departments are considered among the most innovative. However, when this thesis went to print, only a few laboratories in the Netherlands had implemented a fully digitalized workflow. This means that in daily practice, Dutch pathologists are generally not used to performing primary diagnostics on-screen instead of behind a microscope. We would therefore recommend any review panel seeking to digitalize its workflow, to perform at least one validation study for the tissue type that is to be reviewed. This will become of less importance as more laboratories implement a fully digital workflow. Nevertheless, the set-up of this national digital review panel started by validating the use of whole slide imaging for the assessment of BE biopsies, hereby elaborating on many earlier validation studies for different tissue types. [18-20] Regarding the viewing platform, it is important that the interface used is intuitive, compatible with firewalls of different hospitals, and open to different slide scanning formats. Over the course of this thesis, we have noticed that the pathologists needed to adjust to the interface more than to the digital microscopic viewing of slides itself. For panel administration, it is highly recommended to choose an interface where some kind of 'dashboard' can be implemented, so that the panel administrator can keep track of which cases have been uploaded, which cases have been assessed by which pathologists and which diagnoses they assigned. In other words, there should be an opportunity to upload or produce a CRF within the dashboard, so that pathologists can fill in their case assessments and the administrator can extract the information later on. In our case, we co-developed our panel module with Pathology Image Exchange (PIE), a program launched by the Dutch Pathology Society (NVVP) and linked to the national pathology biobank 'PALGA'. [21]

The second prerequisite is adequate expertise of the panel members. BE guidelines describe that review of dysplastic BE should be performed by a 'second, preferably gastrointestinal (expert) pathologist'. [22-25] However, up until the results of this thesis were published, the definition of a 'expert BE pathologist' was very vague, namely 'considered an expert by his/her peers'. We feel that this is quite a dangerous proposition to use as a base for guideline recommendations, and we are confident that guideline development can be aided by the results of this thesis in the future. To strengthen the panel performance, quantitatively defining and improving performance of the panel pathologists has been a sub-aim from the beginning. We started the panel with 5 'core' expert BE pathologists, whose diagnoses had earlier on been proven to concur with the clinical follow-up of dysplastic BE patients. Therefore, the assessments of these pathologists could be used as a baseline for panel assessments and performance quantification of future panel members (see also the third prerequisite below). We believe that the answer to the question 'what is an expert Barrett's pathologist?' is multi-facetted, and that the studies described in this thesis have sought answers to this question from different angles. First, from the multivariate analyses in **Chapter 10**, we have quantified three demographic histopathologist factors associated with diagnostic concordance at expert level, namely; at least five years of experience commensurate with age, working at a teaching hospital or using p53 IHC when one is not. This study also allowed us to test the veracity of expert self-identification. Two-thirds of the 51 pathologists participating in this study self-identified as experts, however, the univariate analysis showed that self-identification as an expert did not predict diagnostic concordance at expert level. Even though it is difficult to translate the results of this study into objective recommendations, they are an important proof of concept that these factors lead to better pathologist performance. Second, from the sequential studies conducted in the course of the national digital review panel set-up in the Netherlands, we have improved group pathologist performance by hands-on-training. Depending on the demands placed on a panel, facilitating the improvement of general panel performance in such a way can be important.

Therefore, the third prerequisite is training, expanding and maintaining panel member performance. In our case, this third prerequisite was the most challenging, to objectify as well as to attain. All GI pathologists responsible for the work-up of BE patients that were working at the Barrett's expert centers eventually joined the

panel after its initial set-up by the core pathologists. For the first time in the history of histopathologist expert panels, we aimed to objectively quantify baseline levels of pathologist performance, to thereafter improve it if necessary. Therefore, we developed four benchmark quality criteria. These were: intra-observer agreement (measuring consistency of the pathologist over multiple assessment rounds), number of diagnoses indefinite for dysplasia (measuring the contribution of the pathologist to the panel diagnoses), percentage agreement with the consensus gold standard diagnosis and number of significant misdiagnoses, i.e. diagnosing NDBE as HGD or vice versa. Over the course of five years, all pathologists that would eventually participate in the panel followed a training program consisting of multiple assessment rounds of multiple BE biopsy case sets enriched for difficult dysplastic cases. All case sets had a consensus gold standard diagnosis available per benchmark criterion, for every case, generated by the five core expert pathologists. After each assessment round, a face-to-face group discussion of all discrepant cases together with the five core expert pathologists took place. In this way, we exposed the GI pathologists to continuous dysplastic BE biopsy histology education. Currently, performance is maintained by monthly group teleconferences in which review cases without a majority diagnosis are reviewed and discussed together in an online environment. We would strongly recommend defining benchmark quality criteria and testing these in multiple study sets if the panel includes a larger group of pathologists, in order to supply every participant with the necessary training to attain the appropriate expertise. We would also recommend planning maintenance of performance and periodic retesting. Earlier studies have shown that the endoscopists of the Barrett expert centers have also strongly benefited from endoscopy training. [26] By ensuring that all pathologists of the Barrett expert centers have been quantitatively trained, we also strengthened the position of the Barrett expert centers as a whole.

IMPLICATIONS OF RESULTS

The set-up of this panel has resulted in more accurate stratification and therefore improved care of patients with dysplastic BE. This can be pinpointed to two specific areas. First of all, the set-up of this panel has led to improved multidisciplinary BE teams in all Barrett expert centers and to a much anticipated collaboration between the pathologists of these centers. Together with the endoscopists, they now form a team of experts for the diagnosis and treatment of all patients with dysplastic BE in the Netherlands. The multidisciplinary meetings that are held on a regular basis,

where common research efforts and protocols are discussed for diagnostics as well as treatment, have improved pathologist and endoscopist expertise for dysplastic BE in all expert centers, as well as the communication between these two groups of medical specialists. Second, improved pathologist performance within the panel has led to more accurate surveillance intervals per patient. Within the panel, preliminary results indicate that two-thirds of the community dysplastic BE cases sent for review could be downstaged to NDBE, leading to a significant decrease in the number of surveillance endoscopies. Restricting the surveillance endoscopies to patients that truly need them has direct consequences for the BE patient group at large. It means a lower disease burden and higher quality of life for downstaged patients requiring less endoscopies, and a shorter waiting list and higher accessibility to care for patients that do need an increased frequency of surveillance endoscopies and/or (endoscopic) treatment. Moreover, performing less surveillance endoscopies in general leads to a significant health care cost reduction. If two-thirds out of the 1,000 patients annually diagnosed with LGD in community hospitals can be downstaged to NDBE, their surveillance interval can be extended to 3 years. This signifies a reduction of 1,980 endoscopies or € 2,376,000 per year in unnecessary Barrett surveillance endoscopies.

PATIENT STRATIFICATION IN OTHER FIELDS

Over the past 20 years, many research efforts have concentrated on reaching a more objective BE patient stratification level. These efforts have been concentrated in different areas such as histology, endoscopy, technology, and artificial intelligence, among others.

Histology

Histopathological assessment of esophageal biopsies is subjective and therefore hampered by observer variation, as discussed above. Ideally, risk stratification in BE is based on objective measures such as biomarker panels. Many research effort are focusing on identifying and validating biomarkers to predict neoplastic progression in BE patients. Several leading research groups in this field have published promising biomarker panels. The Seattle group published studies on a panel including abnormal DNA ploidy and loss of heterozygosity for genes 9p and 17p, which was highly associated with neoplastic progression (relative risk 38.7; 95% CI 11-139). [27-29] The Baltimore group has identified a panel of genes that was methylated early and often in neoplastic progression. [30] Subsequent studies with different prediction models

yielded promising results, with an area under the receiver-operating characteristic curve of 0.84. In multivariate analysis, odds ratios for individual biomarkers ranged from 1.74 to 1.80. [31 32] The Cambridge group identified a biomarker panel consisting of a consensus diagnosis of LGD, abnormal DNA ploidy, and Aspergillus oryzae lectin. In patients with consensus LGD, the adjusted OR for progression was 3.74 (95% CI 2.43-5.79) for each additional biomarker and the risk increased by 2.99 (95% CI 1.72-5.20) for each additional marker in patients without LGD. [33] The Amsterdam Tytgat group constructed a prediction model for progression of NDBE to HGD/EAC based on age, BE length and the markers p16, MYC and aneusomy. On multivariate analysis, these three markers identified a high-risk group with a 8.7-fold increased hazard ratio of neoplastic progression compared to the low-risk group. [34]

Most of the previously published case-control studies, however, have several methodological shortcomings that mainly relate to the adequacy of their sample selection. A major limitation in most studies is an insufficient interval between baseline endoscopy and the time point of progression. In a case control setting, cases (patients who eventually progressed to HGD or carcinoma) should be free of disease in the material that is tested for biomarkers at baseline. Many studies included progressor patients with a relatively short interval between baseline and progression diagnosis (i.e. 6 months) and have included patients with advanced cancer for whom such a short interval is even more inappropriate. Additionally, most studies did not include quality of baseline sampling as an inclusion criterion. These two issues are likely to have led to inclusion of progressor patients in whom the neoplasia was simply already present at baseline. This makes these studies prone to overestimating the ability of tested biomarkers in the prediction of future progression to neoplasia. Therefore, our group set up a cohort using stringent inclusion criteria for baseline sampling, with which a validation study for a biomarker panel previously identified by the Cambridge group was performed. [35 36] This panel originally included aneuploidy, expert confirmed LGD, and immunohistochemistry for p53, aspergillus oryzae lectin (AOL) and cyclin A. The biomarker in this study that demonstrated the highest predictive ability was expert confirmed LGD. It might be suboptimal, however, to incorporate this markedly subjective assessment in a, preferably, objective biomarker panel. Additionally, clinical management of LGD is evolving rapidly and current international guidelines recommend considering patients with confirmed LGD for prophylactic ablation. Future

biomarker panels should aim for correct risk stratification among Barrett's patients who have no dysplasia.

A different technique that also uses the principles of biomarker assessment is termed TissueCypher. It is a multi-biomarker panel based on quantification of epithelial and stromal variables by visualizing them through multiplex fluorescence staining. [37] Different features for each biomarker have been analyzed using automated image analysis software, which utilizes algorithms for collection of quantitative biomarker and morphology feature data at the cellular and subcellular level, and within tissue compartments such as epithelium, metaplasia, and lamina propria. The results of this study were promising. The test incorporates 3-tier stratification to classify patients as low-, intermediate-, or high- risk for progression. The predicted high-risk group had a 9.4-fold increased risk of developing HGD or adenocarcinoma compared to the low-risk group. This biomarker test has several advantages over previously described biomarker panels. It is performed on formalin fixed paraffin embedded biopsies and uses immunohistochemical and immunofluorescent labeling of slides. Subsequent analysis is performed using objective, automated software. These are all characteristics that are essential for implementation in the community setting of Barrett's surveillance. However, actual clinical implementation of this test requires additional research.

Another non-endoscopic, cytology-based technique incorporates a cell-collection device called the Cytosponge, to scrape cells from the esophageal lining via a foam sphere on a string that is swallowed by the patient before being pulled back up through the esophagus. The resultant tissue lacks the tissue architecture of a biopsy, but is able to sample the epithelial layer of the whole Barrett segment by means of loose cells and tissue fragments. This non-endoscopic screening test was recently coupled to a biomarker panel involving protein biomarkers, methylation markers, TP53 mutation status and glandular atypia in order to identify patients with low risk of progression that could be surveilled with the Cytosponge without having to undergo an endoscopy with biopsies. [38] After multivariate analysis, the device coupled with certain biomarkers and patient characteristics could stratify patients into three low-, intermediate- or high-risk groups, whereby high-risk patients would need to undergo endoscopic surveillance with 4-quadrant biopsies to confirm the presence of dysplasia. The Cytosponge would therefore be a useful tool to stratify patients in a primary care setting, which is the diagnostic step preceding panel review.

A relatively new technique is the wide area transepithelial sample with 3D analysis (WATS 3D). The WATS 3D device uses an abrasive brush to obtain a trans-epithelial specimen of the full thickness of the esophageal mucosa. The WATS specimen is then optically imaged in 50 1-mm slices, which are integrated together to form a 3-dimensional image, after which computer-assisted analysis makes use of multi-layered focus and abnormal cell localization algorithms to generate a diagnosis. Hence, as opposed to standard brushing techniques, WATS 3D allows for evaluation of glandular morphology and leads to improved dysplasia assessment compared to cytology alone. All specimens undergo computer-assisted three-dimensional tissue analysis utilizing neural networks specifically optimized for the esophageal mucosa. The computer is able to detect very small numbers of atypical cells on the slide, which are then clearly displayed separately with location correlation to the glass slide. An expert gastrointestinal pathologist, specially trained for this technique, evaluates the selected cells as well as the WATS 3D slides. Several recent studies have demonstrated an increase in detection of BE as well as dysplasia when the WATS technique was added to random biopsies. [39-41] In a multicenter US study among BE surveillance patients, WATS was found to be approximately four times more effective in the detection of HGD/EAC than the Seattle protocol with random biopsies. [41] The sensitivity of exfoliative cytology in the esophagus is traditionally low. Standard cytology brushes are typically soft, as they are primarily designed to gently remove spontaneously exfoliated squamous cells. These brushes could be insufficiently abrasive to sample deeper layers of the more firmly attached glandular epithelium as found in Barrett's esophagus. Moreover, the sensitivity of esophageal cytology is limited by the absence of tissue architecture in the specimen cells. Detection of abnormality hereby often relies on the success of a manual search for a few dysplastic cells randomly scattered among the benign glandular cells typically found in the specimen. Therefore, the WATS 3D device makes use of a stiff brush and is able to obtain a disintegrated, trans-epithelial specimen of the full mucosal thickness, neutralizing these traditional cytology limitations. In the disaggregated tissue specimen obtained by a stiff biopsy brush, the analysis of the specimen could have been complicated by the fact that the resultant smear directly deposited on the microscope slide is up to 100 microns in thickness. This is in contrast to the 2-4 micron thickness of a standard histology section. This limitation is overcome by the fact that the WATS specimen is optically imaged and then re-integrated to form a 3-dimensional image.

Endoscopy

In the field of endoscopy, over the last ten years many imaging techniques have been discovered or significantly advanced. Volumetric laser endoscopy (VLE) incorporates optical coherence technology, which provides cross-sectional images of (glandular) structures underneath the tissue surface. In BE, it detects changes at the mucosal surface and in buried glandular architecture. [42] The result is a VLE image consisting of different shades of grey, based on the amount of backscatter reflected from different types of tissues. First, these images have been related to their histological counterparts. This proved to be quite a challenge. [43] Recently, the interpretation of these images for patient stratification purposes has been made easier by a prediction score combined with machine learning methods to stratify BE patients with early neoplasia. [44 45] VLE has the advantage that there is no sampling error, since the whole BE segment is visualized. A significant disadvantage is its high cost and complicated quantitative analyses that are necessary before any stratifying prediction can be made. Up until now, histological correlation is still necessary and VLE is an addition to, rather than a replacement of histopathological assessment.

Deep learning

The last technique discussed here is termed 'deep learning' and refers to a set of computer models that are being developed in order to aid clinicians in diagnostic decisions. Up until now, most research efforts in this field have concentrated on the fields of radiology and pathology, as both of these consist of digitalized images of some sort. In pathology, convolutional neural networks, which are computer algorithms that are self-learning after being 'taught' by whole slide images annotated by pathologists, have been able to correctly identify lymph node metastases in whole-slide images from breast cancer patients. [46] A significant advantage of the development of these techniques is, that computers are able to transform fixed diagnostic categories employed by a pathologist (i.e. NDBE, LGD, HGD and IND) into a continuous diagnostic scale. Moreover, they can integrate additional visual information, clinical and molecular data to this stratification. However, this technique is still based on expert pathologist assessment, because the convolutional neural networks make use of annotated whole slide images to train themselves. Therefore, the quality of the final algorithm depends on the expertise of the pathologists that trained it. In the future, convolutional neural networks that stratify BE biopsies according to dysplasia grade might be developed. If they evolve to be sensitive enough to accurately discriminate between different

diagnostic categories they could possibly serve as an additional expert pathologist within our national digital review panel in the future.

In conclusion, over the coming years, many newly discovered techniques will be unleashed unto the field of dysplastic BE lesions and concurrent efforts for the improvement of patient stratification. Many techniques are promising, but most will still require validation by the gold standard diagnostic techniques, which are endoscopic and histopathological (expert) assessment. Improved understanding of BE biology coupled to discovery of solid biomarkers, in combination with minimally invasive techniques and computer algorithms based on pathologist performance will help clinicians to make informed decisions about surveillance strategies.

FUTURE PERSPECTIVES

Currently, risk stratification in BE surveillance is solely based on the subjective histological assessment of surveillance biopsies. Promising objective biological markers that may estimate the risk of neoplastic progression are being investigated. These are, however, not yet sufficiently validated to be implemented in regular clinical practice. One of the most important issues in current risk stratification is, therefore, the improvement of histological assessment, as investigated during the course of this thesis. The term 'expert pathologist' is frequently used in literature but is generally ill-defined. To describe the expert BE pathologist more accurately, we have developed four histopathological benchmark criteria and deduced objective demographic histopathologist factors, and related both to pathologist performance. Subsequently, we developed benchmark quality values for BE biopsy assessment in order to evaluate pathologist performance within the context of the national digital review panel we set up. This process enables a certain quantification of expert pathology in BE. Subsequent research is required in order to periodically monitor the quality of our expert pathology panel. The clinical implementation algorithm, described in **Chapter 9** and discussed further below, needs to be validated on an independent dataset with a larger sample size. Following validation, the algorithm can be implemented in clinical practice and will subsequently also require periodic evaluation to ensure that quality and homogeneity of the panel's assessments is maintained.

The future of histopathological assessment

Now that the panel is set up, we need to ensure that it will be future-proof. It should be attractive for laboratories nationwide to submit their community dysplastic cases to the national digital review panel. This includes the aspects discussed below.

First, embedding the panel in the standard work-up of dysplastic BE nationwide. Even though the hospitals in the Amsterdam region, as well as a number of hospitals mid-country, submit their dysplastic cases to the panel, some regional panels are still in operation. In order to achieve homogenous review assessments, these regional panels ought to gradually merge into the national digital review panel. In order to be able to cope with the expanded number of cases on a national level, adequate secretary support is paramount. Therefore, the review panel should ideally be reimbursed by insurance companies or by integration in the 'diagnosis and treatment' combinations (DBCs) that are used by Dutch healthcare providers to declare care-related costs. Support from the Dutch Society of Pathology (NVVP) will further help to achieve this.

Second, optimizing assessment interface and workflow. This is one of the most important prerequisites to make the panel run smoothly. It should be easy to submit as well as assess a review case. The panel currently runs within Pathology Image Exchange (PIE), a program developed by Sectra (Almere, Benelux) and is linked to PALGA, the national pathology database. [21] PIE uses a central, secured server to which the digitalized slides are uploaded, currently by the central laboratory at the Amsterdam University Medical Centers. With the increasing number of Dutch pathology laboratories digitizing their workflow in the coming years, more and more laboratories will be able to upload their own cases to PIE, which will make the workflow more efficient and lean. PIE contains the digitalized images as well as the CRF the panel members fill in with every assessment. The panel has been running on PIE for 6 months now, and we feel confident that future improvements will make it even more easy to use and run smoothly. This will also aid in minimizing the through-put time of the cases. From the start of the panel in January 2015 up to now, the through-put time has decreased from 14 weeks to eight weeks, but the ideal through-put time would be approximately four weeks.

Third, optimizing involvement and performance of all panel members. We feel that optimal performance of panel members is achieved by periodic re-testing according to

the benchmark quality criteria we developed on earlier study sets. This testing can be performed on real-time review cases assessed by the panel members, and discussed during one of the bi-annual face-to-face meetings we organize. Also, periodic re-testing of the algorithm is necessary. Optimal involvement of panel members can be achieved by a circulating Chair during the monthly online consensus meetings.

Fourth, expansion of the panel with additional panel members. It is important to anticipate on panel member transitions and number of review cases, by timely training new expert BE pathologists. Currently, at least one expert BE pathologist is involved in the panel for every Barrett expert center. With the number of Barrett expert centers expanding in the future, pathologists from those centers are already following the extensive BE expertise training described in this thesis. These new panel members have concurrently started assessing real time review cases, and participating in the monthly online consensus meetings, since they do not have the feedback by peers in the form of per-set group discussions about cases discrepant from the consensus gold standard diagnosis, like the first panel pathologists had. Besides training new panel members, we aim to engage the time and expertise of the existing panel members as efficiently as possible. Therefore, we have developed a clinical implementation algorithm to calculate the number of assessing expert BE pathologists necessary to yield a reliable diagnosis per diagnostic category. Importantly, with the help of this algorithm we have been able to deduce that a reliable diagnosis, consistent with the consensus gold standard diagnosis, can be yielded when only a subset of the panel pathologists assesses each case. From the first study in **Chapter 9**, we could conclude that only a minority of dysplastic review requests needs to be confirmed by a second panel pathologist. After the algorithm is validated in a new case set consisting of dysplastic review requests, it will greatly increase the work efficiency of the review panel and its members.

Lastly, European expansion of the panel. This is an ambition, however, that cannot yet be realized, since harmonized pan-European grading of dysplastic BE has not been reached yet. In this respect, the focus should first lie on disseminating expertise through courses and training, before reaching out to selected centers for panel expansion. If the time is there, expansion should occur in consultation with the European Society of Pathology and the European Society of Gastrointestinal Endoscopy.

Review panel expansion

The national expert pathology platform is currently operative for review of BE surveillance biopsies demonstrating IND, LGD or HGD. In the near future, the indication for referral to the platform should be expanded to include EMR and ESD specimens. While endoscopic treatment of BE neoplasia is adequately centralized in the Netherlands, which implies that endoscopic resection specimens are evaluated by an expert pathologist, evaluation of these specimens by an additional expert pathologist may further improve quality of patient management. In general, it is assumed that there is less observer variation when more tissue is available for assessment (as is the case with endoscopic resections). However, this assumption has never been well researched and established for specific risk factors that increase the risk of lymph node metastases and can be present in endoscopic resection specimens. It is essential that these risk factors (i.e. poor differentiation grade, submucosal infiltration depth, presence of lympho-vascular invasion and irradicality of the basal resection margin) are accurately assessed, since the presence of (one of) these factors can tip the scale towards a minimally invasive endoscopic resection or invasive surgical esophagectomy. Our descriptive study on this subject, **Chapter 11** of this thesis, whereby expert BE pathologists from the national review panel assessed ER specimens, showed low observer agreement for each feature. We discovered these discordances to be due mostly to poorly defined diagnostic criteria. After extensive group discussions we were able to propose clearer definitions for each feature and suggest clinically relevant groupings of diagnostic categories per feature. The most important implication of our study is, however, that ER specimens containing high-risk histological features should be reviewed by a second, expert pathologist, just like dysplastic BE biopsies. The national digital review panel for biopsies is the perfect vehicle to expand this to.

The most important propositions distilled from the group discussions, that are still to be validated in a follow-up study, are the following: in order to find the deepest tumor invasion, perform a desmin stain to highlight the course of the muscularis mucosae, and provide a range of invasion depth measurements in microns. In order to score differentiation grade as objectively as possible, the most advanced component of the tumor should be scored. Lymphovascular invasion and resection radicality at the basal margin should be assessed with the help of additional cuts and immunohistochemical stainings. Using the results of this study, we were able to suggest clinically relevant groupings of diagnostic categories per feature. For example, assessment discordances

in depth of invasion only matter when they result in a different treatment regimen. In the Dutch Barrett expert centers, this cut-off point currently sits at sm1 – sm2 invasion depth. In the future, it is even likely that all submucosal cancers will be treated endoscopically, shifting the cut-off point to sm3 invasion versus muscularis propria invasion unless a high-grade feature is present. The presence of lymphovascular invasion, poor differentiation grade or an irradical endoscopic resection will then still lead to upstaging of treatment. [10] In order to validate and implement review of EMR and ESD specimens in the setting of our digital expert review panel, future research should ideally follow a sequence of studies, similar to the sequence followed for the development of the dysplastic BE biopsy expert panel. These studies should incorporate the propositions and diagnostic category groupings explained above. Furthermore, a teaching module for histological assessment of EMR and ESD specimens can then be developed, similar to what is described below.

In the current situation, our national digital review panel is strictly operational on a national level. The digital setting, however, would lend itself perfectly for international expansion. International expert pathologists could qualify for the expert panel after assessing the digitalized slides and attaining the benchmark values for quality criteria we developed and described in **Chapter 6 to 8** of this thesis. Digital pathology is rapidly evolving and it is to be expected that in the upcoming years, laboratories will all undergo full digitalization and adopt a digital workflow. This would further facilitate the digital review process, and would more easily enable international expansion, since slides would not require to be scanned at a central location.

Development of a training module for general pathologists

While improving and maintaining assessment quality of expert pathology in BE is important, the vast majority of BE patients undergo endoscopic surveillance in community centers and will never develop dysplasia. This implies that improving diagnostic accuracy among community-based pathologists, especially in the distinction between reactive changes and dysplasia, is paramount. The digital pathology review panel we have developed is the ideal setting to achieve this goal. First, the expert pathologist can supply direct feedback to the referring pathologist by using annotated slides. Second, in this way the digital pathology review panel automatically generates annotated slides, that can be incorporated into a teaching tool for histological assessment of BE biopsies, similar to the previously discussed BORN

training program for endoscopic recognition of early neoplastic changes. In order for a teaching module to be developed, the expert panel pathologists need to generate annotated slides containing architectural and cytological features that are important in the assessment of dysplasia. Examples of these criteria, incorporated into a teaching module could eventually lead to a more uniform evaluation of BE biopsies nationwide.

Conclusion

In conclusion, over the course of this thesis we have developed a national digital review panel for dysplastic BE and embedded it into the diagnostic work-up of patients with dysplastic BE. For the first time worldwide, the set-up of this panel incorporated quantification and improvement of pathologist performance using observer-related benchmark quality criteria and direct peer-feedback, making it qualitatively robust. Together with the demographic histopathologist factors deduced from the BOLERO study, incorporating assessments of 51 pathologists worldwide, these pathologist characteristics can serve as a basis for future guideline development. Due to the plethora of new techniques that are currently being validated as adjuncts to diagnostic work-up of histopathological BE biopsy assessment, in my opinion, future guidelines shall most certainly start to incorporate these additional techniques for patient stratification besides histopathological assessment. However, the basis for many of these techniques will remain the expert eye of the pathologist. I am sure that the combination of ingenious new technologies and traditional histopathology is most favorable to improve stratification of this patient group.

REFERENCES

1. Coleman HG, Xie SH, Lagergren J. The Epidemiology of Esophageal Adenocarcinoma. *Gastroenterology* 2018;154(2):390-405. doi: 10.1053/j.gastro.2017.07.046 [published Online First: 2017/08/07]

2. Thrift AP. The epidemic of oesophageal carcinoma: Where are we now? *Cancer Epidemiol* 2016;41:88-95. doi: 10.1016/j.canep.2016.01.013 [published Online First: 2016/02/07]

3. Haggitt RC, Tryzelaar J, Ellis FH, et al. Adenocarcinoma complicating columnar epithelium-lined (Barrett's) esophagus. *Am J Clin Pathol* 1978;70(1):1-5.

4. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. *The New England journal of medicine* 2011;365(15):1375-83. doi: 10.1056/NEJMoa1103042

5. Bhat S, Coleman HG, Yousef F, et al. Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. *Journal of the National Cancer Institute* 2011;103(13):1049-57. doi: 10.1093/jnci/djr203

6. Desai TK, Krishnan K, Samala N, et al. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut* 2012;61(7):970-6. doi: 10.1136/gutjnl-2011-300730 [published Online First: 2011/10/15]

7. Curvers WL, ten Kate FJ, Krishnadath KK, et al. Low-grade dysplasia in Barrett's esophagus: overdiagnosed and underestimated. *The American journal of gastroenterology* 2010;105(7):1523-30. doi: 10.1038/ajg.2010.171

8. Duits LC, Phoa KN, Curvers WL, et al. Barrett's oesophagus patients with low-grade dysplasia can be accurately risk-stratified after histological review by an expert pathology panel. *Gut* 2014 doi: 10.1136/gutjnl-2014-307278

9. Levine DS, Blount PL, Rudolph RE, et al. Safety of a systematic endoscopic biopsy protocol in patients with Barrett's esophagus. *The American journal of gastroenterology* 2000;95(5):1152-7. doi: 10.1111/j.1572-0241.2000.02002.x

10. Weusten B, Bisschops R, Coron E, et al. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2017;49(2):191-98. doi: 10.1055/s-0042-122140

11. Offerhaus GJ, Correa P, van Eeden S, et al. Report of an Amsterdam working group on Barrett esophagus. *Virchows Archiv : an international journal of pathology* 2003;443(5):602-8. doi: 10.1007/s00428-003-0906-z

12. Phoa KN, van Vilsteren FG, Weusten BL, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA : the journal of the American Medical Association* 2014;311(12):1209-17. doi: 10.1001/jama.2014.2511

13. van Sandick JW, van Lanschot JJ, Kuiken BW, et al. Impact of endoscopic biopsy surveillance of Barrett's oesophagus on pathological stage and clinical outcome of Barrett's carcinoma. *Gut* 1998;43(2):216-22.

14. Phoa KN, Pouw RE, Bisschops R, et al. Multimodality endoscopic eradication for neoplastic Barrett oesophagus: results of an European multicentre study (EURO-II). *Gut* 2015 doi: 10.1136/gutjnl-2015-309298

15. van Vilsteren FG, Pouw RE, Seewald S, et al. Stepwise radical endoscopic resection versus radiofrequency ablation for Barrett's oesophagus with high-grade dysplasia or early cancer: a multicentre randomised trial. *Gut* 2011;60(6):765-73. doi: 10.1136/gut.2010.229310

16. Pouw RE, Gondrie JJ, Sondermeijer CM, et al. Eradication of Barrett esophagus with early neoplasia by radiofrequency ablation, with or without endoscopic resection. *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract* 2008;12(10):1627-36; discussion 36-7. doi: 10.1007/s11605-008-0629-1

17. Duits LC, van der Wel MJ, Cotton CC, et al. Barrett's esophagus patients with confirmed and persistent low-grade dysplasia are at increased risk of neoplastic progression. *Gastroenterology* 2017 doi: 10.1053/j.gastro.2016.12.008 [published Online First: 21 Dec 2016]

18. Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics of gastrointestinal tract pathology: a feasibility study. *Human pathology* 2012;43(5):702-7. doi: 10.1016/j.humpath.2011.06.017

19. Snead DR, Tsang YW, Meskiri A, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016;68(7):1063-72. doi: 10.1111/his.12879 [published Online First: 2015/09/27]

20. Mukhopadhyay S, Feldman MD, Abels E, et al. Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). *The American journal of surgical pathology* 2018;42(1):39-52. doi: 10.1097/PAS.0000000000000948 [published Online First: 2017/09/30]

21. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* 2007;29(1):19-24.

22. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 2014;63(1):7-42. doi: 10.1136/gutjnl-2013-305372

23. Shaheen NJ, Falk GW, Iyer PG, et al. ACG Clinical Guideline: Diagnosis and Management of Barrett's Esophagus. *The American journal of gastroenterology* 2016;111(1):30-50. doi: 10.1038/ajg.2015.322

24. Maag-Darm-Leverartsen Nvv. Richtlijn Barrett-oesofagus. www.richtlijnendatabase.nl: IKNL, 2018:1-71.

25. Sharma P, Katzka DA, Gupta N, et al. Quality indicators for the management of Barrett's esophagus, dysplasia, and esophageal adenocarcinoma: international consensus recommendations from the American Gastroenterological Association Symposium. *Gastroenterology* 2015;149(6):1599-606. doi: 10.1053/j.gastro.2015.08.007 [published Online First: 2015/08/25]

26. Bergman J, de Groof AJ, Pech O, et al. An Interactive Web-Based Educational Tool Improves Detection and Delineation of Barrett's Esophagus-Related Neoplasia. *Gastroenterology* 2019;156(5):1299-308 e3. doi: 10.1053/j.gastro.2018.12.021 [published Online First: 2019/01/06]

27. Maley CC, Galipeau PC, Finley JC, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature genetics* 2006;38(4):468-73. doi: 10.1038/ng1768 [published Online First: 2006/03/28]

28. Galipeau PC, Li X, Blount PL, et al. NSAIDs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma. *PLoS medicine* 2007;4(2):e67. doi: 10.1371/journal.pmed.0040067

29. Li X, Galipeau PC, Sanchez CA, et al. Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression. *Cancer prevention research* 2008;1(6):413-23. doi: 10.1158/1940-6207.CAPR-08-0121 [published Online First: 2009/01/14]

30. Schulmann K, Sterian A, Berki A, et al. Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. *Oncogene* 2005;24(25):4138-48. doi: 10.1038/sj.onc.1208598 [published Online First: 2005/04/13]

31. Jin Z, Cheng Y, Gu W, et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. *Cancer research* 2009;69(10):4112-5. doi: 10.1158/0008-5472.CAN-09-0028 [published Online First: 2009/05/14]

32. Sato F, Jin Z, Schulmann K, et al. Three-tiered risk stratification model to predict progression in Barrett's esophagus using epigenetic and clinical features. *PloS one* 2008;3(4):e1890. doi: 10.1371/journal.pone.0001890 [published Online First: 2008/04/03]

33. Bird-Lieberman EL, Dunn JM, Coleman HG, et al. Population-based study reveals new risk-stratification biomarker panel for Barrett's esophagus. *Gastroenterology* 2012;143(4):927-35 e3. doi: 10.1053/j.gastro.2012.06.041

34. Timmer MR, Martinez P, Lau CT, et al. Derivation of genetic biomarkers for cancer risk stratification in Barrett's oesophagus: a prospective cohort study. *Gut* 2016;65(10):1602-10. doi: 10.1136/gutjnl-2015-309642

35. Duits LC, Klaver E, Bureo Gonzalez A, et al. The Amsterdam ReBus progressor cohort: identification of 165 Barrett's surveillance patients who progressed to early neoplasia and 723 nonprogressor patients. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2018 doi: 10.1093/dote/doy037 [published Online First: 2018/06/07]

36. Duits LC, Lao-Sirieix P, Wolf WA, et al. A biomarker panel predicts progression of Barrett's esophagus to esophageal adenocarcinoma. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2019;32(1) doi: 10.1093/dote/doy102 [published Online First: 2018/11/30]

37. Prichard JW, Davison JM, Campbell BB, et al. TissueCypher(): A systems biology approach to anatomic pathology. *J Pathol Inform* 2015;6:48. doi: 10.4103/2153-3539.163987 [published Online First: 2015/10/03]

38. Ross-Innes CS, Chettouh H, Achilleos A, et al. Risk stratification of Barrett's oesophagus using a non-endoscopic sampling method coupled with a biomarker panel: a cohort study. *Lancet Gastroenterol* 2017;2(1):23-31. doi: 10.1016/S2468-1253(16)30118-2

39. Gross SA, Smith MS, Kaul V, et al. Increased detection of Barrett's esophagus and esophageal dysplasia with adjunctive use of wide-area transepithelial sample with three-dimensional computer-assisted analysis (WATS). *United European gastroenterology journal* 2018;6(4):529-35. doi: 10.1177/2050640617746298 [published Online First: 2018/06/09]

40. Smith MS, Ikonomi E, Bhuta R, et al. Wide-area transepithelial sampling with computer-assisted 3-dimensional analysis (WATS) markedly improves detection of esophageal dysplasia and Barrett's esophagus: analysis from a prospective multicenter community-based study. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2019;32(3) doi: 10.1093/dote/doy099 [published Online First: 2018/12/13]

41. Vennalaganti PR, Kaul V, Wang KK, et al. Increased detection of Barrett's esophagus-associated neoplasia using wide-area trans-epithelial sampling: a multicenter, prospective, randomized trial. *Gastrointestinal endoscopy* 2018;87(2):348-55. doi: 10.1016/j.gie.2017.07.039 [published Online First: 2017/08/02]

42. Mosko JD, Pleskow D. Evaluation of NinePoint Medical's Nvision VLE device for gastrointestinal applications. *Expert Rev Med Devices* 2017;14(7):495-503. doi: 10.1080/17434440.2017.1343142 [published Online First: 2017/06/16]

43. Swager A, Boerwinkel DF, de Bruin DM, et al. Volumetric laser endomicroscopy in Barrett's esophagus: a feasibility study on histological correlation. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2016;29(6):505-12. doi: 10.1111/dote.12371

44. Swager AF, Tearney GJ, Leggett CL, et al. Identification of volumetric laser endomicroscopy features predictive for early neoplasia in Barrett's esophagus using high-quality histological correlation. *Gastrointestinal endoscopy* 2017;85(5):918-26 e7. doi: 10.1016/j.gie.2016.09.012 [published Online First: 2016/09/24]

45. van der Sommen F, Klomp SR, Swager AF, et al. Predictive features for early cancer detection in Barrett's esophagus using Volumetric Laser Endomicroscopy. *Comput Med Imaging Graph* 2018;67:9-20. doi: 10.1016/j.compmedimag.2018.02.007 [published Online First: 2018/04/24]

46. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA : the journal of the American Medical Association* 2017;318(22):2199-210. doi: 10.1001/jama.2017.14585 [published Online First: 2017/12/14]