Quality assessment and improvements in pathology practice

Chantal C.H.J. Kuijpers

Quality assessment and improvements in pathology practice

ISBN: 978-90-393-6506-9

Lay-out en cover design: Guus Gijben, Proefschrift-AIO, Utrecht
Printing: DPP, Houten

Funding was received by the Dutch Cancer Society (KWF) for the studies presented in Chapters 3 and 4.

Printing of this thesis was financially supported by Symbiant Pathology Expert Centre and the pathology department of the University Medical Centre Utrecht.

Quality assessment and improvements in pathology practice

Analyse en verbetering van de kwaliteit binnen de pathologie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op donderdag 31 maart 2016
des middags te 4.15 uur

door

Chantal Catharina Henrica Johanna Kuijpers
geboren op 16 mei 1986 te Oosterhout

Promotor:	Prof.dr. P.J. van Diest
Copromotoren:	Dr. N.M. Jiwa
	Dr. S.M. Willems

Table of contents

General introduction

Every patient has the right to receive optimal quality health care. To ensure this, effort must be made at every stage of the health care process by every person involved, including general practitioners, clinicians, nurses, secretaries, hospital directors, etc. With regard to pathology practice, a small (interpretational) difference can have major impact for the patient, because prognosis and treatment selection are often based on the pathology report. Quality assurance and patient safety should be top priority within the pathology department.

Patient safety

Patient safety is commonly defined as freedom from accidental injury due to medical care or medical errors [1], and is determined by the frequency and seriousness of errors. Error is defined as the failure of a planned action to be completed as intended or the use of a faulty plan to achieve an aim [1]. A timely detected error or one that does not cause harm is called a near miss or close call. Errors resulting in patient injury are referred to as preventable adverse events. An adverse event is defined as injury that results from a medical intervention and is responsible for harm to the patient [2]. Adverse events may not only be caused by medical errors, but by other, not preventable factors as well, such as complications.

Quality improvement, error reduction and patient safety in medicine have increasingly received attention since the publication of the Institute of Medicine report 'To err is human; building a safer health system' in 1999 [1], which stated that the incidence of errors in medicine is high. An important conclusion of this report was that the majority of medical errors result from faulty systems, processes, and conditions, instead of from individuals' mistakes. The reports' results changed the perspective from the 'individuals approach' into the 'systems approach'. The systems approach focuses on finding situations or factors that are prone to result in human error and changing the underlying systems.

Individual and system-related errors

Traditionally, errors were mainly believed to result from persons' unsafe acts, and interventions, such as education, poster campaigns, and blaming, were mainly addressed to the individual. However, according to the Swiss cheese model, or cumulative act effect model, by Reason (1990), nearly all errors involve a combination of latent (system) and active (individual) failures. The model describes four defensive layers – organizational influences, unsafe

supervision, preconditions for unsafe acts, and unsafe acts – organized as slices of Swiss cheese (Figure 1). The holes represent possible failures. Failure in one layer would normally not lead to an error, because of the protection by the other subsequent layers. An error will then only occur when subsequent defensive layers fail together.

As most errors are due to shortcomings in the system or organisation, blaming the individual will not be effective to prevent future errors to occur. Instead, a patient safety culture, where individuals can report their errors blame-free, is essential to determine the actual number of errors, to learn from them, and to improve systems and processes. Factors that are likely to result in human errors should be identified in order to change the underlying systems to minimize errors.

Diagnostic errors in medicine

In response to the Institute of Medicine report, a large study on adverse events was performed by NIVEL and EMGO in The Netherlands, and published in 2007 [3]. Extrapolation of the study results indicated that a total of 30,000 possibly preventable adverse events and 1,735 possibly preventable deaths had occurred in The Netherlands in 2004. Although the majority of adverse events were observed in surgical specialties, the authors mentioned the diagnostic process as an important focus area, because of the high preventability of the adverse events caused by diagnostic errors.

Diagnostic error is usually defined as a diagnosis that is missed, wrong or delayed as concluded from more definitive information (e.g. autopsy studies) [4]. A systematic review of 42 autopsy studies reported a median major discrepancy rate between clinical and autopsy diagnoses of 23.5% [5]. Although diagnostic errors are less common than other medical errors like surgical errors and drug related errors, they are more likely to result in injury and are the leading cause of malpractice claims [6-8]. Furthermore, diagnostic errors are more likely to be preventable than other types of medical errors [6,8]. These reasons make the problem of diagnostic errors an important issue to address.

Initially, the topic of diagnostic errors has received less attention than other patient safety matters. This lack of attention is mostly due to the challenges in measuring diagnostic errors and the difficulties of finding solutions for diagnostic errors, as they were thought to be mainly cognitive errors and not

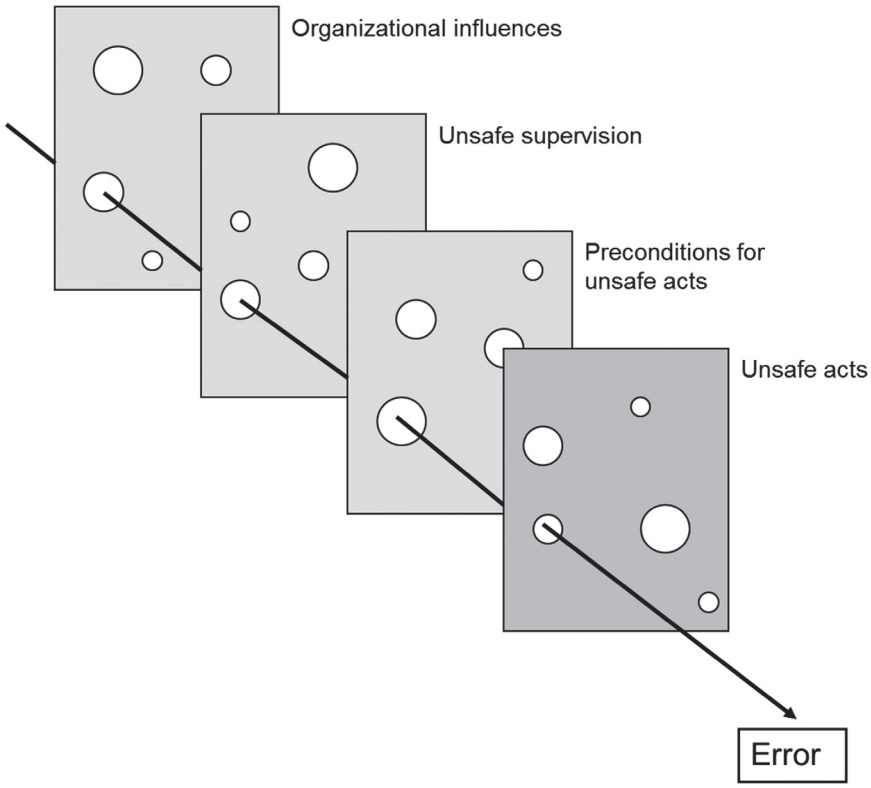


Figure 1. The Swiss cheese model by Reason (1990). An error will most likely occur when subsequent defensive layers fail together. Failure in one defensive layer would normally not lead to an error.

so much system errors [9,10]. However, in order to reduce diagnostic errors, diagnostic accuracy must be considered as a system property [11]. Systems-based solutions to avoid diagnostic errors include cognitive debiasing, clinical guidelines, computerized decision-support systems, second opinions and availability of expertise [12,13]. Moreover, regular feedback to the clinicians on their diagnostic performance is essential to improve future diagnostics. A recent review study showed that publications involving research on diagnostic errors increased in the past few years, and the most frequently studied areas were second review methods, technology-based systems and structured processes [14].

Diagnostic challenges in pathology practice

An adequate diagnosis and pathology report is of paramount importance for prognostication and treatment decisions. To this end, pathology reports should be accurate, timely, and complete [15]. In this thesis we mainly focus on the accurateness. Problems can occur at each phase of the diagnostic process (i.e. the pre-analytical, analytical, and post-analytical phase).

The pre-analytical phase includes specimen acquisition, arrival at the pathology department, and specimen handling. Identification or labeling errors can lead to specimen mix up between patients [16,17]. Furthermore, variation in pre-analytical parameters, such as tissue fixation and decalcification, can influence tissue quality and results of immunohistochemical (IHC) stainings and molecular tests [18,19].

This thesis will mainly focus on problems in the analytical phase that may lead to an inaccurate diagnosis and pathology report. The analytical phase comprises the process from grossing until the formation of a diagnosis, and includes histology (or cytology), IHC, molecular testing, and microscopic interpretation.

- During grossing, gross descriptions are very important for accurate orientation and microscopic examination, and the use of standardized formats is common practice in many institutions. Due to the increasing workload of pathologists, pathologists' assistants (PAs), under supervision by a pathologist, play an increasingly important role in grossing, taking over some routine tasks from pathologists, such as dissecting a selection of specimens and harvesting lymph nodes (LNs).
- Diagnostic pathology may suffer from interobserver variation, as microscopic examination of histology, cytology, and IHC is not always as objective, and may therefore lead to interpretational errors. Double reading or peer review is commonly used as a measure of diagnostic accuracy, and is a potentially valuable tool for reducing diagnostic errors and thereby improving the quality of patient care. Different methods of peer review are being used, including intradepartmental review of a percentage of randomly selected cases, focused review of specific organ systems or diagnoses, interdepartmental conferences, intradepartmental review of material before release to other institutions, and in-house review of outside material [15]. Besides variation in main diagnosis, interobserver variation also exists in evaluating critical parameters, such as histological grade, subtype, and completeness of excision [20].

- Targeted therapies and personalized medicine are becoming increasingly important in cancer care, and adequate selection of patients eligible for targeted treatment is essential. Examples are selecting breast cancer patients that are likely to be responsive to HER-2 targeted therapy (trastuzumab and lapatinib) and lung cancer patients that are likely to be responsive to drugs targeting EGFR (gefitinib and erlotinib) or ALK (crizotinib). Accurate determination of tumor cell percentage, choosing the right test, and accurate interpretation of the test results are essential.

The post-analytical phase starts with dictation of macroscopic and microscopic examination and the final diagnosis, and includes report correction, verification, and delivery. As the pathology report is often the basis for treatment decisions, especially in oncology, one of the major problems is that of incomplete reports. This can be significantly improved by the use of standardized checklists or synoptic reports [21,22]. In The Netherlands, synoptic reporting was first introduced in 2008 by PALGA (the Dutch Pathology Registry) for colorectal cancer and breast cancer resections and is from then onwards applied to an increasing number of malignancies. Although no synoptic report is available for autopsies, these reports should be complete and comprehensive as well, in order to give proper feedback to clinicians, as the autopsy is the gold standard for the evaluation of medical practice.

Aim of this thesis

The aim of this thesis was to assess and improve the quality of several important aspects of pathology practice, and thereby improve quality of health care and patient safety. To this end, several diagnostic processes in pathology practice were assessed, and the added value of multiple interventions or strategies, with focus on oncology, was investigated. We addressed several problems that may lead to an inaccurate diagnosis and pathology report. Firstly, we addressed the problem of high workload of pathologists. The second very common problem we addressed was interobserver variation. Finally, we determined the role of autopsies in quality improvement of health care.

Outline of this thesis

The p**ro**blem of high workloads of pathologists is addressed in **Chapter 2**. Symbiant has been training PAs to take over certain routine tasks from pathologists to decrease workload. We assessed whether PAs contribute to improved quality of care, by assessing one of their tasks, namely harvesting LNs in colorectal cancer resection specimens. In colorectal cancer, LN metastasis is an important prognostic parameter and an indication for adjuvant chemotherapy. The Dutch guideline states that at least 10 LNs should be examined in order to reliably stage the tumor [23]. Inadequate LN harvest (< 10) may result in indeterminate staging and overtreatment, as these patients are considered high-risk and are eligible for adjuvant chemotherapy as well.

Interpretation of pathology specimens (histology, cytology, and IHC) is unfortunately not always as objective, leading to interobserver variation, a problem which we address in **Chapters 3-7.**

Chapters 3 and 4 focus on variability in daily practice between and within Dutch pathology laboratories with regard to grading of (pre-)malignant lesions. Using data from PALGA, we investigated the interlaboratory and intralaboratory variability in the grading of dysplasia in colorectal adenomas (**Chapter 3**) and in the grading of colorectal adenocarcinomas (**Chapter 4**).

Chapters 5 and 6 report the added value of two intradepartmental double reading strategies by determining (degree of) concordance with the initial diagnoses. **Chapter 5** focuses on second review of histopathology specimens prior to discussion at a multidisciplinary meeting. In **Chapter 6**, the added value of secondary slide review of clinical cytology specimens by specialized cytopathologists is reported.

Also evaluation of IHC staining may suffer from interobserver variation. One example is the evaluation of HER-2 overexpression in breast cancer, which identifies patients likely to benefit from treatment with HER-2 targeted therapies. The ASCO/CAP guidelines [24,25] stipulate that trastuzumab therapy is only applicable for patients who strongly overexpress the HER-2 protein (3+) and those who present with equivocal HER-2 protein levels (2+) with confirmed gene amplification. Normally, IHC negative (0-1+) and positive (3+) cases are not reflex tested. In **Chapter 7**, we assessed the added value of routinely co-testing every invasive breast cancer case by IHC and the more

quantitative PCR-based multiplex ligation-dependent probe amplification (MLPA) technique.

In **Chapter 8**, the role of autopsies in quality improvement of health care is investigated by determining the frequency of discrepancies between clinical diagnoses and autopsy diagnoses according to the modified Goldman classification.

Chapter 9 includes a general discussion of the main findings of this thesis, and we conclude with a summary in Dutch in **Chapter 10**.

References

1. Kohn LT, Corrigan JM, Donaldson MS. To Err is Human; Building a Safer Health System. 1999. Institute of Medicine, National Academy Press, Washington DC
2. Garrouste-Orgeas M, Philippart F, Bruel C, Max A, Lau N, Misset B. (2012) Overview of medical errors and adverse events. Ann Intensive Care 2(1):2.
3. de Bruijne MC, Zegers M, Hoonhout LHF, Wagner C. (2007) Onbedoelde schade in Nederlandse ziekenhuizen. Dossieronderzoek van ziekenhuisopnames in 2004. Available from: http://www.nivel.nl/sites/default/files/bestanden/onbedoelde-schade-in-nederlandse-ziekenhuizen-2007.pdf.
4. Newman-Toker DE, Pronovost PJ. (2009) Diagnostic errors--the next frontier for patient safety. JAMA 301(10): 1060-2.
5. Shojania KG, Burton EC, McDonald KM, Goldman L. (2003) Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. JAMA 289(21): 2849-56.
6. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. (1991) The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med 324(6): 377-84.
7. Chandra A, Nundy S, Seabury SA. (2005) The growth of physician medical malpractice payments: evidence from the National Practitioner Data Bank. Health Aff (Millwood) Suppl Web Exclusives: p. W5-240-9.
8. Landro L (2013) The Biggest Mistake Doctors Make. Misdiagnoses are harmful and costly. But they're often preventable. The Wall Street Journal.
9. Wachter RM. (2010) Why diagnostic errors don't get any respect--and what can be done about them. Health Aff (Millwood) 29(9): 1605-10.
10. Graber M (2005) Diagnostic errors in medicine: a case of neglect. Jt Comm J Qual Patient Saf 31(2): 106-13.
11. Schiff GD, Kim S, Abrams R, Cosby K, Lambert B, Elstein AS, et al. (2005) Diagnosing Diagnosis Errors: Lessons from a Multi-institutional Collaborative Project. Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology).
12. Croskerry P. (2003) The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med 78(8): 775-80.
13. Graber M, Gordon R, Franklin N. (2002) Reducing diagnostic errors in medicine: what's the goal? Acad Med 77(10): 981-92.
14. McDonald KM, Matesic B, Contopoulos-Ioannidis DG, Lonhart J, Schmidt E, Pineda N, et al. (2013) Patient safety strategies targeted at diagnostic errors: a systematic review. Ann Intern Med 158(5 Pt 2): 381-9.
15. Nakhleh RE. (2006) What is quality in surgical pathology? J Clin Pathol 59(7) 669-72.
16. Nakhleh RE, Idowu MO, Souers RJ, Meier FA, Bekeris LG. (2011) Mislabeling of cases, specimens, blocks, and slides: a college of american pathologists study of 136 institutions. Arch Pathol Lab Med 135(8): 969-74.
17. Brown RW, Speranza VD, Alvarez JO, Eisen RN, Frishberg DP, Rosai J, et al. (2015) Uniform Labeling of Blocks and Slides in Surgical Pathology: Guideline From the College of American Pathologists Pathology and Laboratory Quality Center and the National Society for Histotechnology. Arch Pathol Lab Med 139(12): 1515-24.

18. College of American Pathologists1, Valenstein PN, Raab SS, Walsh MK. (2006) Identification errors involving clinical laboratories: a College of American Pathologists Q-Probes study of patient and specimen identification errors at 120 institutions. Arch Pathol Lab Med 130(8): 1106-13.
19. Darvishian F, Singh B, Krauter S, Chiriboga L, Gangi MD, Melamed J. (2011) Impact of decalcification on receptor status in breast cancer. Breast J 17(6): 689-91.
20. Turner JK, Williams GT, Morgan M, Wright M, Dolwani S. (2013) Interobserver agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists in Wales. Histopathology 62(6): 916-24.
21. Karim RZ, van den Berg KS, Colman MH, McCarthy SW, Thompson JF, Scolyer RA. (2008) The advantage of using a synoptic pathology report format for cutaneous melanoma. Histopathology 52(2): 130-8.
22. Messenger DE, McLeod RS, Kirsch R. (2011) What impact has the introduction of a synoptic report for rectal cancer had on reporting outcomes for specialist gastrointestinal and nongastrointestinal pathologists? Arch Pathol Lab Med 135(11): 1471-5.
23. Dutch guideline for colorectal cancer. Available from: http://www.oncoline.nl/index.php?pagina=/richtlijn/item/pagina.php&id=37094&richtlijn_id=933.
24. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. (2007) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med 131(1): 18-43.
25. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. (2014) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Arch Pathol Lab Med 138(2): 241-56.

Better retrieval of lymph nodes in colorectal resection specimens by pathologists' assistants

Chantal C.H.J. Kuijpers
Henk-Jan van Slooten
Hermien Schreurs
Gisela R.H.M. Moormann
Mohammad A. Abtahi
André Slappendel
Vincent Cliteur
Paul J. van Diest
Mehdi Jiwa

Published in Journal of Clinical Pathology 2013; 66(1): 18-23

Abstract

Introduction

Errors in surgical pathology are partly due to the increasing workload of pathologists. To reduce this workload, "pathologists' assistants" (PAs) have been trained to take over some of the pathologists' recurrent tasks. One of these tasks is the precise examination of ≥ 10 lymph nodes (LNs), which is of paramount importance to reduce the risk of understaging of colorectal cancer patients. This study evaluates the role of PAs in harvesting LNs in colorectal resection specimens and by doing so in improving patient safety.

Methods

LN harvest was retrospectively reviewed in 649 pathology reports on colorectal resection specimens collected in two Dutch hospitals from 2008 until 2011.

Results

PAs sampled ≥ 10 LNs in significantly more cases than pathologists did (83.2% vs. 60.9% in hospital A and 79.2% vs. 67.6% in hospital B) and recovered on average significantly more LNs than pathologists did (18.5 vs. 12.2 in hospital A and 16.6 vs. 13.2 in hospital B). PAs harvested a significantly higher percentage of LNs < 5 mm than pathologists did (64.2% vs. 53.7%). The percentages of colon cancer patients eligible for adjuvant chemotherapy due to inadequate LN sampling alone were significantly higher for cases dissected by pathologists than for those dissected by PAs (17.3% vs. 1.1% in hospital A and 13.1% vs. 3.4% in hospital B).

Conclusion

PAs contribute to patient safety since they recover more and, in particular, smaller LNs from colorectal resection specimens than pathologists do. Moreover, they help to reduce costs and morbidity by reducing the number of patients eligible for adjuvant chemotherapy due to inadequate LN sampling alone.

Introduction

In the past decade, patient safety and error reduction in surgical pathology have increasingly received attention. Several authorities developed quality standards and guidelines that contribute to ensuring patient safety. Regarding the pathologist, several factors contribute to errors, including distractions and workload increase [1]. The authors of a Turkish study proposed to decrease pathologists' workload by deploying "pathologists' assistants" (PAs) [2].

There is a shortage of pathologists in the Netherlands. Therefore, starting from 2009, our institution has been training PAs to take over certain routine activities to save pathologists valuable time, which they can use instead to focus on microscopy and on interdisciplinary meetings. Based on our own experience and on results of other studies, we expect that adequately trained PAs can contribute to patient safety [3,4].

For the scope of this study, we focussed on lymph node (LN) harvest from colorectal resection specimens, which is a labor-intensive and time-consuming activity. Due to the nature of their job, PAs can usually spend more time focussing on harvesting LNs than pathologists can. This is the main reason why we expect PAs, on average, to harvest more LNs per specimen than pathologists do.

Metastases in LNs form an important prognostic factor in determining eligibility for adjuvant chemotherapy [5], as 5-year survival considerably decreases when $\geq$ 1 LNs are malignant. Also, the number of harvested LNs has been shown to be closely related to recurrence and survival [6-11]. In a study on 480 node-negative patients 5-year survival increased from 51% when < 10 LNs were sampled to 69% when 10-19 LNs were harvested and rose to 71% when > 19 LNs were examined respectively [12].

Nonetheless, 20-40% of patients with presumed negative LN status eventually die as a result of their cancer [13]. A significant number of particularly small (metastatic) LNs might be missed during grossing, as 45-78% of metastatic LNs measure < 5 mm in diameter [14-21]. Precise examination and retrieval of as many LNs as possible is therefore of paramount importance for reducing the risk of understaging.

The recommended minimum number of LNs varied from 6 to 17 LNs in previous studies [8,9,13,22-30]. The Dutch guideline [31] recommends a minimum of 10 LNs, whereas several international guidelines recommend a minimum of 12 LNs [32,33]. According to the seventh edition of the American AJCC Cancer Staging Manual, a minimum of 10-14 LNs must be examined [34]. In The Netherlands, in 2007 only 65% of all stage I and stage II cancer cases had adequate LN sampling (≥ 10 LNs evaluated) [35]. The Dutch Surgical Colorectal Audit (DSCA) of 2009 showed that in 73% of colon cancer cases and in 58% of rectal cancer cases ≥ 10 LNs were evaluated. These percentages rose to 83% and 68% respectively in 2011 [36]. At our institution, we also aim at retrieving more LNs per specimen. Previous studies show percentages of ≥ 12 LNs per rectal cancer case varying between 5-64% after neo-adjuvant chemo-radiotherapy and between 55-88% without prior chemo-radiotherapy [37-43].

Colon cancer patients (≥ T2 stage) receive adjuvant chemotherapy if they present with ≥ 1 metastatic LNs or if they are at a high risk of developing recurrences. Patients at a high risk have either T4 tumors, poor differentiation, bowel obstruction, lymphovascular invasion, perineural invasion, bowel wall perforation or an insufficient number of sampled LNs.

The purpose of this study was to evaluate the role of PAs in the improvement of patient safety by comparing their LN harvest results from colorectal resection specimens with those of pathologists. Moreover, we evaluated whether PA deployment could theoretically result in a reduced proportion of colon cancer patients who were eligible for adjuvant chemotherapy due to inadequate LN sampling alone.

Methods

Patients

This study included colorectal cancer patients in two hospitals (A and B), whose pathology services were provided by Symbiant. We retrospectively reviewed 557 pathology reports of patients who were surgically treated for colorectal cancer between January 2008 and November 2011 (Table 1). The surgical technique did not change during the study period and was the same for every surgeon. In each hospital, 10 pathologists and 7 PAs were responsible for harvesting LNs.

Pathologists' assistants

The PAs in this study were officially trained by Symbiant and the *Centrum Bioscience en Diagnostiek, Hogeschool Leiden*. The PAs started their training by observing a pathologist processing resection specimens. They then learned to dissect specimens and sample LNs themselves, under direct supervision of a pathologist. After proving their skills to the pathologist, PAs began working independently, with the pathologist acting as a back-up. In 2009, PAs had taken over from pathologists in 19.8% of cases. By 2010, they sampled LNs in 53.2% of the cases and by 2011, 66.7% of all specimens were grossed by PAs.

Table 1. Clinicopathological features of 557 colorectal cancer patients.

	Hospital A Pathologist N=115	PA N=125	Hospital B Pathologist N=139	PA N=178	Total N=557
Age					
Average (range)	69 (43-90)	70 (35-90)	71 (42-91)	71 (41-92)	70 (35-92)
Sex					
Male	56 (48.7%)	72 (57.6%)	67 (48.2%)	109 (61.2%)	304 (54.6%)
Female	59 (51.3%)	53 (42.4%)	72 (51.8%)	69 (38.8%)	253 (45.4%)
Tumor site					
Right	50 (43.5%)	63 (50.4%)	42 (30.2%)	59 (33.1%)	214 (38.4%)
Transverse colon	3 (2.6%)	5 (4.0%)	6 (4.3%)	6 (3.4%)	20 (3.6%)
Left	27 (23.5%)	18 (14.4%)	50 (36.0%)	52 (29.2%)	147 (26.4%)
Rectum	34 (29.6%)	38 (30.4%)	39 (28.1%)	59 (33.1%)	170 (30.5%)
*Neo-adj therapy**	*13 (38.2%)*	*13 (34.2%)*	*NA*	*NA*	*NA*
No neo-adj therapy	*19 (55.9%)*	*25 (65.8%)*	*NA*	*NA*	*NA*
Not mentioned	*2 (5.9%)*	-	*NA*	*NA*	*NA*
Subtotal colectomy	1 (0.9%)	1 (0.8%)	2 (1.4%)	1 (0.6%)	5 (0.9%)
Not mentioned	-	-	-	1 (0.6%)	1 (0.2%)
Diagnosis					
Adenoma	-	1 (0.8%)	2 (1.4%)	1 (0.6%)	4 (0.7%)
Adenocarcinoma	112 (97.4%)	124 (99.2%)	135 (97.1%)	176 (98.9%)	547 (98.2%)
Squamous cell carcinoma	1 (0.9%)	-	-	-	1 (0.2%)
Adenosquamous carcinoma	1 (0.9%)	-	-	-	1 (0.2%)
Undifferentiated carcinoma	1 (0.9%)	-	-	-	1 (0.2%)
Metastasis adenocarcinoma	-	-	2 (1.4%)	1 (0.6%)	3 (0.5%)
Tumor stage					
Tis	-	2 (1.6%)	1 (0.7%)	-	3 (0.5%)
T1	3 (2.6%)	12 (9.6%)	6 (4.3%)	8 (4.5%)	29 (5.2%)
T2	23 (20.0%)	20 (16.0%)	28 (20.1%)	43 (24.2%)	114 (20.5%)
T3	73 (63.5%)	72 (57.6%)	71 (51.5%)	90 (50.6%)	306 (54.9%)
T4	16 (13.9%)	19 (15.2%)	32 (23.0%)	36 (20.2%)	103 (18.5%)
Not mentioned	-	-	1 (0.7%)	1 (0.6%)	2 (0.4%)

* Short-course radiotherapy (5x5 Gy) was not considered neo-adjuvant therapy.
NA = not available

Resection specimen processing

Colon and rectal resection specimens were randomly assigned to a pathologist or a PA, based on their schedules, and processed routinely. After fixation in neutral buffered formaldehyde, the adipose tissue was sectioned at 1-2 mm intervals to gather the LNs, and the number of LNs per cassette was accurately counted and registered.

Statistical analysis

The independent samples t-test or ANOVA was performed to compare the average number of harvested LNs per specimen between the subgroups. Chi-squared analysis was used to compare the percentages of cases with ≥ 10 LNs evaluated, the percentages of LNs < 5 mm dissected by PAs and pathologists and the percentages of patients eligible for adjuvant chemotherapy due to inadequate LN sampling alone.

We performed logistic regression (odds ratio, 95% CI and p-value) for univariable and multivariable analysis. Factors that contributed significantly to adequate LN sampling in the univariable analysis were considered for multivariable analysis. P-values < 0.05 were considered as statistically significant. All p-values reported are two-sided.

Results

LN harvest by pathologists and PAs

Table 2 shows the percentages of colorectal cancer cases which were adequately sampled by pathologists and PAs as well as the average numbers of harvested LNs per specimen. Compared with pathologists, PAs more often succeeded in sampling ≥ 10 LNs per specimen (83.2% vs. 60.9% in hospital A, $p < 0.0001$, and 79.2% vs. 67.6% in hospital B, $p = 0.019$). These differences remained statistically significant ($p < 0.0001$ and $p = 0.023$ for hospitals A and B, respectively) after correction for T-stage, tumor site, tumor diameter and specimen length.

Not only did PAs more often harvest ≥ 10 LNs per specimen than pathologists did, their average LN count per specimen also proved higher (18.5 vs. 12.2 and 16.6 vs. 13.2 in hospitals A and B, respectively with $p < 0.0001$ for both hospitals).

Table 2. Percentage of adequately sampled cases and average numbers of LNs per specimen recovered by pathologists and PAs in hospital A and B.

	n	≥ 10 LNs	p-value	Average # LNs	p-value
Hospital A					
Pathologist	115	60.9%	< 0.0001	12.2	< 0.0001
PA	125	83.2%		18.5	
Hospital B					
Pathologist	139	67.6%	0.019	13.2	< 0. 0001
PA	178	79.2%		16.6	

Table 3. Nodal yield according to tumor site for hospital A and hospital B.

	n	≥ 10 LNs	p-value	Average # LNs	p-value
Hospital A					
Right colon	113	86.7%	< 0.0001	18.4	< 0.0001
Left colon	45	68.9%		14.5	
Rectum	72	51.4%		10.6	
Hospital B					
Right colon	101	89.1%	< 0.0001	18.2	< 0. 0001
Left colon	102	70.6%		13.5	
Rectum	98	61.2%		13.4	

Table 4. Percentage of adequately sampled rectal resection specimens by pathologists and PAs in two hospitals.

	n	≥ 10 LNs	p-value
Hospital A			
Pathologist	34	38.2%	0.035
PA	38	63.2%	
Hospital B			
Pathologist	39	62.5%	0.959
PA	59	61.0%	

The role of tumor location in LN harvest

Table 3 shows the relation between tumor site and LN sampling. The percentage of cases in which ≥ 10 LNs were sampled was significantly lower in rectal resection specimens than in right colon resection specimens ($p < 0.0001$ for both hospitals).

Also, the average number of LNs per specimen harvested in rectal resection specimens was significantly lower than in resection specimens from the right colon ($p < 0.0001$ for both hospitals) and in hospital A it was also lower than in resection specimens from the left colon ($p = 0.003$).

LN harvest in the rectal cancer subgroup

As rectal cancer cases especially seem to be among the most difficult to sample adequately, we evaluated this subgroup separately (Table 4). In hospital A, PAs retrieved ≥ 10 LN in 63.2% of cases and pathologists harvested ≥ 10 LNs in 38.2% of cases ($p = 0.035$), whereas in hospital B the percentages achieved by PAs (61.0%) were almost equal to those of pathologists (62.5%).

Neo-adjuvant chemotherapy had no significant effect on the percentage of adequately sampled cases (46.7% vs. 61.0% with and without neo-adjuvant chemotherapy respectively, $p = 0.197$).

Factors contributing to adequate LN harvest

Univariable analyses show that the following factors contribute significantly to adequate LN harvest (based on crude odds ratios): tumor site, tumor diameter, T-stage and pathologist vs. PA. Multivariable analysis showed that tumor site, tumor diameter and pathologist vs. PA still contributed to the harvest of ≥ 10 LNs (based on adjusted odds ratios). After correction for other factors, PA was associated with a greater effect on sampling ≥ 10 LNs (OR 2.671; 95% CI 1.726 – 4.135) than was shown in univariable analysis (OR 2.318; 95% CI 1.578 – 3.406) (Table 5).

Recovery of LNs < 5 mm

Hospital A counted 113 patients with ≥ 1 metastatic LNs. Comparison of the maximum diameters of all LNs from these patients revealed that PAs had recovered a significantly higher percentage of LNs < 5 mm than pathologists had (Table 6).

Table 5. Factors contributing to the percentage of cases with ≥ 10 harvested LNs.

	UV analysis Odds ratio	95% CI	p-value	MV analysis Odds ratio	95% CI	p-value
Age	0.987	0.091 – 1.004	0.131			
Sex	1.059	0.743 – 1.509	0.752			
Tumor site						
Right	1		< 0.0001	1		< 0.0001
Transverse colon	0.459	0.157 – 1.339		0.501	0.161 – 1.560	
Left	0.386	0.235 – 0.634 *		0.456	0.266 – 0.783 *	
Rectum	0.159	0.123 – 0.315 *		0.287	0.169 – 0.487 *	
Subtotal colectomy Δ	-	-		-	-	
Tumor diameter	1.391	1.245 – 1.554 *	< 0.0001	1.286	1.138 – 1.453 *	< 0.0001
Specimens length	1.014	0.998 – 1.030	0.096	1.001	0.983 – 1.019	0.944
Tumor stage						
Tis	0.305	0.041 – 2.276	< 0.0001	0.207	0.021 – 2.003	0.085
T1	0.343	0.153 – 0.770 *		0.445	0.174 – 1.141	
T2	0.479	0.270 – 0.850 *		0.863	0.438 – 1.701	
T3	1.287	0.768 – 2.157		1.214	0.681 – 2.162	
T4	1			1		
Hospital	1.047	0.735 – 1.491	0.800			
Pathologist vs. PA	2.318	1.578 – 3.406 *	< 0.0001	2.671	1.726 – 4.135 *	< 0.0001

* factors that significantly contribute to the chance of harvesting ≥ 10 LNs.
Δ odds ratio and 95% CI not mentioned, as this group comprises only five patients.

Table 6. Percentages of small LNs dissected by pathologists and PAs in hospital A.

	n	LNs < 5 mm	p-value
Pathologist	633	340 (53.7%)	< 0.0001
PA	822	566 (64.2%)	

Table 7. Percentages of colon cancer patients eligible for adjuvant chemotherapy solely because of inadequate LN sampling.

	Colon cancer patients	< 10 LNs harvested	NO (LN) mets and T2-3	Eligible for adjuvant chemotherapy	p-value
Hospital A					
Pathologist	81	24	14	17.3%	< 0.0001
PA	87	7	1	1.1%	
Hospital B					
Pathologist	99	30	13	13.1%	0.007
PA	119	14	4	3.4%	

Eligibility for adjuvant chemotherapy due to inadequate LN sampling

Table 7 shows the percentages of colon cancer patients who could have been eligible for adjuvant chemotherapy based solely on inadequate LN harvest. In the subgroup of patients dissected by PAs, significantly fewer patients were eligible for adjuvant chemotherapy due to inadequate LN harvest alone than in the subgroup of patients dissected by pathologists (1.1% vs. 17.3% in hospital A, $p < 0.0001$, and 3.4% vs. 13.1% in hospital B, $p = 0.007$).

Discussion

The aim of this study was to determine the value of PAs for patient safety by comparing the performance of PAs and pathologists in harvesting LNs from colorectal resection specimens. Sampling LNs is time-consuming and in particularly difficult to achieve in rectal cancer.

Our study showed a significant improvement in LN harvest after deployment of PAs in two hospitals. We evaluated the average number of LNs per specimen and the percentage of cases with ≥ 10 LNs, which in the Netherlands is considered to be the minimum number needed to assess prognosis and the indication for adjuvant chemotherapy. Multivariable analysis confirmed that the PAs were responsible for improved LN yield. Results were comparable when the AJCC threshold of 12 LNs was used (data not shown). Reasons for these results were not researched but are likely to be varied and to include practical circumstances. The main reason for the use of PAs is to optimise the use of pathologists' time [44]. PAs are generally dedicated to this type of task, whereas pathologists' increasing workloads, can lead to time constraints. Due to the nature of their job, PAs can usually spend more time, with fewer distractions, searching for LNs than pathologists can [44].

Results of the 2011 DSCA showed that ≥ 10 LNs had been evaluated in 83% of colon cancer cases included in the audit. The PAs working in the two hospitals in which this study was performed found ≥ 10 LNs in 92.0% and 88.1% of colon cases, whereas the pathologists found ≥ 10 LNs in 70.4% and 70.0% of colon cases. Also, when considering the more widely used recommendation of ≥ 12 LNs, the respective success rates do not change (≥ 12 LNs in 87.4% and 77.1% of colon cases by PAs and in 58.0% and 50.0% by pathologists). We should add, however, that pathologists and PAs had the Dutch guideline of ≥ 10 LNs in mind [31].

Rectal cancer patients generally receive neo-adjuvant radiotherapy and/or chemotherapy, resulting in tumor downstaging [45]. Neo-adjuvant therapy, however, induces LN shrinkage [46]. Added to the fact that LNs in the rectal mesentery are already inherently smaller, as is the mesentery itself (compared to colonic mesentery), it becomes increasingly difficult to meet the criterion of examining ≥ 10 LNs in rectal resection specimens [47,48]. The problem of insufficient LN harvesting is therefore particularly relevant for rectal cancer. In our study, all metastatic LNs were < 5 mm in 16.7% (4/24) of the rectal cancer cases with ≥ 1 metastatic LNs. A previous study showed that all metastatic LNs measured < 5 mm in 32% of 98 rectal cancer cases with metastatic LNs, and in 8% of these cases all metastatic LNs measured < 2 mm [48]. Nonetheless, PAs in hospital A significantly contributed to harvesting ≥ 10 LNs (63.2% compared to 38.2% by pathologists) in the subgroup of rectal cancer cases in the present study.

This study concurs with comparable studies [3,4]. Reese *et al* reported a significant increase in LN harvest when a single PA was made responsible for LN harvest in a single centre [3]. The average numbers of harvested LNs increased, from pathologists averaging 13.6 LNs per specimen to the PA averaging 19.7 LNs. In terms of percentages, pathologists adequately sampled 58.4% of cases (≥ 12 LNs evaluated) while the PAs correctly sampled 84.0% of the cases. Galvis *et al* studied a single centre employing two PAs, who retrieved more negative LNs from axillary dissection and colorectal resection specimens than pathology residents did [4]. As these studies analyzed the work of just one or two PAs respectively, it cannot be deduced that PAs in general perform better than pathologists. However, our two-centre study, which measured results of seven PAs per hospital, corroborated the validity of the findings of these previous studies.

Our study is the first to show a direct link between PA deployment and improved harvest of small (< 5 mm) LNs (64.2% vs. 53.7% by PAs and pathologists, respectively). In the final year of our study, by which time PAs were responsible for 66.7% of cases, 46.2% of all metastatic LNs measured < 5 mm, which is consistent with previous studies (45-78%) [14,16-20].

Negative LN status (after adequate LN sampling) can spare colon cancer patients adjuvant chemotherapy. Theoretically, in hospital A the percentages of colon cancer patients who were eligible for adjuvant chemotherapy because of inadequate sampling alone could have been reduced from 17.3% with

pathologists harvesting LNs to 1.1% with PAs harvesting LNs. In hospital B, this average percentage would have fallen from 13.1% to 3.4%.

We demonstrate that PA deployment significantly improves patient safety and leaves pathologists more valuable time to spend on their remaining tasks requiring their specific expertise. Another reason to employ PAs can be to cut costs, as pathologists are relatively expensive [44]. Future studies will need to examine time and cost effectiveness of PAs, focussing on turnaround times and resubmission rates.

In conclusion, PAs certainly contribute to the patient safety since they recover more and, in particular smaller LNs from colorectal resection specimens than pathologists do. Moreover, they help to reduce costs and morbidity, by reducing the number of patients eligible for adjuvant chemotherapy due to inadequate LN sampling alone.

References

1. Santell JP, Hicks RW, McMeekin J, Cousins DD. (2003) Medication errors: experience of the United States Pharmacopeia (USP) MEDMARX reporting system. J Clin Pharmacol 43(7): 760-7.
2. Usubütün A, Uner S, Harorlu F, Ozer E, Tuzlali S, Ruacan A, et al. (2011) Pathology laboratories staff workload evaluation in Turkey: a survey study. Turk Patoloji.Derg 27(2): 98-105.
3. Reese JA, Hall C, Bowles K, Moesinger RC. (2009) Colorectal surgical specimen lymph node harvest: improvement of lymph node yield with a pathology assistant. J Gastrointest.Surg 13(18): 1459-63.
4. Galvis CO, Raab SS, D'Amico F, Grzybicki DM. (2001) Pathologists' assistants practice: a measurement of performance. Am J Clin Pathol 116(6): 816-22.
5. Le Voyer TE, Sigurdson ER, Hanlon AL, Mayer RJ, Macdonald JS, Catalano PJ, et al. (2003) Colon cancer survival is associated with increasing number of lymph nodes analyzed: a secondary survey of intergroup trial INT-0089. J Clin Oncol 21(15): 2912-9.
6. Wolmark N, Fisher B, Wieand HS. (1986) The prognostic value of the modifications of the Dukes' C class of colorectal cancer. An analysis of the NSABP clinical trials. Ann Surg 203(2): 115-22.
7. Newland RC, Chapuis PH, Smyth EJ. (1987) The prognostic value of substaging colorectal carcinoma. A prospective study of 1117 cases with standardized pathology. Cancer 60(4): 852-7.
8. Luna-Perez P, Rodriguez-Ramirez S, Alvarado I, Gutiérrez de la Barrera M, Labastida S. (2003) Prognostic significance of retrieved lymph nodes per specimen in resected rectal adenocarcinoma after preoperative chemoradiation therapy. Arch Med Res 34(4): 281-6.
9. Swanson RS, Compton CC, Stewart AK, Bland KI. (2003) The prognosis of T3N0 colon cancer is dependent on the number of lymph nodes examined. Ann Surg Oncol 10(1): 65-71.
10. Chang GJ, Rodriguez-Bigas MA, Skibber JM, Moyer VA. (2007) Lymph node evaluation and survival after curative resection of colon cancer: systematic review. J Natl.Cancer Inst 99(6): 433-41.
11. Rivadulla-Serrano MI, Martinez-Ramos D, Armengol-Carrasco M, Escrig-Sos J, Paiva-Coronel GA, Fortea-Sanchís C. (2010) Impact of the total number of harvested lymph nodes after colon cancer resections on survival in patients without involved lymph node. Rev Esp Enferm Dig 102(5): 296-301.
12. Sarli L, Bader G, Iusco D, Salvemini C, Mauro DD, Mazzeo A. (2005) Number of lymph nodes examined and prognosis of TNM stage II colorectal cancer. Eur J Cancer 41(2): 272-9.
13. Cserni G, Vinh-Hung V, Burzykowski T. (2002) Is there a minimum number of lymph nodes that should be histologically assessed for a reliable nodal staging of T3N0M0 colorectal carcinomas? J Surg Oncol 81(2): 63-9.
14. Herrera-Ornelas L, Justiniano J, Castillo N, Petrelli NJ, Stulc JP, Mittelman A. (1987) Metastases in small lymph nodes from colon cancer. Arch Surg 122(11): 1253-6.
15. Haboubi NY, Clark P, Kaftan SM, Schofield PF. (1992) The importance of combining xylene clearance and immunohistochemistry in the accurate staging of colorectal carcinoma. J R Soc Med 85(7): 386-8.

16. Herrera L, Villarreal JR. (1992) Incidence of metastases from rectal adenocarcinoma in small lymph nodes detected by a clearing technique. Dis Colon Rectum 35(8): 783-8.
17. Kotanagi H, Fukuoka T, Shibata Y, Yoshioka T, Aizawa O, Saito Y, et al. (1993) The size of regional lymph nodes does not correlate with the presence or absence of metastasis in lymph nodes in rectal cancer. J Surg Oncol 54(4): 252-4.
18. Andreola S, Leo E, Belli F, Bufalino R, Tomasic G, Lavarino C, et al. (1996) Manual dissection of adenocarcinoma of the lower third of the rectum specimens for detection of lymph node metastases smaller than 5 mm. Cancer 77(4): 607-12.
19. Rodriguez-Bigas MA, Maamoun S, Weber TK, Penetrante RB, Blumenson LE, Petrelli NJ. (1996) Clinical significance of colorectal cancer: metastases in lymph nodes < 5 mm in size. Ann Surg Oncol 3(2): 124-30.
20. Wong JH, Severino R, Honnebier MB, Tom P, Namiki TS. (1999) Number of nodes examined and staging accuracy in colorectal carcinoma. J Clin Oncol 17(9): 2896-900.
21. Fan L, Levy M, Aguilar CE, Mertens RB, Dhall D, Frishberg DP, et al. (2011) Lymph node retrieval from colorectal resection specimens for adenocarcinoma: is it worth the extra effort to find at least 12 nodes? Colorectal Dis 13(12): 1377-83.
22. Beresford M, Glynne-Jones R, Richman P, Makris A, Mawdsley S, Stott D, et al. (2005) The reliability of lymph-node staging in rectal cancer after preoperative chemoradiotherapy. Clin Oncol (R.Coll Radiol) 17(6): 448-55.
23. Caplin S, Cerottini JP, Bosman FT, Constanda MT, Givel JC. (1998) For patients with Dukes' B (TNM Stage II) colorectal carcinoma, examination of six or fewer lymph nodes is related to poor prognosis. Cancer 83(4): 666-72.
24. Cianchi F, Palomba A, Boddi V, Messerini L, Pucciani F, Perigli G, et al. (2002) Lymph node recovery from colorectal tumour specimens: recommendation for a minimum number of lymph nodes to be examined. World J Surg 26(3): 384-9.
25. Goldstein NS, Sanford W, Coffey M, Layfield LJ. (1996) Lymph node recovery from colorectal resection specimens removed for adenocarcinoma. Trends over time and a recommendation for a minimum number of lymph nodes to be recovered. Am J Clin Pathol 106(2): 209-16.
26. Goldstein NS. (2002) Lymph node recoveries from 2427 pT3 colorectal resection specimens spanning 45 years: recommendations for a minimum number of recovered lymph nodes based on predictive probabilities. Am J Surg Pathol 26(2): 179-89.
27. Hernanz F, Revuelta S, Redondo C, Madrazo C, Castillo J, Gómez-Fleitas M. (1994) Colorectal adenocarcinoma: quality of the assessment of lymph node metastases. Dis Colon Rectum 37(4): 373-6.
28. Maurel J, Launoy G, Grosclaude P, Gignoux M, Arveux P, Mathieu-Daudé H, et al. (1998) Lymph node harvest reporting in patients with carcinoma of the large bowel: a French population-based study. Cancer 82(8): 1482-6.
29. Prandi M, Lionetto R, Bini A, Francioni G, Accarpio G, Anfossi A, et al. (2002) Prognostic evaluation of stage B colon cancer patients is improved by an adequate lymphadenectomy: results of a secondary analysis of a large scale adjuvant trial. Ann Surg 235(4): 458-63.
30. Yoshimatsu K, Ishibashi K, Umehara A, Yokomizo H, Yoshida K, Fujimoto T, et al. (2005) How many lymph nodes should be examined in Dukes' B colorectal cancer? Determination on the basis of cumulative survival rate. Hepatogastroenterology 52(66): 1703-6.

31. Dutch guideline for colorectal cancer. Available from: http://www.oncoline.nl/index.php?pagina=/richtlijn/item/pagina.php&id=37094&richtlijn_id=933.
32. Guidance on Cancer Services. Improving Outcomes in Colorectal Cancers. Manual Update. 2004; Available from: https://www.nice.org.uk/guidance/csg5/resources/improving-outcomes-in-colorectal-cancer-update-773376301.
33. Compton CC, Fielding LP, Burgart LJ, Conley B, Cooper HS, Hamilton SR, et al. (2000) Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 124(7): 979-94.
34. AJCC Cancer Staging Manual. 7th edition. 2010
35. Progress Report on Cancer Control in the Netherlands by the Dutch National Cancer Control Programme 2005-2010. 2010; Available from: http://www.epaac.eu/from_heidi_wiki/Netherlands_Progress_Report_on_Cancer_Control_2010_English.pdf.
36. DICA rapportages 2011. Available from: http://www.clinicalaudit.nl/jaarrapportage/archief/DICA%20Jaarrapportage%202011.pdf.
37. Baxter NN, Morris AM, Rothenberger DA, Tepper JE. (2005) Impact of preoperative radiation for rectal cancer on subsequent lymph node evaluation: a population-based analysis. Int J Radiat.Oncol Biol Phys 61(2): 426-31.
38. Kerwel TG, Spatz J, Anthuber M, Wünsch K, Arnholdt H, Märkl B. (2009) Injecting methylene blue into the inferior mesenteric artery assures an adequate lymph node harvest and eliminates pathologist variability in nodal staging for rectal cancer. Dis Colon Rectum 52(5): 935-41.
39. Taflampas P, Christodoulakis M, Gourtsoyianni S, Leventi K, Melissas J, Tsiftsis DD. (2009) The effect of preoperative chemoradiotherapy on lymph node harvest after total mesorectal excision for rectal cancer. Dis Colon Rectum 52(8): 1470-4.
40. Latkauskas T, Lizdenis P, Janciauskiene R, Pranys D, Tamelis A, Pavalkis D. (2010) Lymph node retrieval after resection of rectal cancer following preoperative chemoradiotherapy. Medicina (Kaunas.) 46(5): 299-304.
41. Marks JH, Valsdottir EB, Rather AA, Nweze IC, Newman DA, Chernick MR. (2010) Fewer than 12 lymph nodes can be expected in a surgical specimen after high-dose chemoradiation therapy for rectal cancer. Dis Colon Rectum 53(7): 1023-9.
42. Morcos B, Baker B, Al Masri M, Haddad H, Hashem S. (2010) Lymph node yield in rectal cancer surgery: effect of preoperative chemoradiotherapy. Eur J Surg Oncol 36(4): 345-9.
43. Martijnse IS, Dudink RL, Kusters M, Rutten HJ, Nieuwenhuijzen GA, Wasowicz-Kemps DK. (2012) Patent blue staining as a method to improve lymph node detection in rectal cancer following neoadjuvant treatment. Eur J Surg Oncol 38(3): 252-8.
44. Grzybicki DM, Vrbin CM, Reilly TL, Zarbo RJ, Raab SS. (2004) Use of physician extenders in surgical pathology practice. Arch Pathol Lab Med 128(2): 165-72.
45. Wichmann MW, Muller C, Meyer G, Strauss T, Hornung HM, Lau-Werner U, et al. (2002) Effect of preoperative radiochemotherapy on lymph node retrieval after resection of rectal cancer. Arch Surg 137(2): 206-10.
46. Mekenkamp LJ, van Krieken JH, Marijnen CA, van de Velde CJ, Nagtegaal ID; Pathology Review Committee and the Co-operative Clinical Investigators. (2009) Lymph node retrieval in rectal cancer is dependent on many factors--the role of the tumor, the patient, the surgeon, the radiotherapist, and the pathologist. Am J Surg Pathol 33(10): 1547-53.
47. Sarashina H, Inoue I, Saitoh N, Nunomura M, Nakayama H, Okui K. (1991) Preoperative radiotherapy in rectal carcinoma. 2. The effect of irradiation on lymph node involvement. Strahlenther.Onkol 167(6): 361-5.

48. Dworák O. (1989) Number and size of lymph nodes and node metastases in rectal carcinomas. Surg Endosc 3(2): 96-9.

Interlaboratory variability in the grading of dysplasia in a nationwide cohort of colorectal adenomas

Chantal C.H.J. Kuijpers
Caro E. Sluijter
Jan H. von der Thüsen
Katrien Grünberg
Martijn G.H. van Oijen
Paul J. van Diest
Mehdi Jiwa
Iris D. Nagtegaal
Lucy I.H. Overbeek
Stefan M. Willems

Published in Histopathology, 2015 Dec 28. doi: 10.1111/his.12923

Abstract

Introduction

Although high-grade dysplasia (HGD) is a risk factor for malignant transformation and future development of adenomas/carcinomas, grade is not incorporated in the Dutch guidelines for colonoscopy surveillance, partly due to presumed interobserver variability. In a nationwide cohort of colorectal adenomas, we analyzed the interlaboratory variability in the grading of dysplasia in daily practice.

Methods

Using the Dutch Pathology Registry (PALGA), all synoptically reported classic adenomas in The Netherlands in 2013 were identified. Proportion of adenomas with HGD was determined for biopsies and polypectomies and compared between 37 laboratories by multivariable logistic regression analyses.

Results

In total, 21,030 colonoscopies of 20,270 patients were included. HGD was reported in 530 (3.6%) out of 14,866 adenomas diagnosed on biopsies (range between laboratories: 0%-13.6%) and in 983 (11.8%) out of 8,346 adenomas diagnosed on polypectomies (range 3.1%-42.9%). After adjustment for case mix, thirteen (35%) laboratories reported a significantly lower or higher frequency of HGD than average.

Conclusion

We observed considerable interlaboratory variation in the grading of dysplasia in colorectal adenomas, which could be only partly explained by differences in case mix. Therefore, better standardization of grading criteria is needed before grade of dysplasia can usefully be incorporated in colonoscopy surveillance guidelines.

Introduction

Colorectal adenomas, the precursor lesions of colorectal cancer (CRC) [1,2], are usually detected by colonoscopy. During colonoscopy, adenomas are either biopsied or removed by polypectomy. Colonoscopic removal of adenomas reduces the incidence and mortality of CRC [3-6].

Due to high recurrence rate of colorectal adenomas and the increased risk of developing new adenomas or carcinoma, regular colonoscopy surveillance is recommended for patients with adenomas [7-14]. Histopathological examination of adenomas is essential to assess risk factors that determine surveillance intervals. International variation in both surveillance intervals and risk factors is reported [7-14]. Common risk factors are multiplicity of adenomas, polyp size $\geq$ 1.0 cm, proximal localization, villous features, and high-grade dysplasia (HGD).

In the Dutch guidelines for colonoscopy surveillance [11], HGD is not incorporated since it was reported not to be an independent risk factor [15,16]. The fact that interpretation of subjective criteria that define dysplasia results in variation between pathologists probably plays a role here. Dysplasia is classified as low-grade or high-grade, taking both architectural and cytological criteria into account. This subdivision is artificial, as the development of dysplasia occurs as a continuum, rather than in discrete steps [17]. Only slight to substantial interobserver agreement in the grading of adenomas has been described in numerous previous studies in which several pathologists assessed the same set of colorectal adenomas [18-28]. The degree of variability in the grading of dysplasia in daily practice has never been studied.

We investigated whether histological grading of colorectal adenomas varies in daily practice between Dutch pathology laboratories and individual pathologists in a nationwide pathology database. Furthermore, we aimed to explain variation in grading through a questionnaire among pathologists, surveying the criteria that they use to grade dysplasia.

Methods

Data extraction

Data were extracted from PALGA, the nationwide network and registry of histo- and cytopathology in The Netherlands, which contains excerpts of

all pathology reports from Dutch pathology laboratories, with nationwide coverage since 1991 [29]. The PALGA database does not contain personally identifiable data, as the data are pseudomized in the laboratories and by a trusted third party (ZorgTTP). Data from patients who object against the use of their data for scientific research are not included in the PALGA database as well. The protocol of this study was approved by the scientific and privacy committee of PALGA.

Since the gradual implementation of synoptic reporting starting in 2009, an increasing number of pathology reports are made in a standardized manner [30,31]. All parameters of the synoptic pathology report are stored in the PALGA database as separate variables, enabling easy data extraction and analysis of large numbers of pathology reports.

We identified all synoptic pathology reports with ≥ 1 biopsied lesions and/or lesions removed by polypectomy during colonoscopy, histologically diagnosed in 2013 as a classic adenoma (i.e. tubular, tubulovillous or villous adenoma). Pathology reports of revised cases were excluded. In The Netherlands the national bowel cancer screening program was introduced in 2014, i.e. after the study period.

Per colonoscopy, we extracted patients' age and sex, the total number of reported adenomas in that colonoscopy and whether CRC was diagnosed during the same colonoscopy (synchronous CRC). Per adenoma, the following data were extracted: grade of dysplasia (low-grade dysplasia (LGD) or high-grade dysplasia (HGD) (with intramucosal carcinoma included as HGD)), whether a biopsy/biopsies or a polypectomy was evaluated, localization, and histological subtype. If applicable, the number of biopsies taken per adenoma, polyp size and the presence or absence of pseudo-invasion (displacement of adenomatous glands together with lamina propria through the muscularis mucosae into the submucosa) were collected as well.

The pathology laboratories where the adenomas were diagnosed were extracted anonymously. In addition, pathology laboratories that gave consent for further analyses were asked to participate in research on variation in grading between individual pathologists within the laboratory. The participating laboratories defined per pathology report which pathologist (either with their full names or in a coded fashion) was responsible for the microscopic evaluation of the adenomas.

Data analysis

Subgroups of adenomas diagnosed on biopsies and adenomas diagnosed on polypectomies were created. These subgroups were analyzed separately. We solely included adenomas of the laboratories that synoptically reported ≥ 100 adenomas on biopsies and polypectomies together in 2013. We assumed this number to be high enough to draw conclusions on the whole laboratory. Percentages of adenomas with HGD diagnosed on biopsies and on polypectomies were determined.

In both subgroups, analyses were performed at the colonoscopy level, including only one adenoma per colonoscopy (i.e. the one that was described first in the pathology report, or the first described adenoma with HGD). We chose to include only one adenoma per colonoscopy to correct for multiple paired measurements (adenomas) at the same time (colonoscopy). With the inclusion of paired measurements, some variables (e.g. age and number of adenomas per colonoscopy) would have been overrepresented in the multivariable analyses. Percentages of adenomas with HGD were determined per laboratory and per pathologist. On the basis of exploratory data analysis, we chose to solely analyze data of pathologists who synoptically reported ≥ 28 adenomas diagnosed on biopsies and ≥ 9 adenomas diagnosed on polypectomies.

To determine possible confounding variables that might explain variation between laboratories, we performed association analyses of a diagnosis of HGD with clinicopathological variables, based on the literature [32-36] and pathologists' experience. These variables included age, sex, total number of reported adenomas in the colonoscopy, synchronous CRC, localization, and histological subtype. For the biopsied adenomas, we also analyzed the variable number of biopsies taken from each adenoma. For adenomas diagnosed on polypectomies, the variables polyp size and presence of pseudo-invasion were analyzed as well. For the analyses, the following variables were categorized: age (< 70 and ≥ 70 years), total number of adenomas in the colonoscopy (1, 2 to 4, and ≥ 5), localization (rectum, left-sided colon, right-sided colon, and unknown), number of biopsies taken from an adenoma (1, 2, and ≥ 3), and polyp size (< 1.0 cm, ≥ 1.0 cm and unknown (in case of a piecemeal resection)).

Questionnaire among pathologists

To identify which criteria pathologists use to determine the grade of dysplasia, a questionnaire was performed, including all Dutch pathology laboratories. The questionnaire consisted of questions concerning the pathology laboratory, the years of experience as a pathologist, and whether the pathologists consider

themselves a specialized gastrointestinal (GI)-pathologist. Furthermore, we asked them to estimate the proportion of colorectal biopsies and polypectomies that they report synoptically and to state possible reasons to report a case in a narrative instead of a synoptic manner.

The crux of the questionnaire concerned questions about the manner of grading dysplasia. We asked whether the pathologist evaluates architectural criteria, cytological criteria, or both to determine the grade of dysplasia and which architectural and cytological criteria he or she evaluates. In a multiple choice question, one or more of the following architectural criteria could be ticked: cribriform growth, irregular crypts, enlarged crypts, and the option 'other' where the pathologist could fill in other architectural criteria. Cytological criteria included in a multiple choice question were enlarged hyperchromatic nuclei, nuclear-cytoplasmic ratio, elongated nuclei, and mucus depletion, as well as the option 'other'. Finally, we asked to state which books, articles or guidelines that they use as reference for the grading of dysplasia.

Statistical analysis

Statistical analyses were performed using SPSS version 21. Laboratories were tested as a categorical variable. Two reference laboratories were chosen to compare the other laboratories with in logistic regression analyses, selected on the basis of the average percentage of HGD found in adenomas diagnosed on biopsies and polypectomies, respectively. To analyze the association between laboratory and a diagnosis of HGD, clinicopathological variables were taken into account. Firstly, crude odds ratios (OR) and 95% confidence intervals (CI) were calculated by univariable logistic regression analyses. Variables were considered statistically significant if the 95% CI of the crude OR did not include 1. Statistically significant variables were checked for multicollinearity. If multicollinearity was not present, variables were included in multivariable logistic regression analyses. Adjusted ORs and 95% CIs were calculated. Adjusted ORs were compared between laboratories.

Results

Characteristics of adenomas

A flowchart of the adenomas included in this study is shown in Figure 1. Forty-two (76%) Dutch pathology laboratories reported classic colorectal adenomas using synoptic reporting in 2013. Five laboratories (12%) that synoptically reported less than 100 adenomas were excluded. Finally, from 37 laboratories,

21,030 colonoscopies with 32,391 adenomas were included (Supplementary Table 1). HGD was reported in 557 (2.6%) of 21,460 adenomas diagnosed on biopsies and in 1,048 (9.6%) of 10,931 adenomas diagnosed on polypectomies. Only one adenoma per colonoscopy was analyzed, and the characteristics of these included adenomas (n=23,212) are summarized in Table 1. Compared to the total cohort of adenomas, in these subgroups of one adenoma per colonoscopy, we observed significantly more adenomas with HGD ($p < 0.0001$; in the biopsies and polypectomies subgroups), adenomas from the right-sided colon ($p = 0.025$; in the biopsies subgroup), adenomas with ≥ 3 biopsies taken ($p < 0.0001$), and polypectomies ≥ 1.0 cm ($p = 0.037$). Histological subtype and pseudoinvasion were similar in the total cohort and the subgroup. All tested variables, except age, significantly differed between the subgroups of biopsies and polypectomies.

HGD was diagnosed in 530 (3.6%) out of the 14,866 adenomas diagnosed on biopsies and in 983 (11.8%) out of the 8,346 adenomas diagnosed on polypectomies ($p < 0.0001$).

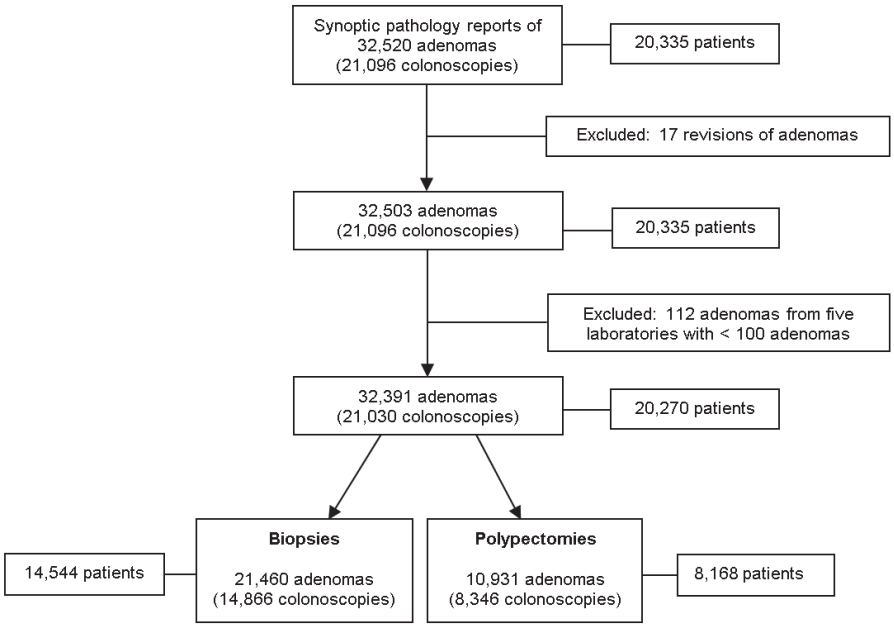


Figure 1. Flowchart of included adenomas. The sum of colonoscopies in the biopsies and polypectomies subgroups is higher than 21,030, as in a single colonoscopy adenomas can both be present in a biopsy and in a polypectomy. The same holds true for the number of patients.

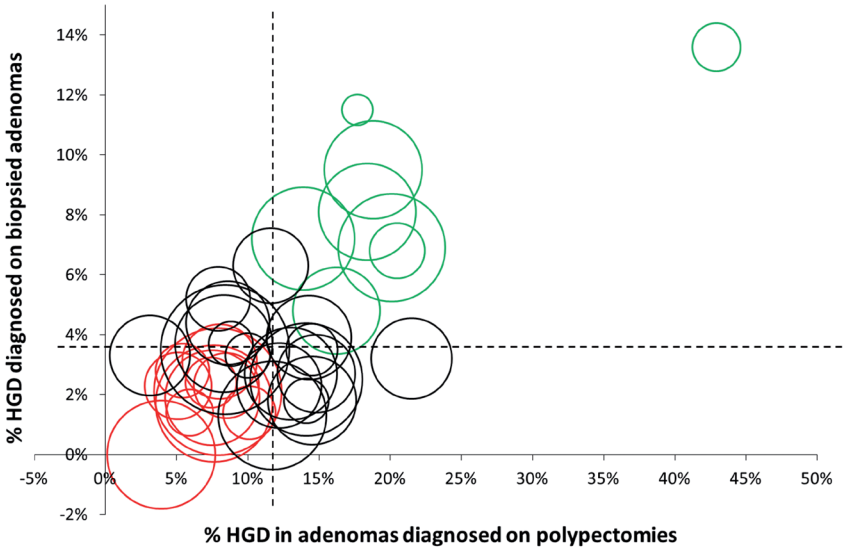


Figure 2. Interlaboratory variation between 37 pathology laboratories in the grading of dysplasia. Each circle represents a single laboratory, and the circle size indicates the total number of synoptically reported colonoscopies with one or more adenomas in the laboratory. Dotted lines illustrate the average percentages of adenomas diagnosed with HGD, i.e. 3.6% for biopsies and 11.8% for polypectomies. The red and green bubbles illustrate the laboratories that diagnosed HGD less often and more often than average (with a difference of >10% from the average), respectively, both in adenomas found on biopsies and on polypectomies.

Interlaboratory variation in the grading of dysplasia

Variation in the grading of dysplasia between the 37 laboratories (hereafter referred to as laboratories 1 to 37) was analyzed. The number of synoptically reported colonoscopies with 1 or more adenomas diagnosed on a biopsy varied from 26 to 938 per laboratory (median: 429). The number of synoptically reported colonoscopies with 1 or more adenomas diagnosed on a polypectomy varied from 37 to 605 per laboratory (median: 214). The uptake of synoptic reporting for colorectal adenomas in 2013 varied from 10% to 53% between laboratories.

Figure 2 shows the interlaboratory variation in the grading of dysplasia. The percentage of adenomas with HGD diagnosed on biopsies varied from 0% to 13.6% between laboratories, and the percentage of adenomas with HGD diagnosed on polypectomies varied from 3.1% to 42.9%. Eleven laboratories diagnosed HGD less often than average whereas eight laboratories reported HGD more often than average on both biopsies and polypectomies (with a > 10% difference from the average).

Table 1. Characteristics of the 23,212 included adenomas (the first adenoma or the first adenoma with HGD per colonoscopy).

	Biopsies 14,866 adenomas	Polypectomies 8,346 adenomas	p-value
Age; n (%)			0.556
< 70 yr	8,922 (60.0)	4,976 (59.6)	
≥ 70 yr	5,944 (40.0)	3,370 (40.4)	
Sex; n (%)			0.026
Male	8,470 (57.0)	4,881 (58.5)	
Female	6,396 (43.0)	3,465 (41.5)	
Number of adenomas in the colonoscopy; n (%)			< 0.0001
1 adenoma	9,170 (61.7)	4,903 (58.7)	
2-4 adenomas	5,321 (35.8)	3,171 (38.0)	
≥ 5 adenomas	375 (2.5)	272 (3.3)	
Synchronous CRC ; n (%)			< 0.0001
Absent	14,384 (96.8)	7,998 (95.8)	
Present	482 (3.2)	348 (4.2)	
Degree of dysplasia; n (%)			< 0.0001
Low-grade	14,336 (96.4)	7,363 (88.2)	
High-grade	530 (3.6)	983 (11.8)	
Localization *; n (%)			< 0.0001
Right-sided colon	7,570 (50.9)	2,458 (29.5)	
Left-sided colon	4,538 (30.5)	3,838 (46.0)	
Rectum	1,784 (12.0)	1,719 (29.5)	
Unknown	974 (6.6)	331 (4.0)	
Histological subtype; n (%)			< 0.0001
Tubular	12,529 (84.3)	5,065 (60.7)	
Tubulovillous	2,263 (15.2)	3,192 (38.2)	
Villous	74 (0.5)	89 (1.1)	
Number of biopsies taken from an adenoma; n (%)			na
1 biopsy	7,130 (48.0)	na	
2 biopsies	3,610 (24.3)	na	
≥ 3 biopsies	4,126 (27.8)	na	
Polyp size; n (%)			na
< 1.0cm	na	3,893 (46.6)	
≥ 1.0 cm	na	3,042 (36.5)	
Unknown	na	1,411 (16.9)	
Pseudoinvasion; n (%)			na
Absent	na	1,165 (14.0)	
Present	na	82 (1.0)	
Not mentioned	na	7,099 (85.1)	

* Right-sided colon: transverse colon, hepatic flexure, ascending colon and coecum or > 60 cm from the anus. Left-sided colon: sigmoid, descending colon and splenic flexure or 16-60 cm from the anus. Rectum: rectum and rectosigmoid or 0-15 cm from the anus.
na = not applicable

Univariable analysis showed that in adenomas diagnosed on biopsies, three and eight laboratories reported significantly less and more HGD, respectively, than the reference laboratory. For adenomas diagnosed on polypectomies, three and six laboratories reported significantly less and more HGD, respectively, than the reference laboratory.

All clinicopathological variables that were significantly associated with a diagnosis of HGD on univariable analyses (Supplementary Tables 2 and 3) were included in multivariable logistic regression analyses to correct for differences between laboratories. For the subset of adenomas diagnosed on biopsies, the variables corrected for were age, number of adenomas in the colonoscopy, the presence of synchronous CRC, localization, histological subtype, and number of biopsies taken from an adenoma. For the subset of adenomas diagnosed on polypectomies, the variables corrected for were age, number of adenomas in the colonoscopy, localization, histological subtype, polyp size, and the presence of pseudoinvasion. All variables, including the variable pathology laboratory, remained significantly associated with a diagnosis of HGD on multivariable analysis (data not shown).

In Figures 3a and 3b the association of the variable pathology laboratory (laboratories 1 to 37) with a diagnosis of HGD after multivariable analyses is plotted in forest plots showing adjusted ORs and their 95% CIs per laboratory. For the biopsied adenomas (Figure 3a), four laboratories reported a significantly lower proportion of adenomas with HGD than the reference laboratory. Laboratory 3 had not reported HGD on 748 biopsied adenomas at all. Furthermore, six laboratories reported a significantly higher proportion of adenomas with HGD than the reference laboratory. For the adenomas removed by polypectomy (Figure 3b), nine laboratories reported a significantly higher proportion of adenomas with HGD than the reference laboratory, including the six laboratories that reported significantly more HGD in the biopsy subgroup as well.

Variation in the grading of dysplasia within pathology laboratories

Twenty-one laboratories gave consent for their laboratory name to be included with the excerpts of pathology reports. Thirteen of these (4, 7, 11, 15, 17, 23, 24, 26, 27, 30, 31, 33, and 37) participated in further research on variation in the grading of dysplasia between individual pathologists and disclosed which pathologist microscopically evaluated which adenomas. In these 13 laboratories

a total of 88 pathologists reported adenomas synoptically in 2013. Fifty-six pathologists (from 13 laboratories) synoptically reported ≥ 28 adenomas diagnosed on biopsies and ≥ 9 adenomas diagnosed on polypectomies. One pathologist was excluded, as he or she was the only pathologist of his or her laboratory with sufficient synoptically reported adenomas. Therefore, intralaboratory variation was studied among 55 pathologists (from 12 laboratories). The number of included pathologists varied from 2 to 11 per

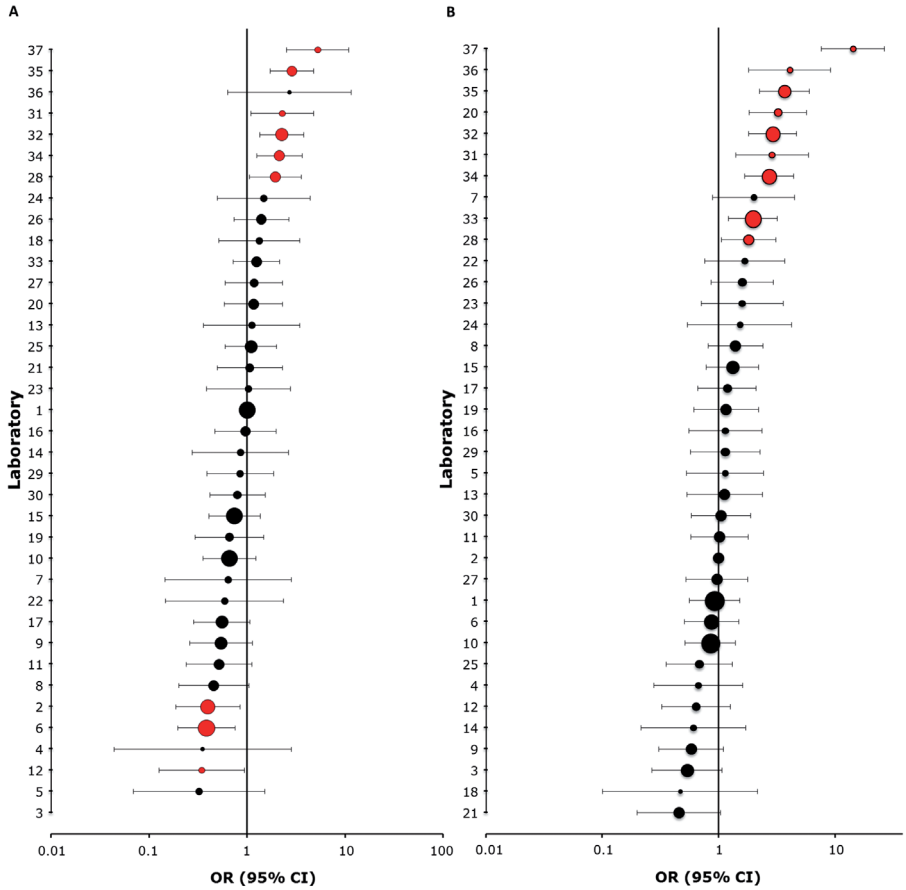


Figure 3. Forest plots showing adjusted odds ratios and 95% confidence intervals of laboratories regarding frequency of diagnosing HGD in adenomas diagnosed on biopsies (A) and polypectomies (B). Dots sizes indicate the total number of synoptically reported biopsies and polypectomies, respectively. The red dots indicate laboratories that were significantly aberrant with regard to frequency of diagnosing HGD, compared to the reference laboratory. In Figure A, no dot could be presented for laboratory 3, as no adenoma found on a biopsy was diagnosed with HGD.

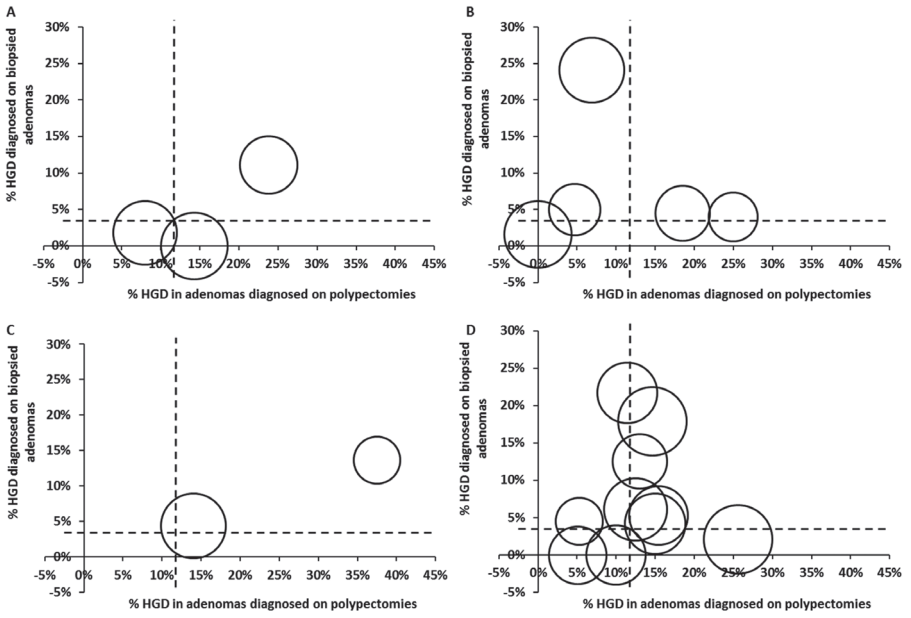


Figure 4. Variability between pathologists within four pathology laboratories in the grading of dysplasia. A. laboratory 23; B. laboratory 27; C. laboratory 31; D. laboratory 33. Each circle represents a single pathologist and the size of the circle represents the total number of adenomas diagnosed synoptically by the pathologist in 2013. Dotted lines illustrate the average nationwide percentages of adenomas diagnosed with HGD, i.e. 3.6% for biopsies and 11.8% for polypectomies.

laboratory (median: 3.5). In all laboratories, we observed, to a greater or lesser extent, variation between pathologists with regard to the percentage of diagnosing HGD. Most variation was observed in diagnosing HGD in adenomas found on polypectomies, but in a few laboratories (23, 27, 31, and 33), the percentage of HGD also varied considerably between pathologists in adenomas diagnosed on biopsies (Figure 4).

Results of questionnaire

Table 2 summarizes the main results of the questionnaire among Dutch pathologists. Fifty-one pathologists responded, of whom one third was a specialized gastrointestinal (GI) pathologist. All pathologists stated that they use architectural criteria to determine the grade of dysplasia, and 47 (92.2%) stated that they use cytological criteria as well. Furthermore, two pathologists (4.0%) stated that a villous adenoma is by definition high-grade. Cribriform growth was an architectural criterion for HGD utilized by all pathologists. Less agreement was observed for the use of other architectural

Table 2. Results of 51 Dutch pathologists responding to our questionnaire concerning the manner of grading dysplasia.

	General pathologist (n = 34)	GI pathologist (n = 17)	p-value
Laboratory; n (%)			0.119
Peripheral	30 (88.2)	12 (70.6)	
Academic	4 (11.8)	5 (29.4)	
Years of experience; n (%)			0.722
0-5	8 (23.5)	3 (17.6)	
6-10	9 (26.5)	4 (23.5)	
11-20	13 (38.2)	4 (23.5)	
>20	4 (11.8)	6 (35.3)	
How do you define the degree of dysplasia in a colorectal adenoma? *			
Evaluation of architectural criteria; n (%) *	34 (100.0)	17 (100.0)	na
Architectural criteria evaluated:			
Cribriform growth	34 (100.0)	17 (100.0)	na
Irregular crypts	15 (44.1)	6 (35.3)	0.546
Enlarged crypts	4 (11.8)	0	0.141
Evaluation of cytological criteria; n (%) *	32 (94.1)	15 (88.2)	0.461
Cytological criteria evaluated:			
Enlarged hyperchromatic nuclei	27 (84.4)	10 (66.7)	0.167
Nuclear-cytoplasmic ratio	23 (71.9)	10 (66.7)	0.716
Elongated nuclei	19 (59.4)	3 (20.0)	0.012
Mucus depletion	14 (43.8)	3 (20.0)	0.114
A villous adenoma is per definition high-grade; n (%) *	2 (5.9)	0	0.308

na = not applicable

* multiple answers possible

and cytological criteria, and general pathologists reported using these criteria more often than GI pathologists. Other architectural criteria that pathologists mentioned were crypt aberrations (complex, crowding, and solid), debris, and influx of neutrophils. Additional cytological criteria mentioned were aberrant nucleoli (enlarged, and vesicular) and nuclei (stratification, loss of polarity, and heterochromatin). Furthermore, the pathologists reported many different guidelines/articles/books that they use as reference for the grading of dysplasia.

Discussion

In this large nationwide cohort of synoptically reported colorectal adenomas, HGD was reported in 2.6% and 9.6% of the adenomas diagnosed on biopsies and polypectomies, respectively. In the analyses at the colonoscopy level,

including only one adenoma per colonoscopy, the percentages of HGD were somewhat higher (3.6% and 11.8%). The percentage of HGD in adenomas removed by polypectomy largely corresponds with that described in a meta-analysis of six studies (9.5%) [15]. This makes it plausible that our subset of synoptically reported adenomas diagnosed on polypectomies is representative of the whole cohort, and we assume the same holds true for those diagnosed on biopsies.

The main finding of our study is the considerable interlaboratory variation in the grading of dysplasia between 37 Dutch pathology laboratories, which could be only partly explained by differences in case mix between the laboratories. Even after adjustment for case mix, thirteen (35%) laboratories diagnosed a significantly lower or higher frequency of HGD: four laboratories with less HGD on biopsies only, six laboratories with more HGD on biopsies and polypectomies, and three laboratories with more HGD on polypectomies only. Although the selection of the two reference laboratories was arbitrary, and the average frequency of HGD might not necessarily indicate greater accuracy of diagnosis, comparison of the laboratories with these reference laboratories was deemed to be the best possible approach. Unfortunately, because of our study design, including only synoptic pathology reports, it was not possible to adjust for all variables that could be possibly associated with a diagnosis of HGD, such as previous adenomas with HGD or previous CRC. However, the variables deemed to be most important were included in the analyses and adjusted for: age, the number of adenomas during the colonoscopy, rectal localization, and a villous component (in both the biopsy and the polypectomy subgroup); synchronous CRC and the number of biopsies taken from an adenoma (in the biopsy subgroup); and polyp size and the presence of pseudoinvasion (in the polypectomy subgroup). These findings largely correspond with previous studies [32-37], which were all based on polypectomies.

Different hospitals or gastroenterologists might use various protocols or standards for removal and/or dismissal of (small) adenomas by polypectomy, possibly affecting the percentage of HGD within a laboratory. To adjust for possible differences, a subset of polyps of ≥ 2.0 cm was analyzed, as we assumed that all gastroenterologists would remove such large adenomas. Interlaboratory variation in the grading of dysplasia persisted: overall 27.5% HGD with a range of 4.8% to 63.0% between laboratories.

It is important to emphasize that this study aimed to investigate variation in daily practice, rather than determining interobserver or intraobserver variability through review of a set of adenomas by multiple pathologists. Previous studies that had this goal already described substantial interobserver and intraobserver variability in the grading of dysplasia [18-27]. In addition to interlaboratory variation, we observed intralaboratory variation between individual pathologists in a subgroup of twelve pathology laboratories. The results of our questionnaire show considerable heterogeneity in criteria used by pathologists to determine the grade of dysplasia, which might explain part of the variation, and emphasizes the need for better standardization of criteria for the grading of dysplasia.

Because the role of HGD in an adenoma removed during colonoscopy in the risk of developing future (high-risk) adenomas or CRC is unclear, national and international colonoscopy surveillance guidelines are not consistent. Some include HGD as a risk factor, whereas others, like the Dutch guideline, do not. Inconsistency of, and lack of clarity in guidelines might have led to underestimation of the importance of grading adenomas by some pathologists. However, standardization is needed, not only to improve diagnostics, but also to enable reliable research on HGD as a risk factor.

The results of this study and individual feedback to laboratories might increase the awareness among pathologists that the grading of dysplasia is probably not performed optimally, which is certainly a first step towards standardization. Osmond *et al* studied interobserver agreement before and after pathologists read the Canadian guidelines for standardized reporting of adenomas [28]. Interobserver agreement improved for the histological subtype. However, a significant decrease in interobserver agreement for grade of dysplasia was observed, although Kappa values were moderate before and after the intervention. In The Netherlands, an interactive online learning module and examination were introduced together with the introduction of the national bowel cancer screening program in 2014, and these are expected to contribute to synchronization of judgement. In further research, the effect of the learning module and examination will be studied, and the most reproducible criteria for dysplasia grading will be assessed.

In conclusion, this large nationwide cohort of colorectal adenomas demonstrates considerable interlaboratory and intralaboratory variation in the grading of dysplasia. Better standardization of grading criteria is needed before grade of dysplasia can usefully be incorporated in colonoscopy surveillance guidelines to determine the surveillance interval.

References

1. Muto T, Bussey HJ, Morson BC. (1975) The evolution of cancer of the colon and rectum. Cancer 36(6): 2251-70.
2. Morson BC. (1984) The evolution of colorectal carcinoma. Clin Radiol 35(6): 425-31.
3. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al. (1993) Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 329(27): 1977-81.
4. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L. (2009) Association of colonoscopy and death from colorectal cancer. Ann Intern Med 150(1): 1-8.
5. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. (2011) Protection from colorectal cancer after colonoscopy: a population-based, case-control study. Ann Intern Med 154(1): 22-30.
6. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. (2012) Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 366(8): 687-96.
7. Cairns SR, Scholefield JH, Steele RJ, Dunlop MG, Thomas HJ, Evans GD, et al. (2010) Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). Gut 59(5): 666-89.
8. Lieberman DA, Rex DK, Winawer SJ, Giardiello FM, Johnson DA, Levin TR, et al. (2012) Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 143(3): 844-57.
9. Leddin D, Enns R, Hilsden R, Fallone CA, Rabeneck L, Sadowski DC, et al. (2013) Colorectal cancer surveillance after index colonoscopy: guidance from the Canadian Association of Gastroenterology. Can J Gastroenterol 27(4): 224-8.
10. Hassan C, Quintero E, Dumonceau JM, Regula J, Brandão C, Chaussade S, et al. (2013) Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy 45(10): 842-51.
11. Dekker EVL, van Leerdam ME, Hazewinkel Y, Sanduleanu S, Vasen HF, Lansdorp-Vogelaar I, et al. (2013) Nederlandse richtlijn coloscopie surveillance. Available from: www.mdl.nl.
12. Arditi C, Gonvers JJ, Burnand B, Minoli G, Oertli D, Lacaine F, et al. (2009) Appropriateness of colonoscopy in Europe (EPAGE II). Surveillance after polypectomy and after resection of colorectal cancer. Endoscopy 41(3): 209-17.
13. Kohn LT, Corrigan JM, Donaldson MS. To Err is Human; Building a Safer Health System. 1999. Institute of Medicine, National Academy Press, Washington DC
14. Allsbrook WC Jr, Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JL. (2001) Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. Hum Pathol 32(1): 81-8.
15. Martinez ME, Baron JA, Lieberman DA, Schatzkin A, Lanza E, Winawer SJ, et al. (2009) A pooled analysis of advanced colorectal neoplasia diagnoses after colonoscopic polypectomy. Gastroenterology 136(3): 832-41.
16. van Heijningen EM, Lansdorp-Vogelaar I, Kuipers EJ, Dekker E, Lesterhuis W, Ter Borg F, et al. (2013) Features of adenoma and colonoscopy associated with recurrent colorectal neoplasia based on a large community-based study. Gastroenterology 144(7): 1410-8.

17. Bosman FT. (2001) Dysplasia classification: pathology in disgrace? J Pathol 194(2): 143-4.
18. Brown LJ, Smeeton NC, Dixon MF. (1985) Assessment of dysplasia in colorectal adenomas: an observer variation and morphometric study. J Clin Pathol 38(2): 174-9.
19. Fenger C, Bak M, Kronborg O, Svanholm H. (1990) Observer reproducibility in grading dysplasia in colorectal adenomas: comparison between two different grading systems. J Clin Pathol 43(4): 320-4.
20. Jensen P, Krogsgaard MR, Christiansen J, Braendstrup O, Johansen A, Olsen J. (1995) Observer variability in the assessment of type and dysplasia of colorectal adenomas, analyzed using kappa statistics. Dis Colon Rectum 38(2): 195-8.
21. Terry MB, Neugut AI, Bostick RM, Potter JD, Haile RW, Fenoglio-Preiser CM. (2002) Reliability in the classification of advanced colorectal adenomas. Cancer Epidemiol Biomarkers Prev 11(7): 660-3.
22. Yoon H, Martin A, Benamouzig R, Longchampt E, Deyra J, Chaussade S, et al. (2002) Inter-observer agreement on histological diagnosis of colorectal polyps: the APACC study. Gastroenterol Clin Biol 26(3): 220-4.
23. van Putten PG, Hol L, van Dekken H, van Krieken HJ, van Ballegooijen M, Kuipers EJ, et al. (2011) Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening. Histopathology 58(6): 974-81.
24. Foss FA, Milkins S, McGregor AH. (2012) Inter-observer variability in the histological assessment of colorectal polyps detected through the NHS Bowel Cancer Screening Programme. Histopathology 61(1): 47-52.
25. Turner JK, Williams GT, Morgan M, Wright M, Dolwani S. (2013) Interobserver agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists in Wales. Histopathology 62(6): 916-24.
26. Mahajan D, Downs-Kelly E, Liu X, Pai RK, Patil DT, Rybicki L, et al. (2013) Reproducibility of the villous component and high-grade dysplasia in colorectal adenomas <1 cm: implications for endoscopic surveillance. Am J Surg Pathol 37(3): 427-33.
27. Lasisi F, Mouchli A, Riddell R, Goldblum JR, Cummings OW, Ulbright TM, et al. (2013) Agreement in interpreting villous elements and dysplasia in adenomas less than one centimetre in size. Dig Liver Dis 45(12): 1049-55.
28. Osmond A, Li-Chang H, Kirsch R, Divaris D, Falck V, Liu DF, et al. (2014) Interobserver variability in assessing dysplasia and architecture in colorectal adenomas: a multicentre Canadian study. J Clin Pathol 67(9): 781-6.
29. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, et al. (2007) Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 29(1): 19-24.
30. Srigley J, Lankshear S, Brierley J, McGowan T, Divaris D, Yurcan M, et al. (2013) Closing the quality loop: facilitating improvement in oncology practice through timely access to clinical performance indicators. J Oncol Pract 9(5): 255-61.
31. Williams CL, Bjugn R, Hassell LA. (2015) Current status of discrete data capture in synoptic surgical pathology and cancer reporting. Pathol Lab Med Int 7: 11-22.
32. Konishi F, Morson BC. (1982) Pathology of colorectal adenomas: a colonoscopic survey. J Clin Pathol 35(8): 830-41.

33. O'Brien MJ, Winawer SJ, Zauber AG, Gottlieb LS, Sternberg SS, Diaz B, et al. (1990) The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology 98(2): 371-9.
34. Gschwantler M, Kriwanek S, Langner E, Göritzer B, Schrutka-Kölbl C, Brownstone E, et al. (2002) High-grade dysplasia and invasive carcinoma in colorectal adenomas: a multivariate analysis of the impact of adenoma and patient characteristics. Eur J Gastroenterol Hepatol 14(2): 183-8.
35. Kristjansdottir S, Jonasson JG, Cariglia N, Thjodleifsson. (2010) Colonic adenomas found via colonoscopy: yield and risk factors for high-grade dysplasia. Digestion 82(4): 252-7.
36. Zare-Mirzaie A, Abolhasani M, Aryamanesh A. (2013) Left sided colorectal adenomatous polyps have more risk for high grade dysplasia. Acta Med Iran 51(3): 172-7.
37. van Enckevort CC, de Graaf AP, Hollema H, Sluiter WJ, Kleibeuker JH, Koornstra JJ. (2014) Predictors of colorectal neoplasia after polypectomy: based on initial and consecutive findings. Neth J Med 72(3): 139-45.

Supplementary tables

Supplementary table 1. Characteristics of the 21,030 included colonoscopies and 32,391 included adenomas. The sum of colonoscopies in the biopsies and polypectomies subgroups is higher than 21,030, as from a single colonoscopy adenomas can both be present in a biopsy and in a polypectomy.

	Biopsies	Polypectomies	
Reported at the colonoscopy level	**14,866 colonoscopies**	**8,346 colonoscopies**	**p-value**
Age; n (%)			0.556
< 70 yr	8,922 (60.0)	4,976 (59.6)	
≥ 70 yr	5,944 (40.0)	3,370 (40.4)	
Sex; n (%)			0.026
Male	8,470 (57.0)	4,881 (58.5)	
Female	6,396 (43.0)	3,465 (41.5)	
Number of adenomas in the colonoscopy; n (%)			< 0.0001
1 adenoma	9,170 (61.7)	4,903 (58.7)	
2-4 adenomas	5,321 (35.8)	3,171 (38.0)	
≥ 5 adenomas	375 (2.5)	272 (3.3)	
Synchronous CRC ; n (%)			< 0.0001
Absent	14,384 (96.8)	7,998 (95.8)	
Present	482 (3.2)	348 (4.2)	
Reported at the adenoma level	**21,460 adenomas**	**10,931 adenomas**	**p-value**
Degree of dysplasia; n (%)			< 0.0001
Low-grade	20,903 (97.4)	9,883 (90.4)	
High-grade	557 (2.6)	1,048 (9.6)	
Localization *; n (%)			< 0.0001
Right-sided colon	10,407 (48.5)	3,202 (29.3)	
Left-sided colon	7,215 (33.6)	5,091 (46.6)	
Rectum	2,557 (11.9)	2,225 (20.4)	
Unknown	1,281 (6.0)	413 (3.8)	
Histological subtype; n (%)			< 0.0001
Tubular	18,170 (84.7)	6,797 (62.2)	
Tubulovillous	3,197 (14.9)	4,029 (36.9)	
Villous	93 (0.4)	105 (1.0)	
Number of biopsies taken from an adenoma; n (%)			na
1 biopsy	10,568 (49.2)	na	
2 biopsies	5,184 (24.2)	na	
≥ 3 biopsies	5,708 (26.6)	na	
Polyp size; n (%)			na
< 1.0cm	na	5,287 (48.4)	
≥ 1.0 cm	na	3,804 (34.8)	
Unknown	na	1,840 (16.8)	
Pseudoinvasion; n (%)			na
Absent	na	1,554 (14.2)	
Present	na	96 (0.9)	
Not mentioned	na	9,281 (84.9)	

* Right-sided colon: transverse colon, hepatic flexure, ascending colon and coecum or > 60 cm from the anus. Left-sided colon: sigmoid, descending colon and splenic flexure or 16-60 cm from the anus. Rectum: rectum and rectosigmoid or 0-15 cm from the anus.

Supplementary table 2. Association of clinicopathological variables with a diagnosis of HGD in adenomas diagnosed on biopsies. Univariable and multivariable analyses including one adenoma per colonscopy (n=14,866).

	N (total)	N (%) with HGD	Crude OR (95% CI)	Adjusted OR (95% CI)
Age (years)				
<70	8,922	229 (2.6)	1	1
≥ 70	5,944	301 (5.1)	**2.025 (1.700 – 2.412)**	**1.723 (1.415 – 2.099)**
Sex				
Male	8,470	322 (3.8)	1	
Female	6,396	208 (3.3)	0.851 (0.712 – 1.016)	
Number of adenomas in colonoscopy				
1 adenoma	9,170	282 (3.1)	1	1
2-4 adenomas	5,321	226 (4.2)	**1.398 (1.170 – 1.671)**	**2.131 (1.730 – 2.626)**
≥ 5 adenomas	375	22 (5.9)	**1.964 (1.257 – 3.070)**	**3.673 (2.198 – 6.139)**
Synchronous CRC				
Absent	14,384	475 (3.3)	1	1
Present	482	55 (11.4)	**3.772 (2.807 – 5.067)**	**2.787 (1.966 – 3.950)**
Localization				
Right-sided colon	7,570	150 (2.0)	**0.553 (0.441 – 0.693)**	**0.359 (0.279 – 0.460)**
Left-sided colon	4,538	160 (3.5)	1	1
Rectum	1,784	196 (11.0)	**3.377 (2.720 – 4.194)**	**1.694 (1.312 – 2.187)**
Unknown	974	24 (2.5)	0.691 (0.448 – 1.068)	**0.435 (0.271 – 0.697)**
Histological subtype				
Tubular	12,529	197 (1.6)	1	1
Tubulovillous	2,263	316 (14.0)	**10.160 (8.450 – 12.215)**	**6.506 (5.225 – 8.103)**
Villous	74	17 (23.0)	**18.670 (10.668 – 32.672)**	**6.957 (3.610 – 13.406)**
# of biopsies taken from an adenoma				
1 biopsy	7,130	68 (1.0)	1	1
2 biopsies	3,610	70 (1.9)	**2.054 (1.467 – 2.874)**	**1.678 (1.183 – 2.379)**
≥ 3 biopsies	4,126	392 (9.5)	**10.903 (8.402 – 14.147)**	**5.755 (4.343 – 7.626)**

Supplementary table 3. Association of clinicopathological variables with a diagnosis of HGD in adenomas diagnosed on polypectomies. Univariable and multivariable analyses including one adenoma per colonscopy (n=8,346).

	N (total)	N (%) with HGD	Crude OR (95% CI)	Adjusted OR (95% CI)
Age (years)				
<70	4,976	516 (10.4)	1	1
≥ 70	3,370	467 (13.9)	**1.390 (1.217 – 1.589)**	**1.338 (1.156 – 1.548)**
Sex				
Male	4,881	571 (11.7)	1	
Female	3,465	412 (11.9)	1.019 (0.890 – 1.166)	
Number of adenomas in the colonoscopy				
1 adenoma	4,903	542 (11.1)	1	1
2-4 adenomas	3,171	402 (12.7)	**1.168 (1.018 – 1.340)**	**1.445 (1.241 – 1.682)**
≥ 5 adenomas	272	39 (14.3)	1.347 (0.948 – 1.912)	**2.284 (1.549 – 3.368)**
Synchronous CRC				
Absent	7,998	933 (11.7)	1	
Present	348	50 (14.4)	1.271 (0.934 – 1.727)	
Localization				
Right-sided colon	2,458	152 (6.2)	**0.452 (0.374 – 0.547)**	**0.533 (0.432 – 0.657)**
Left-sided colon	3,838	488 (12.7)	1	1
Rectum	1,719	298 (17.3)	**1.440 (1.231 – 1.684)**	**1.276 (1.072 – 1.519)**
Unknown	331	45 (13.6)	1.080 (0.778 – 1.500)	1.119 (0.781 – 1.603)
Histological subtype				
Tubular	5,065	364 (7.2)	1	1
Tubulovillous	3,192	596 (18.7)	**2.965 (2.580 – 3.407)**	**2.231 (1.890 – 2.634)**
Villous	89	23 (25.8)	**4.501 (2.767 – 7.320)**	**2.742 (1.617 – 4.651)**
Polyp size				
< 1.0 cm	3,893	149 (3.8)	1	1
≥ 1.0 cm	3,042	615 (20.2)	**6.367 (5.286 – 7.670)**	**4.532 (3.696 – 5.557)**
Unknown	1,411	219 (15.5)	**4.617 (3.712 – 5.742)**	**4.259 (3.356 – 5.405)**
Pseudoinvasion				
Absent	1,165	114 (9.8)	1	1
Present	82	45 (54.9)	**11.213 (6.966 – 18.047)**	**6.081 (3.594 – 10.288)**
Not mentioned	7,099	824 (11.6)	1.211 (0.985 – 1.488)	1.065 (0.851 – 1.332)

Interlaboratory variability in the histological grading of colorectal adenocarcinomas in a nationwide cohort

Chantal C.H.J. Kuijpers
Caro E. Sluijter
Jan H. von der Thüsen
Katrien Grünberg
Martijn G.H. van Oijen
Paul J. van Diest
Mehdi Jiwa
Iris D. Nagtegaal
Lucy I.H. Overbeek
Stefan M. Willems

Accepted for publication in The American Journal of Surgical Pathology

Abstract

Introduction

Differentiation grade of colorectal adenocarcinoma (CRC) is a prognostic factor and important for therapy selection. In patients with stage II colon cancer, poor differentiation is an indication for adjuvant chemotherapy.

Methods

The variability in daily practice in grading of CRC was assessed in a nationwide cohort. Using the Dutch Pathology Registry (PALGA), all synoptically reported CRC resections from 2010 up to 2013 were identified. Proportions of poorly differentiated (PD) adenocarcinomas were determined and compared between 35 laboratories by univariable and multivariable logistic regression analyses.

Results

In total, 11,719 resections of 11,681 patients were included, of which 1,427 (12.2%) were PD (range between 35 laboratories: 5.0% to 33.2%). After adjustment for case mix, four (11%) laboratories still reported a significantly lower (n=2) or higher (n=2) proportion of PD adenocarcinoma than the reference laboratory. Seven out of eight investigated laboratories showed considerable intralaboratory variation between pathologists as well. In a subgroup of 2,812 patients (2,813 tumors) who could have been eligible for adjuvant chemotherapy solely based on the differentiation grade (stage II colon cancer patients without other high-risk factors (i.e. T4, <10 lymph nodes evaluated, perforation, ileus, or angio-invasion)), 258 (9.2%) were PD (range between laboratories: 0% to 22.7%). In this subgroup, four laboratories still diagnosed significantly more PD adenocarcinomas after multivariable logistic regression analysis, increasing the number of colon cancer patients eligible for adjuvant therapy.

Conclusions

This large nationwide cohort demonstrates considerable interlaboratory and intralaboratory variation in differentiation grading of CRC. Better standardization of grading criteria is needed for optimal determination of prognosis and treatment selection.

Introduction

Differentiation grade reflects the aggressiveness of a tumor: in many malignancies, including colorectal cancer (CRC) grade is a prognostic factor independent of stage [1-5]. Poor differentiation grade is associated with a higher risk of lymph node metastasis, higher recurrence rates and poorer survival [6,7]. Therefore, the histological grade has consequences for therapy selection: poor differentiation of early rectal cancer is a contraindication for local resection [8,9] and an indication for adjuvant chemotherapy in patients with stage II colon cancer [10].

Grading is usually performed on architectural criteria. Traditionally, three- or four-tiered systems were used: CRC were graded into well-, moderately- and poorly differentiated (and in some systems undifferentiated) [11,12]. However, considerable interobserver variability was observed in designating well- and moderately differentiated classes [12-14]. For this reason and because of the similar behaviour of well- and moderately differentiated CRCs, in 1999, the College of American Pathologists (CAP) recommended a two-tiered grading system, in which well- and moderately differentiated tumors were combined as low-grade (> 50% gland formation) and poorly differentiated and undifferentiated tumors as high-grade (< 50% gland formation) [11,15]. Interobserver agreement was expected to increase, but remained at best fair [12].

Guidelines are not consistent with regard to criteria to grade CRC [11,12,15-17]. Although the degree of gland formation is uniformly used for differentiation grading, some guidelines recommend additional criteria, such as irregular, distorted and small glands, and loss of nuclear polarity [16,17]. Furthermore, guidelines differ on whether differentiation grade should be based on the least differentiated component, excluding the invasive margin of focal dedifferentiation [15,16], or on the predominant grade [17].

Because histological grading is important for adequate therapy selection and optimal patient care, optimal interobserver and intraobserver agreement is needed. In this nationwide study, we aimed to determine whether in daily practice histological grading of CRC resection specimens varies between Dutch pathology laboratories and individual pathologists. Finally, we aimed to explain differences in grading through a questionnaire among pathologists.

Methods

Data extraction

Data were extracted from PALGA, the nationwide network and registry of histo- and cytopathology in The Netherlands, which contains excerpts of all pathology reports from Dutch pathology laboratories, with nationwide coverage since 1991 [18]. The PALGA database does not contain personally identifiable data, as the data are pseudomized by a trusted third party (ZorgTTP, Houten, The Netherlands). Data from patients who object against the use of their data for scientific research are not included in the PALGA database. The protocol of this study was approved by the scientific and privacy committee of PALGA. Since the gradual implementation of synoptic reporting starting in 2009, an increasing number of pathology specimens are reported in a standardized manner [19,20]. All parameters of the synoptic pathology report are stored in the PALGA database as separate variables, enabling easy data extraction and analysis of large numbers of pathology reports.

We retrieved all synoptic pathology reports of resection specimens of CRC diagnosed with adenocarcinoma not otherwise specified between January 1st 2010 and December 31st 2013. In The Netherlands, the national bowel cancer screening program was introduced in 2014, which was after the study period. We extracted from PALGA the age and sex of the patient, tumor localization, year of examination, type of resection, whether the patient received neoadjuvant therapy, tumor size, extent of tumor invasion (T stage), the number of lymph nodes evaluated, lymph node invasion (N stage), angio-invasion, clinical presence of distant metastasis, perforation, and ileus, and differentiation grade (well- to moderately differentiated or poorly- to undifferentiated (the latter hereafter referred to as poorly differentiated (PD)). Data were solely extracted from the pathology reports, patient charts were not consulted.

Only original pathology reports were included, thereby excluding pathology reports of reassessments in other laboratories ("revisions"). Reports of patients that received neoadjuvant treatment were excluded as well, since treatment can influence grading [21,22]. Since, according to our national guidelines, rectal cancer should be treated with neoadjuvant radiotherapy or neoadjuvant radiochemotherapy, we excluded all rectal cancers with an unknown status for neoadjuvant treatment. Since neoadjuvant treatment for colon cancer is uncommon (1.5% of our cases), those with unknown neoadjuvant status were included and considered as non-neoadjuvant.

The names of the pathology laboratories were extracted anonymously from PALGA. In addition, pathology laboratories were asked for consent to participate in further research on variation in grading between individual pathologists within the laboratory. Participating laboratories defined per pathology report which pathologist (either with their full names or in a coded fashion) was responsible for the microscopic evaluation of the CRC.

Data analysis

The overall nationwide proportion of PD adenocarcinomas was determined. Multiple synchronous adenocarcinomas, i.e. those resected during the same surgery or within six months, were considered paired measurements. Therefore, only one synchronous adenocarcinoma per patient was included in the analyses: either the first described adenocarcinoma or the first described PD adenocarcinoma. Metachronous adenocarcinomas, i.e. resected > 6 months after each other, were included as separate entities.

The proportion of PD adenocarcinomas was determined per laboratory. Only laboratories that synoptically reported ≥ 100 adenocarcinomas were included in the interlaboratory analyses. This number was based on exploratory data analysis, allowing to draw representative conclusions for a certain laboratory. Furthermore, proportions of PD adenocarcinomas per pathologist were determined. On the basis of exploratory data analysis, we chose to solely analyze data of pathologists who synoptically reported ≥ 10 adenocarcinomas.

A subanalysis was performed in a subgroup of patients who could have been eligible for adjuvant chemotherapy solely based on the differentiation grade, i.e. patients with stage II colon adenocarcinoma with no other high-risk features (T4, <10 lymph nodes evaluated, perforation, ileus, or angio-invasion). Again in this subgroup, proportions of PD adenocarcinomas were compared between laboratories.

To determine possible confounding variables, we performed association analyses of a diagnosis of PD adenocarcinoma with clinicopathological variables, chosen based on the literature [23] and pathologists' experience. The analyzed variables were age, sex, localization, year of examination, tumor size, T stage, N stage, and angio-invasion.

For the association analyses, several variables were categorized. Age was subdivided into four categories (< 60; 60 to 69; 70 to 79; and ≥ 80 years).

Localization was based on the reported localization or the type of resection and subdivided into rectum, left-sided colon, right-sided colon, and unknown. Outlier values of tumor size (0 cm and > 25 cm) were deemed unknown.

Questionnaire among pathologists

To identify in what manner pathologists grade colorectal adenocarcinomas, a questionnaire was spread via the weekly newsletter of the Dutch Society of Pathology (NVVP) to all member pathologists (~320) in The Netherlands. The questionnaire included questions on the criteria used to determine differentiation grade (i.e. architectural, cytological, or both), and how the pathologist deals with intratumoral heterogeneity of differentiation grade. Finally, we asked to state which books, articles or guidelines they use as reference for the grading of colorectal adenocarcinomas. A copy of the questionnaire is included as a supplement.

Statistical analysis

Statistical analysis was performed using SPSS version 21. Laboratories were tested as a categorical variable. A reference laboratory was chosen to compare laboratories by logistic regression analyses. The median laboratory with regard to the proportion of PD adenocarcinoma was set as the reference laboratory. To analyze the association between laboratory and a diagnosis of PD adenocarcinoma, clinicopathological variables were taken into account. Firstly, crude odds ratios (OR) and 95% confidence intervals (CI) were calculated by univariable logistic regression analyses. Variables were considered statistically significant if the 95% CI of the crude OR did not include 1. Statistically significant variables were checked for multicollinearity. If multicollinearity was absent, variables were included in multivariable logistic regression analyses. Adjusted ORs and 95% CIs were calculated. Adjusted ORs were compared between laboratories, which was repeated for the subgroup of patients who could have been eligible for adjuvant chemotherapy solely based on the differentiation grade.

Results

Figure 1 presents a flowchart of the included adenocarcinomas. Forty-one pathology laboratories synoptically reported adenocarcinomas on colorectal resection specimens. Six of these laboratories reported < 100 adenocarcinomas synoptically and were excluded, leaving 11,719 adenocarcinomas from 35 laboratories (Table 1), of which 1,427 (12.2%) were PD.

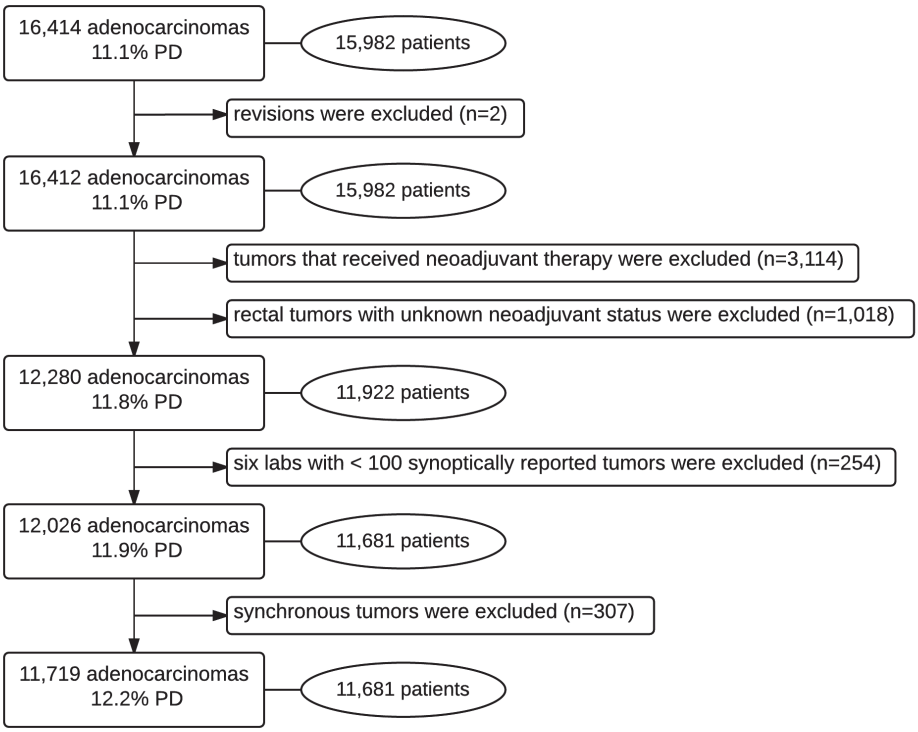


Figure 1. Flowchart of included adenocarcinomas.

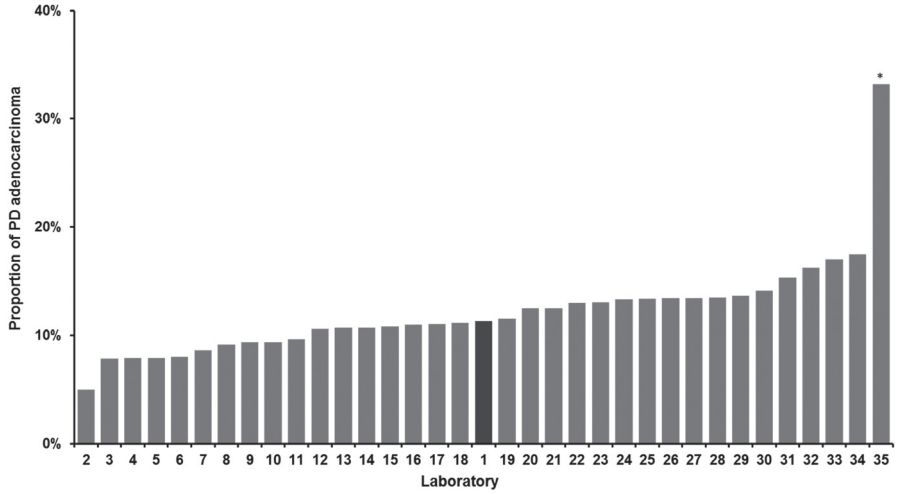


Figure 2. Proportions of PD adenocarcinomas per laboratory. The asterisk indicates the laboratory that significantly differed from the reference laboratory (laboratory 1) on univariable logistic regression analysis.

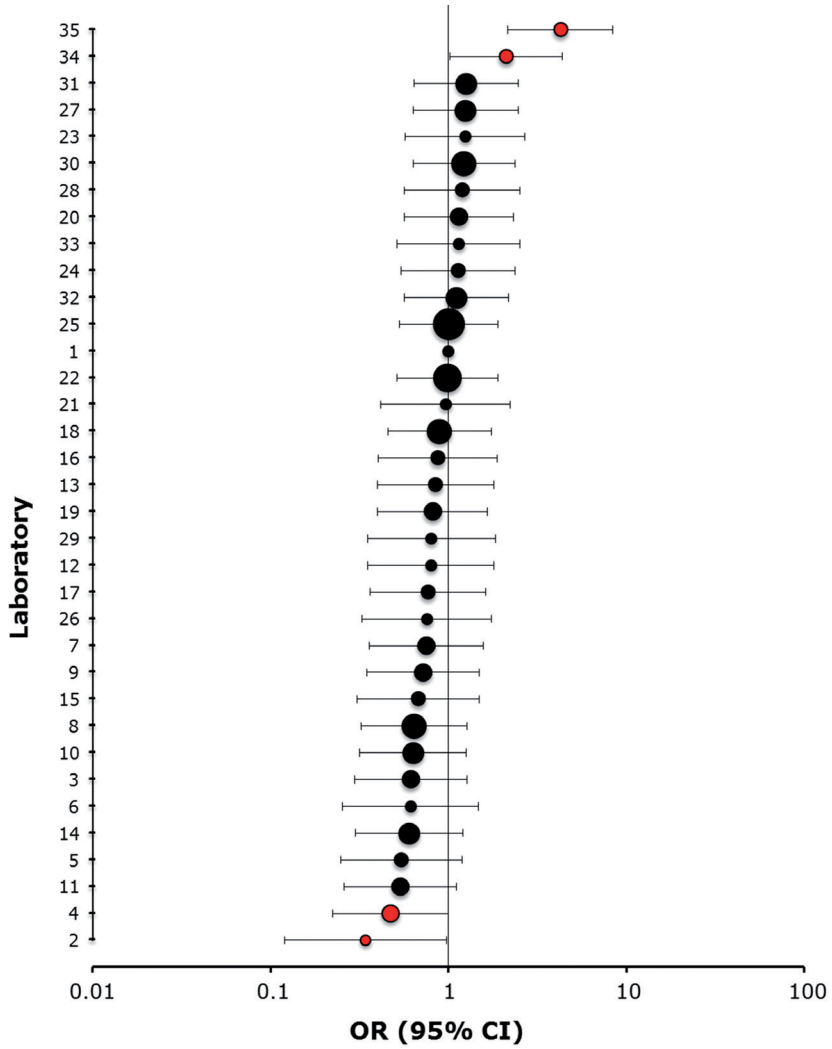


Figure 3. Forest plot showing adjusted odds ratios and 95% confidence intervals of laboratories regarding proportion of diagnosing PD adenocarcinoma. Dots sizes indicate the total number of synoptically reported adenocarcinomas per laboratory. The red dots indicate laboratories that were significantly aberrant with regard to proportion of diagnosing PD adenocarcinoma, compared to the reference laboratory (laboratory 1).

Table 1. Characteristics of the 11,719 included adenocarcinomas and association of clinicopathological variables with a diagnosis of poorly differentiated (PD) adenocarcinoma. Univariable (crude ORs) and multivariable (adjusted ORs) analyses.

	N (total)	N (%) with PD	Crude OR (95% CI)	Adjusted OR (95% CI)[b]
Age (years)				
< 60	1,794	215 (12.0)	1	
60 – 69	3,223	377 (11.7)	0.973 (0.814 – 1.163)	
70 – 79	4,032	478 (11.9)	0.988 (0.832 – 1.173)	
≥ 80	2,670	357 (13.4)	1.134 (0.946 – 1.358)	
Sex				
Male	6,170	608 (9.9)	1	1
Female	5,549	819 (14.8)	**1.584 (1.416 – 1.771)**	**1.602 (1.417 – 1.812)**
Localization				
Right-sided colon	5,531	1,008 (18.2)	**2.755 (2.136 – 3.554)**	**2.172 (1.655 – 2.849)**
Left-sided colon	5,067	320 (6.3)	0.833 (0.636 – 1.092)	**0.720 (0.541 – 0.958)**
Rectum	922	69 (7.5)	1	1
Unknown	199	30 (15.1)	**2.194 (1.386 – 3.474)**	**2.115 (1.294 – 3.458)**
Year of examination				
2010	1,484	184 (12.4)	1	
2011	2,372	291 (12.3)	0.988 (0.812 – 1.204)	
2012	3,604	431 (12.0)	0.960 (0.798 – 1.154)	
2013	4,260	521 (12.2)	0.984 (0.823 – 1.178)	
Tumor size	11,716	4.5 (2.0)[a]	**1.262 (1.231 – 1.294)**	**1.192 (1.159 – 1.227)**
T stage				
T1	586	15 (2.6)	**0.185 (0.110 – 0.310)**	**0.404 (0.237 – 0.688)**
T2	1,990	88 (4.4)	**0.325 (0.260 – 0.407)**	**0.523 (0.413 – 0.663)**
T3	7,351	915 (12.4)	1	1
T4	1,792	409 (22.8)	**2.080 (1.826 – 2.370)**	**1.414 (1.218 – 1.641)**
N stage				
N0	5,993	444 (7.4)	1	1
N1	3,797	563 (14.8)	**2.176 (1.907 – 2.482)**	**1.620 (1.396 – 1.881)**
N2	716	227 (31.7)	**5.802 (4.823 – 6.979)**	**3.146 (2.530 – 3.912)**
N3	123	67 (54.5)	**14.953 (10.351 – 21.600)**	**6.227 (4.106 – 9.442)**
Unknown	1,090	126 (11.6)	**1.634 (1.325 – 2.014)**	**1.391 (1.106 – 1.750)**
Angioinvasion				
No	7,708	625 (8.1)	1	1
Yes	2,536	608 (24.0)	**3.574 (3.162 – 4.039)**	**2.301 (1.979 – 2.675)**
Unknown	1,475	194 (13.2)	**1.716 (1.445 – 2.038)**	**1.706 (1.384 – 2.104)**

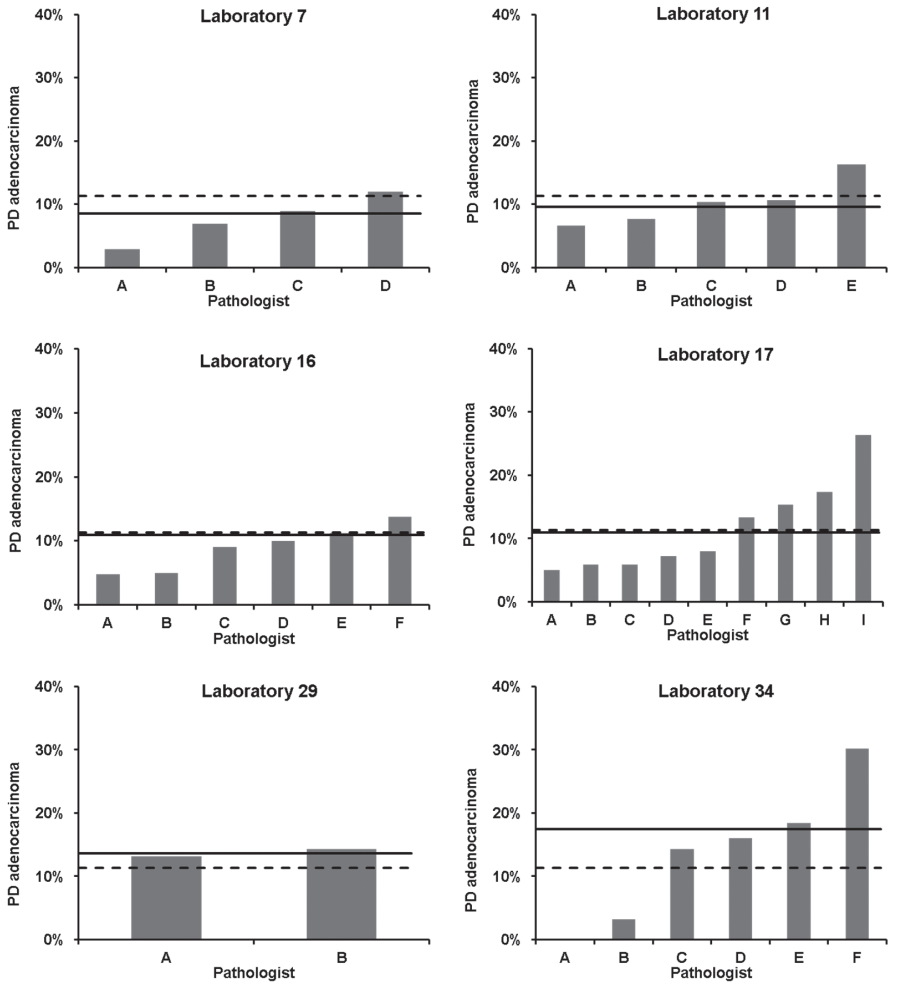
PD = poorly differentiated

OR = odds ratio

CI = confidence interval

[a] mean (SD)

[b] adjusted for sex, localization, tumor size, T stage, N stage, and angio-invasion



The number of synoptically reported resection specimens diagnosed with a colorectal adenocarcinoma varied from 120 to 1,280 per laboratory (median: 253). The proportion of PD adenocarcinoma varied from 5.0% to 33.2% between the laboratories (Figure 2). Univariable analysis showed that one laboratory reported a significantly higher proportion of PD adenocarcinomas (33.2%) than the other laboratories.

To adjust for differences in case mix between laboratories, association analyses of clinicopathological variables with PD adenocarcinoma were performed. Female sex, right-sided localization, increasing tumor size, T stage, N stage, and presence of angio-invasion were significantly associated with PD

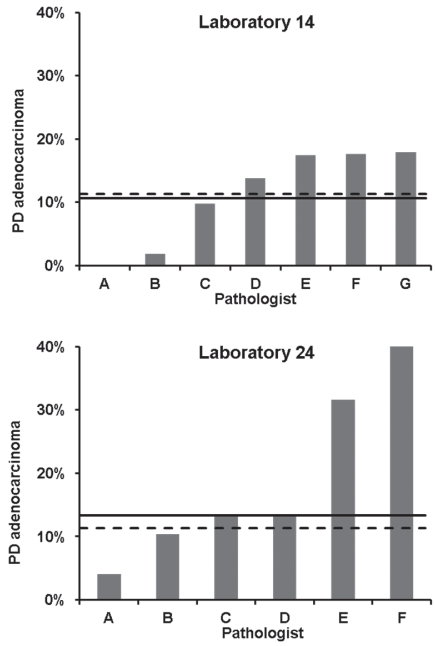


Figure 4. Intralaboratory variation between pathologists in eight Dutch pathology laboratories. Dashed line = proportion of PD adenocarcinoma in the reference laboratory (11.3%). Solid line = proportion of PD adenocarcinoma in the specific laboratory.

adenocarcinoma by univariable logistic regression analyses and included in multivariable analysis (Table 1). All variables remained significantly associated with a diagnosis of PD adenocarcinoma on multivariable analysis. Association of distant metastasis with PD adenocarcinoma was not analyzed, because this was often (67%) unknown to the pathologist.

In Figure 3 the association of pathology laboratory (laboratories 1 to 35) with a diagnosis of PD adenocarcinoma after multivariable analyses is plotted. Two laboratories reported a significantly lower proportion of PD adenocarcinoma and two laboratories a significantly higher proportion of PD adenocarcinoma than the reference laboratory.

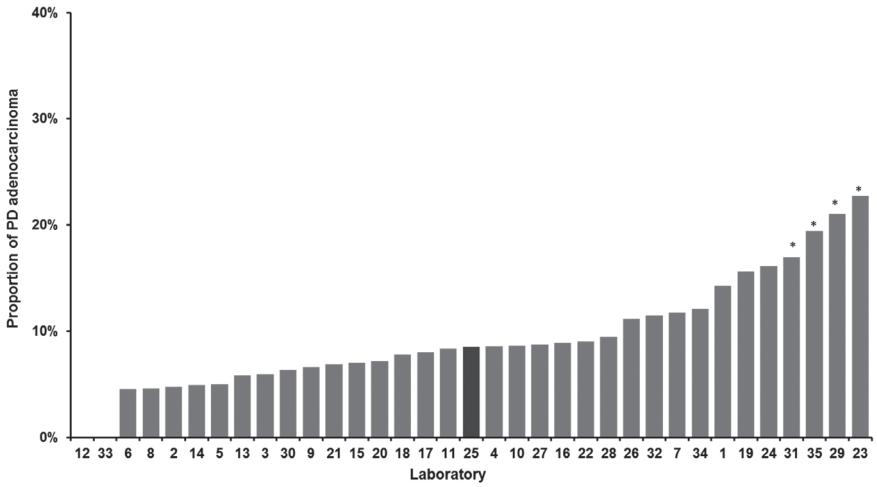


Figure 5. Proportions of PD adenocarcinomas per laboratory in the subgroup of patients who could have been eligible for adjuvant chemotherapy solely based on the differentiation grade (n=2,813 tumors). The asterisks indicate the laboratories that significantly differed from the reference laboratory (laboratory 25) on univariable logistic regression analysis.

Table 2. Results of 52 Dutch pathologists responding to our questionnaire concerning the manner of grading colorectal adenocarcinomas.

	General pathologist (n = 35)	GI pathologist (n = 17)	p-value
Laboratory; n (%)			0.194
Peripheral	30 (85.7)	12 (70.6)	
Academic	5 (14.3)	5 (29.4)	
Years of experience; n (%)			0.232
0-5	8 (22.9)	3 (17.6)	
6-10	10 (28.6)	4 (23.5)	
11-20	13 (37.1)	4 (23.5)	
>20	4 (11.4)	6 (35.3)	
How do you define the differentiation grade of colorectal adenocarcinomas? *			
Evaluation of architectural criteria; n (%)	35 (100.0)	17 (100.0)	na
Evaluation of cytological criteria; n (%)	14 (40.0)	3 (17.6)	0.107
How do you grade a specimen with heterogeneous differentiation?			0.215
Based on the predominant grade	4 (11.4)	2 (11.8)	
Based on the least differentiated grade	27 (77.1)	9 (52.9)	
Based on the overall percentage of gland formation	1 (2.9)	1 (5.9)	
I report the heterogeneity and the percentages of each grade	3 (8.6)	5 (29.4)	

na = not applicable

* = multiple answers possible

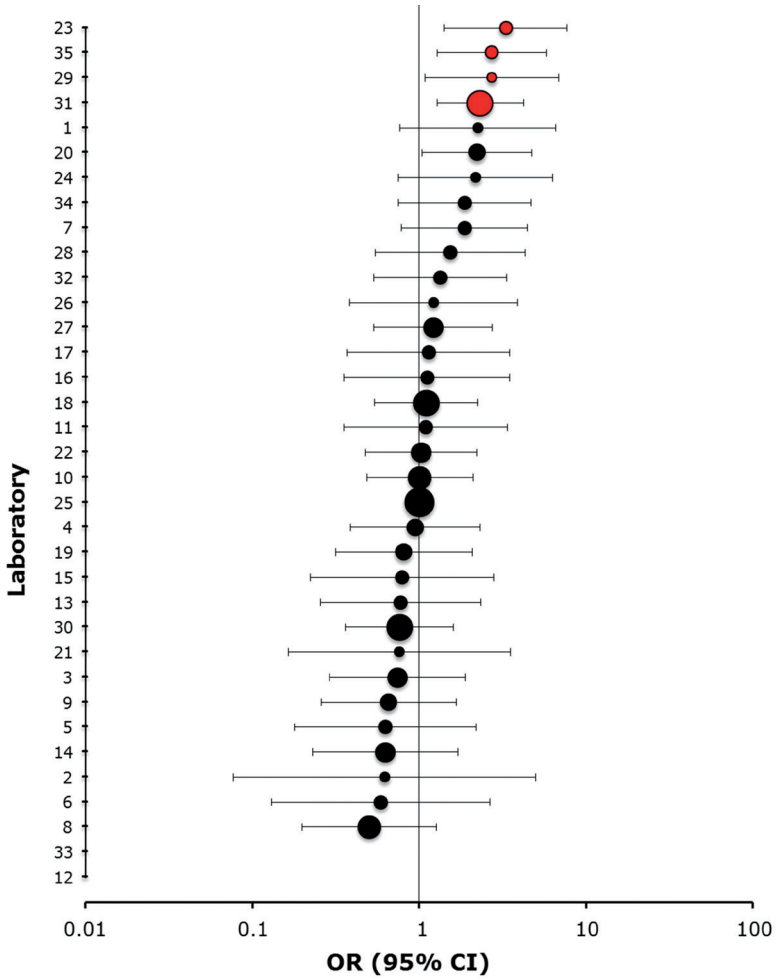


Figure 6. Forest plot showing adjusted odds ratios and 95% confidence intervals of laboratories regarding proportion of diagnosing PD adenocarcinoma in the subgroup of 2,812 patients (2,813 tumors) who could have been eligible for adjuvant chemotherapy solely based on the differentiation grade. Dots sizes indicate the total number of synoptically reported adenocarcinomas per laboratory. The red dots indicate laboratories that were significantly aberrant with regard to proportion of diagnosing PD adenocarcinoma, compared to the reference laboratory (laboratory 25). Laboratories 12 and 33 had not diagnosed PD adenocarcinoma in this subgroup of tumors and could therefore not be plotted.

Twenty-one laboratories gave their consent to disclose their laboratory name with the excerpts of pathology reports. Eight laboratories (7, 11, 14, 16, 17, 24, 29, and 34) participated in further research on variation in grading CRC between individual pathologists and disclosed which pathologist microscopically evaluated which adenocarcinomas. In these eight laboratories, a total of 54 pathologists synoptically reported colorectal adenocarcinomas within the study period. Forty-five pathologists (from eight laboratories) synoptically reported ≥ 10 adenocarcinomas. The number of included pathologists varied from 2 to 9 per laboratory (median: 6). In all but one laboratory (laboratory 29) considerable intralaboratory variation in the grading of colorectal adenocarcinomas was observed, even if laboratories scored average as a whole (Figure 4). Differences in proportion of PD adenocarcinomas between pathologists were not statistically significant, partly because of the relatively low numbers of cases per pathologist, except for pathologist 14A. At univariable logistic regression analysis, this pathologist, who had diagnosed 0 out of 89 adenocarcinomas as PD, significantly differed from the other pathologists in his or her laboratory. At multivariable logistic regression analysis this was no longer statistically significant.

We identified a subgroup of patients who could have been eligible for adjuvant chemotherapy solely based on the differentiation grade. Of the 4,415 stage II colon cancers, we excluded those with other high-risk factors (i.e. T4 (n=579), <10 LNs evaluated (n=353), clinical presence of ileus (n=251), clinical presence of perforation (n=119), and angio-invasion (n=300)) leaving a subgroup of 2,813 tumors. Two hundred fifty-eight out of these 2,813 tumors (9.2%) were PD adenocarcinomas. The number of adenocarcinomas within this subgroup varied from 21 to 329 (median: 61) per laboratory, and the proportion of PD adenocarcinomas varied from 0% to 22.7% (Figure 5). At univariable logistic regression analysis, four laboratories diagnosed significantly more PD adenocarcinoma than the reference laboratory. After adjustment for clinicopathological variables (sex, localization, and tumor size), the four laboratories remained significantly aberrant (Figure 6). Laboratory 35 was significantly aberrant in the total cohort as well, whereas the other three laboratories were only significantly aberrant in this subgroup and not in the total cohort.

Results of questionnaire

Table 2 summarizes the main results of the questionnaire among Dutch pathologists. Fifty-two of the ~320 Dutch pathologists (16%) responded. One third of the responding pathologists were specialized gastrointestinal (GI) pathologists. All pathologists stated that they use architectural criteria to determine the differentiation grade, and 17 (32.7%) stated that they use cytological criteria as well. Although not statistically significantly different, a higher proportion of general pathologists than GI pathologists uses cytological criteria. In case of heterogeneity in differentiation grade, 36 pathologists (69.2%) report the least differentiated grade. A higher proportion of GI pathologists than general pathologists reports that heterogeneity is present with the percentages of each grade, but this difference was also not statistically significant.

Furthermore, the pathologists reported seven different guidelines or books that they use as reference for the grading of CRC. The WHO classification [15] was mentioned by 18 pathologists, 12 pathologists mentioned the book Surgical Pathology of the GI Tract, Liver, Biliary Tract and Pancreas [24], and the Dutch guideline for colorectal cancer [22] and the book Gastrointestinal Pathology: An Atlas and Text [25] were each mentioned by five pathologists. Three other books [26-28] were each mentioned by one pathologist.

Discussion

Using data from a nationwide pathology database, we aimed to study variability in daily practice in the grading of colorectal adenocarcinomas between pathology laboratories in The Netherlands. In our large cohort of 11,719 resected colorectal adenocarcinomas, 12.2% of the cases were reported as PD. Considerable variation in grading was observed between 35 Dutch pathology laboratories, with proportions of PD adenocarcinomas varying between 5.0% and 33.2%. In a similar but smaller study (n=2,046), Blenkinsopp *et al* [13] reported an average proportion of 16% PD adenocarcinoma, with individual laboratories diagnosing 5% up to 30% of PD adenocarcinoma, but no further effort was made to adjust for clinicopathological variables. In literature, the reported proportion of PD adenocarcinomas varies widely (2% up to 25%) [29]. Large nationwide cohort studies from The Netherlands, reported PD adenocarcinoma in 13.4% and 16-19% of rectal and colon carcinomas, respectively [30-33].

The interlaboratory differences in differentiation grading in the present study could be only partly explained by differences in case mix between the laboratories, as after adjustment for these differences four (11%) laboratories still reported a significantly lower (n=2) or higher (n=2) proportion of PD adenocarcinomas. Selection of the median laboratory as reference laboratory is arbitrary, because the proportion of this laboratory might not necessarily indicate greater accuracy of diagnosis. However, comparison of the laboratories to this reference laboratory was considered the best possible approach to study interlaboratory variation.

Unfortunately, it was not possible to adjust for all variables possibly associated with a diagnosis of PD adenocarcinoma, such as presence of distant metastasis. This clinical variable was considered to be unreliably registered in pathology reports, because of the large number of "unknown". However, the variables deemed most important were included in the analyses and adjusted for: female gender, right-sided localization, tumor size, T stage, N stage, and angio-invasion. These findings largely correspond with previous studies [6,23,34-39]. Many studies reported a significant association between right-sided tumor localization and PD adenocarcinoma [23,34-39].

The observed variation in histological grading likely leads to variation in treatment decisions, as poor differentiation has prognostic importance and influences treatment choice. The present study focused on grading CRC on resection specimens, which has implications for selecting stage II colon cancer patients for adjuvant chemotherapy. In the subgroup of patients who could have been eligible for adjuvant chemotherapy solely based on the differentiation grade, considerable interlaboratory variation in grading was observed, with four laboratories diagnosing significantly more PD adenocarcinomas, increasing the number of high-risk stage II patients eligble for adjuvant therapy.

In addition to variation between laboratories, substantial variation between individual pathologists was observed as well in seven out of eight analyzed laboratories, even if a laboratory scored average as a whole. The results of our questionnaire revealed that pathologists are not unanimous with regard to using cytological criteria in differentiation grading. Also in literature, cytological criteria (cellular atypia, nuclear pleomorphism, high mitotic rate) are variably considered in determining differentiation grade [11,12,15-17]. Furthermore, although the majority of pathologists agreed on basing the differentiation grade on the least differentiated area, they were not unanimous on how to deal

with intratumoral heterogeneity in differentiation grade, which is also reflected by conflicting guidelines [15-17]. These reasons might explain part of the variation, emphasizing the need for better standardization of grading criteria.

The variation observed in this study focusing on the grading of CRC resection specimens is probably also present in the grading of CRC biopsies, which may have consequences for therapy selection in early rectal cancer patients, as poor differentiation is a contraindication for local resection. Unfortunately, due to the relatively small number of synoptically reported biopsies with CRC, we were unable to reliably compare the laboratories to confirm this hypothesis.

By the results of this study and individual feedback to laboratories, awareness will be raised among pathologists that differentiation grading of CRC is probably not performed optimally, which may be a first step towards standardization. Laboratories might also consider double reading by a peer in case of PD adenocarcinoma. Recent studies reported improved interobserver agreement and prognostic stratification by assessing the number of poorly differentiated cell clusters (defined as "solid cancer cell nests comprising ≥ 5 cancer cells and lacking a gland-like structure") instead of conventional grading [40-42]. This less subjective method might be promising to standardize grading, but it is unknown how suitable this method is for daily pathology practice.

In conclusion, this large nationwide cohort of CRC demonstrates considerable interlaboratory and intralaboratory variation in differentiation grading. Better standardization of grading criteria is needed for optimal determination of prognosis and treatment selection.

References

1. Freedman LS, Macaskill P, Smith AN. (1984) Multivariate analysis of prognostic factors for operable rectal cancer. Lancet 2(8405): 733-6.
2. Chapuis PH, Dent OF, Fisher R, Newland RC, Pheils MT, Smyth E, et al. (1985) A multivariate analysis of clinical and pathological variables in prognosis after resection of large bowel cancer. Br J Surg 72(9): 698-702.
3. Jass JR, Atkin WS, Cuzick J, Bussey HJ, Morson BC, Northover JM, et al. (1986) The grading of rectal cancer: historical perspectives and a multivariate analysis of 447 cases. Histopathology 10(5): 437-59.
4. Kobayashi H, Kotake K, Sugihara K. (2013) Prognostic scoring system for stage IV colorectal cancer: is the AJCC sub-classification of stage IV colorectal cancer appropriate,? Int J Clin Oncol 18(4): 696-703.
5. Greene FL, Stewart AK, Norton HJ. (2002) A new TNM staging strategy for node-positive (stage III) colon cancer: an analysis of 50,042 patients. Ann Surg 236(4): 416-21.
6. Derwinger K, Kodeda K, Bexe-Lindskog E, Taflin H. (2010) Tumour differentiation grade is associated with TNM staging and the risk of node metastasis in colorectal cancer. Acta Oncol 49(1): 57-62.
7. Cho YB, Chun HK, Yun HR, Kim HC, Yun SH, Lee WY. (2009) Histological grade predicts survival time associated with recurrence after resection for colorectal cancer. Hepatogastroenterology 56(94-95): 1335-40.
8. Heafner TA, Glasgow SC. (2014) A critical review of the role of local excision in the treatment of early (T1 and T2) rectal tumors. J Gastrointest Oncol 5(5): 345-52.
9. Maeda K, Koide Y, Katsuno H. (2014) When is local excision appropriate for "early" rectal cancer? Surg Today 44(11): 2000-14.
10. Benson AB 3rd, Schrag D, Somerfield MR, Cohen AM, Figueredo AT, Flynn PJ, et al. (2004) American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. J Clin Oncol 22(16): 3408-19.
11. Compton CC, Fielding LP, Burgart LJ, Conley B, Cooper HS, Hamilton SR, et al. (2000) Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 124(7): 979-94.
12. Chandler I, Houlston RS. (2008) Interobserver agreement in grading of colorectal cancers-findings from a nationwide web-based survey of histopathologists. Histopathology 52(4): 494-9.
13. Blenkinsopp WK, Stewart-Brown S, Blesovsky L, Kearney G, Fielding LP. (1981) Histopathology reporting in large bowel cancer. J Clin Pathol 34(5): 509-13.
14. Thomas GD, Dixon MF, Smeeton NC, Williams NS. (1983) Observer variation in the histological grading of rectal carcinoma. J Clin Pathol 36(4): 385-91.
15. Bosman FT, Carneiro F, Hruban RH, Theise ND. (2010) Tumours of the colon and rectum, in WHO classification of tumours of the digestive system.
16. Jass JR, O'Brien J, Riddell RH, Snover DC; Association of Directors of Anatomic and Surgical Pathology. (2008) Recommendations for the reporting of surgically resected specimens of colorectal carcinoma: Association of Directors of Anatomic and Surgical Pathology. Am J Clin Pathol 129(1): 13-23.
17. Loughrey MB, Quirke P, Shepherd NA. (2014) Dataset for colorectal cancer histopathology reports. Available from: www.rcpath.org.

18. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, et al. (2007) Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 29(1): 19-24.
19. Srigley J, Lankshear S, Brierley J, McGowan T, Divaris D, Yurcan M, et al. (2013) Closing the quality loop: facilitating improvement in oncology practice through timely access to clinical performance indicators. J Oncol Pract 9(5): e255-61.
20. Williams CL, Bjugn R, Hassell LA. (2015) Current status of discrete data capture in synoptic surgical pathology and cancer reporting. Pathol Lab Med Int 7: 11-22.
21. Nagtegaal ID, Marijnen CA, Kranenbarg EK, Mulder-Stapel A, Hermans J, van de Velde CJ, et al. (2002) Short-term preoperative radiotherapy interferes with the determination of pathological parameters in rectal cancer. J Pathol 197(1): 20-7.
22. Dutch guideline for colorectal cancer. Available from: http://www.oncoline.nl/index.php?pagina=/richtlijn/item/pagina.php&id=37094&richtlijn_id=933.
23. Ishihara S, Watanabe T, Akahane T, Shimada R, Horiuchi A, Shibuya H, et al. (2012) Tumor location is a prognostic factor in poorly differentiated adenocarcinoma, mucinous adenocarcinoma, and signet-ring cell carcinoma of the colon. Int J Colorectal Dis 27(3): 371-9.
24. Odze RD, Goldblum JR. (2014) Surgical Pathology of the GI Tract, Liver, Biliary Tract and Pancreas.
25. Fenoglio CM, Noffsinger AE, Stemmermann GN, Lantz PE, Isaacson PG. (2007) Gastrointestinal Pathology: An Atlas and Text.
26. Greenson J, Polydorides AD, Montgomery EA, Lauwers G, Purdy J, Lamps L, et al. (2009) Diagnostic Pathology: Gastrointestinal.
27. Shepherd NA, Warren BF, Williams GT, Greenson JK, Lauwers GY, Novelli MR. (2013) Morson and Dawson's Gastrointestinal Pathology.
28. Iacobuzio-Donahue C, Montgomery E. (2011) Gastrointestinal and Liver Pathology.
29. Shinji S, Naito Z, Ishiwata T, Yoko M, Seya T, Tajiri T. (2009) Poorly Differentiated Colorectal Adenocarcinoma (Methodology), in Colorectal Cancer.
30. Elferink MA, Siesling S, Lemmens VE, Visser O, Rutten HJ, van Krieken JH, et al. (2011) Variation in lymph node evaluation in rectal cancer: a Dutch nationwide population-based study. Ann Surg Oncol 18(2): 386-95.
31. Elferink MA, Siesling S, Visser O, Rutten HJ, van Krieken JH, Tollenaar RA, et al. (2011) Large variation between hospitals and pathology laboratories in lymph node evaluation in colon cancer and its impact on survival, a nationwide population-based study in the Netherlands. Ann Oncol 22(1): 110-7.
32. Lemmens VE, van Lijnschoten I, Janssen-Heijnen ML, Rutten HJ, Verheij CD, Coebergh JW. (2006) Pathology practice patterns affect lymph node evaluation and outcome of colon cancer: a population-based study. Ann Oncol 17(12): 1803-9.
33. van Erning FN, Crolla RM, Rutten HJ, Beerepoot LV, van Krieken JH, Lemmens VE. (2014) No change in lymph node positivity rate despite increased lymph node yield and improved survival in colon cancer. Eur J Cancer 50(18): 3221-9.
34. Takeuchi K, Kuwano H, Tsuzuki Y, Ando T, Sekihara M, Hara T, et al. (2004) Clinicopathological characteristics of poorly differentiated adenocarcinoma of the colon and rectum. Hepatogastroenterology 51(60): 1698-702.

35. Meguid RA, Slidell MB, Wolfgang CL, Chang DC, Ahuja N. (2008) Is there a difference in survival between right- versus left-sided colon cancers? Ann Surg Oncol 15(9): 2388-94.
36. Benedix F, Kube R, Meyer F, Schmidt U, Gastinger I, Lippert H, et al. (2010) Comparison of 17,641 patients with right- and left-sided colon cancer: differences in epidemiology, perioperative course, histology, and survival. Dis Colon Rectum 53(1): 57-64.
37. Benedix F, Schmidt U, Mroczkowski P, Gastinger I, Lippert H, Kube R, et al. (2011) Colon carcinoma--classification into right and left sided cancer or according to colonic subsite?--Analysis of 29,568 patients. Eur J Surg Oncol 37(2): 134-9.
38. Weiss JM, Pfau PR, O'Connor ES, King J, LoConte N, Kennedy G, et al. (2011) Mortality by stage for right- versus left-sided colon cancer: analysis of surveillance, epidemiology, and end results--Medicare data. J Clin Oncol 29(33): 4401-9.
39. Papagiorgis PC, Zizi AE, Tseleni S, Oikonomakis IN, Sofras L, Nikiteas NI. (2012) Worse histological grade of proximal colorectal tumors and its relation with stage. J BUON 17(1): 79-84.
40. Ueno H, Kajiwara Y, Shimazaki H, Shinto E, Hashiguchi Y, Nakanishi K, et al. (2012) New criteria for histologic grading of colorectal cancer. Am J Surg Pathol 36(2): 193-201.
41. Barresi V, Reggiani Bonetti L, Branca G, Di Gregorio C, Ponz de Leon M, Tuccari G. (2012) Colorectal carcinoma grading by quantifying poorly differentiated cell clusters is more reproducible and provides more robust prognostic information than conventional grading. Virchows Arch 461(6): 621-8.
42. Kim JW, Shin MK, Kim BC. (2015) Clinicopathologic impacts of poorly differentiated cluster-based grading system in colorectal carcinoma. J Korean Med Sci 30(1): 16-23.

Supplemental file. Questionnaire to pathologists.

01. What is the name of your pathology laboratory?

02. Is this an academic or peripheral pathology laboratory?
• Academic
• Peripheral

03. Where (institution / town / country) were you trained as a pathologist?

04. How many years of experience as a pathologist do you have?
• 0-5 years
• 6-10 years
• 11-20 years
• 20 years

05. Would you consider yourself a specialized gastrointestinal (GI) pathologist?
• Yes
• No

06. What percentage of your daily work do you spend on GI pathology?
• 0-25%
• 26-50%
• 51-75%
• 76-100%

07. How many pathologists are employed in your laboratory?

08. How many pathologists in your laboratory perform GI diagnostics?

09. How many specialized GI pathologists are employed in your laboratory?

10. What percentage of colon / rectal resections do you report synoptically?
• 0%
• 1-25%
• 26-50%
• 51-75%
• 76-99%
• 100%

11. What reasons do you have to not report a case synoptically?

12. How do you define the differentiation grade of colorectal adenocarcinomas? Multiple answers possible.
• Evaluation of architectural criteria
• Evaluation of cytological criteria
• Other, namely ..

13. How do you grade a specimen with heterogeneous differentiation?
• Based on the predominant grade
• Based on the least differentiated grade
• I report the heterogeneity and the percentages of each grade
• Other, namely ..

14. Which book/article/guideline do you use as a reference for the grading of colorectal adenocarcinomas?

Improved quality of patient care through routine second review of histopathology specimens prior to multidisciplinary meetings

Chantal C.H.J. Kuijpers
Gerard Burger
Shaimaa Al-Janabi
Stefan M. Willems
Paul J. van Diest
Mehdi Jiwa

Submitted to Journal of Clinical Pathology

Abstract

Introduction

Double reading may be a valuable tool for improving quality of patient care by identifying diagnostic errors before final sign-out, but standard double reading would significantly increase costs of pathology. We assessed the added value of intradepartmental routine double reading of histopathology specimens prior to multidisciplinary meetings.

Methods

Diagnoses, treatment plans and prognoses of patients are often discussed at multidisciplinary meetings. As part of daily routine, all pathology specimens to be discussed at upcoming multidisciplinary meetings undergo prior intradepartmental double reading. We identified all histopathology specimens from 2013 that underwent such double reading and determined major and minor discordance rates based on clinical relevance between initial and consensus sign-out diagnosis.

Results

We included 6,796 histopathology specimens that underwent double reading, representing ~8% of all histopathology cases at our institution in 2013. Double reading diagnoses were concordant in 6,566 specimens (96.6%). Major and minor discordances were observed in 60 (0.9%) and 170 (2.5%) specimens, respectively. Urology specimens encountered significantly more discordances than the other tissues of origin, Gleason grading of prostate cancer biopsies being the most frequent diagnostic problem. Furthermore, pre-malignant and malignant cases showed significantly higher discordance rates than the rest. The vast majority (90%) of discordances represented changes within the same diagnostic category, e.g. malignant to malignant.

Conclusions

Routine double reading of histopathology specimens prior to multidisciplinary meetings prevents diagnostic errors. It resulted in about 1% discordant diagnoses of potential clinical significance, indicating that second review is worthwhile in terms of patient safety and quality of patient care.

Introduction

There is a growing awareness that pathology diagnosis is not infallible and that diagnostic errors may lead to under- or overtreatment and thereby compromise patient safety. Double reading, i.e. second assessment of pathology cases, is a potentially valuable tool for reducing diagnostic errors and thereby improving the quality of patient care. It may reveal inaccurate diagnoses that otherwise might have led to improper patient management. In response to the Institute of Medicine report 'To err is human; building a safer health system' from 1999 [1], the American Society for Clinical Pathology (ASCP) recognized double reading of pathology cases as a key aspect in the assurance of patient safety [2]. The ASCP recommends to consider double reading in highly critical or significant cases, problem prone cases and cases suggested for review by clinicians [2].

Numerous studies have assessed the value of double reading in diagnostic surgical pathology, and reported major diagnostic disagreement rates of 0.1% to 28%, mainly depending on the organ system studied, and the definition of disagreement [3-29]. Intradepartmental second review resulted in 0.1% to 2.8% disagreements with potential clinical significance [3,5,6,15,18,19].

At our institution, pathology specimens of patients who will be discussed at the upcoming multidisciplinary meetings undergo prior intradepartmental double reading, most of the time by an expert pathologist. In this study we retrospectively assessed the added value of this double reading strategy in improving diagnostic accuracy in a large one-year cohort of reviewed histopathology cases by assessing (degree of) concordance between the initial and the consensus sign-out diagnoses.

Methods

Routine double reading prior to multidisciplinary meetings

Symbiant provides pathology services for six public health care non-academic teaching hospitals in the province of North Holland. In these hospitals, diagnoses, treatment plans and prognoses of patients are discussed at multidisciplinary meetings. Cases to be discussed at these multidisciplinary meetings vary from all cases in the specific medical discipline of the multidisciplinary meeting to only exceptional cases which differ from routine guidelines.

At Symbiant's three pathology laboratories (Alkmaar Medical Centre, Westfriesgasthuis Hoorn and Zaandam Medical Centre), all pathology specimens to be discussed at the upcoming multidisciplinary meetings, routinely undergo prior intradepartmental double reading at preparation for each multidisciplinary meeting, based on lists from the responsible clinician. The second review is performed in a non-blinded fashion by the pathologist who will attend the meeting, most of the time an expert pathologist in the field of that multidisciplinary meeting. Cases reported by a resident were routinely checked by a pathologist before sign-out, which was not considered as double reading.

The pathologists register the results of the second review (i.e. concordant or discordant) in a 'hidden' section of the pathology reporting system, which is not visible to clinicians. Discordant diagnoses are fed back to and discussed with the pathologist who made the initial pathology report. Consensus diagnoses are then formulated, achieved either by unanimity or by majority, following consensus joint review by the first and second pathologists, or after consulting other colleague pathologists. The final report only contains the consensus sign-out diagnosis. Previous versions of the report are stored in a separate section of the pathology reporting system, unavailable to clinicians as well, highlighting the changes made.

Data extraction

All pathology reports of histopathology specimens from 2013 that underwent routine double reading prior to a multidisciplinary meeting were identified by automatic search for a specific code (i.e. internal revision). The initial diagnoses and the consensus sign-out diagnoses were extracted from the 'hidden' previous version of the report and the final report, respectively. Specimens for which we were not able to determine the diagnostic concordance of the double reading, due to the unavailability of the original pathology reports, were excluded from analysis.

The study cases were sorted according to their tissue of origin and diagnostic categories. Diagnoses were categorized as no abnormalities, benign (including reactive, inflammation, benign tumor, and other benign abnormalities such as hemochromatosis and amyloidosis), uncertain malignant potential, pre-malignant, suspicious for malignancy, malignant (including primary and metastatic malignancy), and no diagnosis.

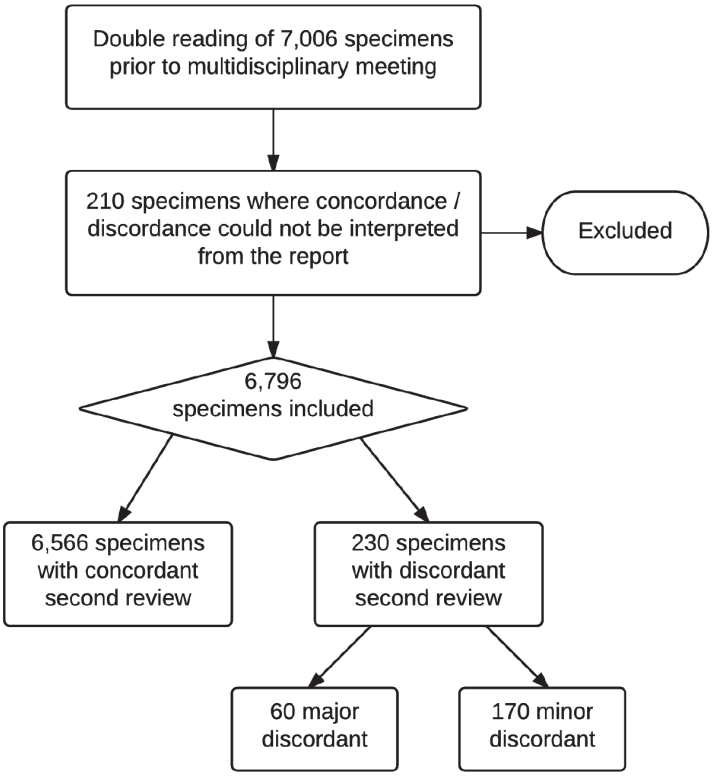


Figure 1. Flowchart of specimens included in this study.

Assessment of concordance between initial and second review diagnoses

Initial and consensus sign-out diagnoses were retrospectively compared to determine (the degree of) concordance after double reading. A discordant double reading was categorized as minor discordant or major discordant based on clinical significance. Minor discordance was defined as a change in diagnosis that would not alter patient management or prognosis, whereas major discordance was defined as a changed diagnosis with a potential effect on patient management or prognosis.

Statistical analysis

Statistical analysis was performed using SPSS Statistics for Windows, version 20.0 (Armonk, NY, USA: IBM Corp.). The percentages of concordance, major discordance and minor discordance between initial and consensus sign-out diagnoses with their 95% confidence intervals (CI) were calculated. The Chi-

squared test statistic was used to compare the percentages of overall, major and minor discordance between each individual tissue of origin and the rest, and between each individual diagnostic category and the rest. Furthermore, the categories of initial diagnosis and consensus sign-out diagnosis were compared using the unweighted Cohen's Kappa coefficient. A Kappa-value of 0.00–0.20 indicates a slight agreement, 0.21–0.40 a fair agreement, 0.41–0.60 a moderate agreement, 0.61–0.80 a substantial agreement, and 0.81–1 (almost) perfect agreement [30,31]. P-values < 0.05 were considered statistically significant.

Results

Figure 1 illustrates a flowchart of the histopathology specimens originating from 2013 that underwent double reading. We included 6,796 unique histopathology specimens from 4,388 patients, representing ~8% of all histopathology cases from 2013. The double reading of these specimens was performed by 22 different pathologists in total. The tissue of origin and the initial diagnoses of the included histopathology specimens are summarized in Table 1. The bulk (56.7%) of the cases that underwent second review originated from the breast and the gastrointestinal tract. With regard to initial diagnosis, the majority (59.5%) of the cases that underwent second review were malignant. Consensus sign-out diagnoses were concordant with initial diagnoses in 6,566 specimens (96.6%; 95% CI: 0.962-0.970), and discordant in 230 specimens (3.4%), with minor and major discordance in 170 specimens (2.5%; 95% CI: 0.021-0.029) and 60 specimens (0.9%; 95% CI: 0.007-0.011), respectively.

Table 2 summarizes the percentages of major, minor and overall discordance observed per tissue of origin. The overall discordance rates of bone and joint, lymph node, and breast specimens were significantly lower, which was mainly explained by a lower frequency of minor discordances. Specimens from the gastrointestinal tract encountered significantly less major discordances than the rest. Urology specimens encountered significantly more discordances than the rest, Gleason grading of prostate cancer biopsies being the most frequent diagnostic problem. Of the 402 prostate specimens, 16 (4.0%) major discordances were encountered. Without second review of these prostate cancer biopsies, patients would probably have been undertreated (n=7) or overtreated (n=7) because of underestimation or overestimation of the Gleason grade, respectively. Treatment of the other two patients would probably have been suboptimal as well (a change from Gleason grade not assessable into Gleason

Table 1. Overview of the 6,796 included histopathology specimens undergoing routine double reading.

Diagnosis Tissue of origin	NA	Benign	UMP	PM	Suspicious for malignancy	Malignant	ND	Total (%)
Breast	95	753	16	354	4	1,400	19	2,641 (38.9)
Gastrointestinal	27	116	4	169	14	882	2	1,214 (17.9)
Lymph node *	460	64	0	1	1	308	6	840 (12.4)
Dermatopathology	10	249	4	47	2	340	5	657 (9.7)
Urology	5	21	1	16	6	582	2	633 (9.3)
Female genital tract	15	25	8	43	1	140	2	234 (3.4)
Pulmonary	7	32	0	5	10	169	7	230 (3.4)
Head and neck	10	14	0	9	4	80	0	117 (1.7)
Bone and joint	21	11	2	9	2	69	3	117 (1.7)
Soft tissue	5	8	1	0	0	32	3	49 (0.7)
Central nervous system	0	18	0	0	0	27	0	45 (0.7)
Endocrine	0	4	0	0	0	13	2	19 (0.3)
Total (%)	655 (9.6)	1,315 (19.3)	36 (0.5)	653 (9.6)	44 (0.6)	4,042 (59.5)	51 (0.8)	6,796

* The category of lymph nodes includes 1 case of spleen
NA = no abnormalities; UMP = uncertain malignant potential; PM = pre-malignant; ND = no diagnosis.

Table 2. Percentages of discordance observed after double reading per tissue of origin.

Tissue of origin	N	Discordance; n (%)					
		Major	p-value	Minor	p-value	Overall	p-value
Endocrine	19	0	0.680	0	0.484	0	0.414
Bone and joint	117	0	0.303	0	0.081	0	**0.041**
Lymph node	840	5 (0.6)	0.341	7 (0.8)	**0.001**	12 (1.4)	**0.001**
Soft tissue	49	0	0.507	1 (2.0)	0.836	1 (2.0)	0.602
Central nervous system	45	1 (2.2)	0.335	0	0.281	1 (2.2)	0.685
Breast	2,641	20 (0.8)	0.378	48 (1.8)	**0.004**	68 (2.6)	**0.003**
Gastrointestinal	1,214	4 (0.3)	**0.023**	30 (2.5)	0.941	34 (2.8)	0.215
Female genital tract	234	1 (0.4)	0.448	8 (3.4)	0.360	9 (3.8)	0.691
Pulmonary	230	1 (0.4)	0.460	8 (3.5)	0.335	9 (3.9)	0.652
Head and neck	117	1 (0.9)	0.974	4 (3.4)	0.522	5 (4.3)	0.592
Dermatopathology	657	9 (1.4)	0.160	21 (3.2)	0.230	30 (4.6)	0.078
Urology	633	18 (2.8)	**< 0.0001**	43 (6.8)	**< 0.0001**	61 (9.6)	**< 0.0001**
Total	6,796	60 (0.9)		170 (2.5)		230 (3.4)	

grade 7, and a change from tumor only present in the left sided biopsies into right sided biopsies containing tumor as well).

Table 3 summarizes the percentages of discordance observed per initial diagnostic category. Second review of specimens with no abnormalities and with a benign diagnosis resulted in significantly lower discordance rates than the other diagnostic categories. For the benign cases, both major and minor discordance rates were significantly lower. Malignant and pre-malignant cases encountered significantly more overall discordances than the rest. For the malignant cases, this was mainly explained by a higher proportion of minor discordances, whereas pre-malignant cases encountered a significantly higher proportion of major discordances than the rest.

Table 4 presents the correlation between diagnostic categories of initial and second review diagnoses. The diagnostic category remained the same for 6,773 specimens (99.7%; Kappa 0.99, $p < 0.0001$). A change in diagnostic category was observed for 23 specimens (0.3%), which represents 10.0% of overall discordances. The other discordances represented a change within the same diagnostic category. These included changes in histological subtype, margin status, grade, TNM stage, tumor diameter, tumor percentage, the number of malignant biopsies, the number of (metastatic) lymph nodes without changing TNM stage, the number of mitoses without changing TNM stage, or HER-2 oncogene status. Other discordances within the same diagnostic category were missed unsolicited findings (e.g. an additional *in situ* lesion in a case of carcinoma), the need for additional material, typographical errors or different terminology used, a change in clinical information, an incomplete pathology report, and other minor differences in interpretation without clinical relevance. All malignant cases retained the diagnosis malignancy after second review. In these cases, only changes within the same diagnostic category, i.e. malignant, were observed (e.g. changes in grade or margin status).

Discussion

This study assessed the added value of intradepartmental routine double reading of histopathology specimens prior to discussion at multidisciplinary meeting. Initial and second review diagnoses were concordant in 96.6% of cases. Major discordances with a potential clinical significance were observed in 60 cases (0.9%). The vast majority (90%) of discordances were changes within the same diagnostic category, rather than changes into another diagnostic category. Our

Table 3. Percentages of discordance observed after double reading per tissue of origin.

Tissue of origin	N	Discordance; n (%)					
		Major	p-value	Minor	p-value	Overall	p-value
No diagnosis	51	0	0.499	0	0.251	0	0.180
No abnormalities	655	2 (0.3)	0.096	7 (1.1)	**0.014**	9 (1.4)	**0.003**
Benign	1,315	5 (0.4)	**0.030**	19 (1.4)	**0.006**	24 (1.8)	**< 0.0001**
Suspicious for malignancy	44	1 (2.3)	0.323	0	0.286	1 (2.3)	0.682
Uncertain malignant potential	36	0	0.570	1 (2.8)	0.915	1 (2.8)	0.840
Malignant	4,042	41 (1.0)	0.160	120 (3.0)	**0.003**	161 (4.0)	**0.001**
Pre-malignant	653	11 (1.7)	**0.021**	23 (3.5)	0.079	34 (5.2)	**0.007**
Total	6,796	60 (0.9)		170 (2.5)		230 (3.4)	

Table 4. Correlation between diagnostic categories of initial diagnoses and consensus sign-out diagnoses.

	Second review diagnosis							
	NA	Benign	UMP	PM	Suspicious for malignancy	Malignant	ND	Total
NA	650	0	1	0	0	4	0	655
Benign	0	1,310	0	4	0	1	0	1,315
UMP	0	0	35	0	0	1	0	36
PM	1	0	2	642	3	5	0	653
Suspicious for malignancy	0	1	0	0	43	0	0	44
Malignant	0	0	0	0	0	4,042	0	4,042
ND	0	0	0	0	0	0	51	51
Total	651	1,311	38	646	46	4,053	51	6,796

NA = no abnormalities; UMP = uncertain malignant potential; PM = pre-malignant; ND = no diagnosis.

observed 0.9% major discordance rate falls within the range of discordance rates (0.1% to 2.8%) with potential clinical significance described in literature concerning intradepartmental second review of surgical pathology specimens [3,5,6,15,18,19].

Urology specimens encountered an overall discordance rate of 9.6% and a major discordance rate of 2.8%, both significantly higher than the other tissues of origin. Weydert *et al* [18] also found most major discordances in the urologic

tract. Contrary to our study, Lind *et al* [6] described no major discordances in genitourinary biopsies, while the highest rates of major discordance were observed in pulmonary and head and neck specimens. The discordances observed in urology specimens particularly concerned prostate biopsies, especially a change in Gleason grade. It is well-known that the interobserver agreement of Gleason grading is unsatisfactory [32-33], although accurate grading may be essential for optimal treatment selection. Several studies comparing prostate biopsy first and second review diagnoses found a change in Gleason score in approximately 40% [10, 17]. Furthermore, Wurzer *et al* [10] showed that in 5% of patients second review resulted in treatment modifications, either due to misdiagnosis of prostate cancer, change in Gleason grade or missed presence of perineural invasion.

In the present study, significantly higher discordance rates were observed in pre-malignant and malignant cases compared to the other diagnostic categories. Still, five cases (0.4%) that were initially diagnosed as benign had a discrepancy after double reading with potential clinical significance, of which three originated from the skin. Romanoff *et al* [34] studied second review of breast biopsies, and found that benign biopsies were more likely to result in a discrepancy than malignant cases. Troxel *et al* [35] demonstrated that most pathology malpractice claims resulted from false-negative diagnoses, especially missed melanomas. These results indicate that ideally a selection of problem-prone benign cases should undergo double reading as well. Benign skin and breast lesions may therefore be interesting for second review. Most of the discordances concerned changes within the same diagnostic category (e.g. benign to benign), rather than changes into another diagnostic category (e.g. benign to malignant). Contrary, in the study of Weydert *et al* [18], the majority of the reported major discordances were due to changes into another diagnostic category.

The double reading strategy described here was performed in addition to usual upfront consultation that might have already been performed at the time of second review, but still resulted in a substantial number of changed diagnoses: 60 major discordances per year where double reading possibly led to a change in patient management. Very recently published recommendations of the College of American Pathologists (CAP) and the Association of Directors of Anatomic and Surgical Pathology (ADASP) for the review of pathology cases emphasize this as well [36]. They recommend that anatomic pathologists develop procedures for the review of selected pathology cases, perform case

reviews in a timely manner (prior to definitive treatment), document case review procedures that are relevant to their practice setting, continuously monitor and document the results of case reviews, and take steps to improve agreement if pathology case reviews show poor agreement within a defined case type [36]. We fully agree with these recommendations, which we deem to be appropriate, feasible and necessary for good pathology practice, depending on the practice setting.

Although the CAP/ADASP recommend that double reading is performed prospectively (i.e. before sign-out), our retrospective double reading strategy is considered timely as well according to the CAP/ADASP guidelines, as it is still performed prior to definitive treatment. Second review was performed after sign-out, but it is clear to the clinicians that the definitive pathology diagnosis is rendered at the multidisciplinary meeting. Prospective second review of the same specimens would probably have been less rework in corrections without changing the results of second review. However, routine second review before sign-out increases turnaround times, which may especially be a problem in cases where small turnaround times are warranted.

There is no single best double reading strategy, as this varies per practice setting. Double reading strategies can either comprise reviewing every (histo)pathology case, a fixed percentage of randomly selected cases or known problem prone tissues of origin. Literature is not consistent on which specimen types are most problem prone [36], and this may also vary between laboratories. Therefore, as part of quality control, pathology laboratories might want to determine annually for their specific case mix which specimens / tracts / organs are problem prone, and focus their double reading strategy on these areas. The cases included in this study were, on average, probably somewhat more difficult than the routine of the department, which contains for example a high percentage of relatively easy skin and gastrointestinal biopsies, of which only a small proportion was reviewed prior to and discussed at a multidisciplinary meeting.

A weakness of this study is that we did not assess on an individual basis whether a discrepancy would lead to an actual change in patient management. The expected effect on prognosis and patient management was determined by specialized pathologists according to guidelines and experience. Furthermore, although we cannot be totally sure, we considered the consensus sign-out diagnosis, involving a specialized pathologist, to be the correct diagnosis.

We consider the increase in workload and costs of double reading and associated administrative actions worth the effort given the large added value in terms of minimizing diagnostic errors to improve quality of patient care and to prevent claims due to diagnostic error. The additional pathologists' time spent on double reading should be incorporated into the cost price of pathology service.

In conclusion, routine double reading of histopathology specimens prior to multidisciplinary meetings resulted in about 1% discordant diagnoses of potential clinical significance, indicating that second review is worthwhile in terms of patient safety and quality of patient care. An adequate quality improvement program requires some routine double reading, which should be financially covered, independent of the funding model.

References

1. Kohn LT, Corrigna JM, Donaldson MS (1999) To Err is Human; Building a Safer Health System, Washington DC: Institute of Medicine, National Academy Press
2. Tomaszewski JE, Bear HD, Connally JA, Epstein JI, Feldman M, Foucar K, et al. (2000) Consensus conference on second opinions in diagnostic anatomic pathology. Who, What, and When. Am J Clin Pathol 114(3): 329-35.
3. Whitehead ME, Fitzwater JE, Lindley SK, Kern SB, Ulirsch RC, Winecoff WF 3rd. (1984) Quality assurance of histopathologic diagnoses: a prospective audit of three thousand cases. Am J Clin Pathol 81(4): 487-91.
4. Ramsay AD, Gallagher PJ. (1992) Local audit of surgical pathology. 18 month's experience of peer review-based quality assessment in an English teaching hospital. Am J Surg Pathol 16(5): 476-82.
5. Safrin RE, Bark CJ. (1993) Surgical pathology sign-out. Routine review of every case by a second pathologist. Am J Surg Pathol 17(11): 1190-2.
6. Lind AC, Bewtra C, Healy JC, Sims KL. (1995) Prospective peer review in surgical pathology. Am J Clin Pathol 104(5): 560-6.
7. Abt AB, Abt LG, Olt GJ. (1995) The effect of interinstitution anatomic pathology consultation on patient care. Arch Pathol Lab Med 119(6): 514-7.
8. Bruner JM, Inouye L, Fuller GN, Langford LA. (1997) Diagnostic discrepancies and their clinical impact in a neuropathology referral practice. Cancer 79(4): 796-803.
9. Santoso JT, Coleman RL, Voet RL, Bernstein SG, Lifshitz S, Miller D. (1998) Pathology slide review in gynecologic oncology. Obstet Gynecol 91(5 Pt 1): 730-4.
10. Wurzer JC, Al-Saleem TI, Hanlon AL, Freedman GM, Patchefsky A, Hanks GE. (1998) Histopathologic review of prostate biopsies from patients referred to a comprehensive cancer center: correlation of pathologic findings, analysis of cost, and impact on treatment. Cancer 83(4): 753-9.
11. Aldape K, Simmons ML, Davis RL, Miike R, Wiencke J, Barger G, et al. (2000) Discrepancies in diagnoses of neuroepithelial neoplasms: the San Francisco Bay Area Adult Glioma Study. Cancer 88(10): 2342-9.
12. Chafe S, Honore L, Pearcey R, Capstick V. (2000) An analysis of the impact of pathology review in gynecologic cancer. Int J Radiat.Oncol Biol Phys 48(5): 1433-8.
13. Arbiser ZK, Folpe AL, Weiss SW. (2001) Consultative (expert) second opinions in soft tissue pathology. Analysis of problem-prone diagnostic situations. Am J Clin Pathol 116(4): 473-6.
14. Hahm GK, Niemann TH, Lucas JG, Frankel WL. (2001) The value of second opinion in gastrointestinal and liver pathology. Arch Pathol Lab Med 125(6): 736-9.
15. Renshaw AA, Pinnar NE, Jiroutek MR, Young ML. (2002) Quantifying the value of in-house consultation in surgical pathology. Am J Clin Pathol 117(5): 751-4.
16. Khalifa MA, Dodge J, Covens A, Osborne R, Ackerman I. (2003) Slide review in gynecologic oncology ensures completeness of reporting and diagnostic accuracy. Gynecol Oncol 90(2): 425-30.
17. Nguyen PL, Schultz D, Renshaw AA, Vollmer RT, Welch WR, Cote K, et al. (2004) The impact of pathology review on treatment recommendations for patients with adenocarcinoma of the prostate. Urol Oncol 22(4): 295-9.
18. Weydert JA, De Young BR, Cohen MB. (2005) A preliminary diagnosis service provides prospective blinded dual-review of all general surgical pathology cases in an academic practice. Am J Surg Pathol 29(6): 801-5.

19. Croce-Kleinmann S, Marcellin L, Neuville A, Onéa A, Lindner V, Casnedi S, et al. (2008) Internal quality control of histological diagnosis in pathology. A nine-year experience. Ann Pathol 28(1): 9-16.
20. LaCasce AS, Kho ME, Friedberg JW, Niland JC, Abel GA, Rodriguez MA, et al. (2008) Comparison of referring and final pathology for patients with non-Hodgkin's lymphoma in the National Comprehensive Cancer Network. J Clin Oncol 26(31): 5107-12.
21. Brimo F, Schultz L, Epstein JI. (2010) The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. J Urol 184(1): 126-30.
22. Jones K, Jordan RC. (2010) Patterns of second-opinion diagnosis in oral and maxillofacial pathology. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 109(6): 865-9.
23. Lee MC, Levin HS, Jones JS. (2010) The role of pathology review of transurethral bladder tumor resection specimens in the modern era. J Urol 183(3):921-7.
24. Lurkin A, Ducimetière F, Vince DR, Decouvelaere AV, Cellier D, Gilly FN, et al. (2010) Epidemiological evaluation of concordance between initial diagnosis and central pathology review in a comprehensive and prospective series of sarcoma patients in the Rhone-Alpes region. BMC Cancer 10: 150.
25. Wayment RO, Bourne A, Kay P, Tarter TH. (2011) Second opinion pathology in tertiary care of patients with urologic malignancies. Urol Oncol 29(2): 194-8.
26. Zembowicz A, Ahmad A, Lyle SR. (2011) A comprehensive analysis of a web-based dermatopathology second opinion consultation practice. Arch Pathol Lab Med 135(3): 379-83.
27. Matasar MJ, Shi W, Silberstien J, Lin O, Busam KJ, Teruya-Feldstein J, et al. (2012) Expert second-opinion pathology review of lymphoma in the era of the World Health Organization classification. Ann Oncol 23(1): 159-66.
28. Gaudi S, Zarandona JM, Raab SS, English JC 3rd, Jukic DM. (2013) Discrepancies in dermatopathology diagnoses: the role of second review policies and dermatopathology fellowship training. J Am Acad Dermatol 68(1): 119-28.
29. Swapp RE, Aubry MC, Salomão DR, Cheville JC. (2013) Outside case review of surgical pathology for referred patients: the impact on patient care. Arch Pathol Lab Med 137(2): 233-40.
30. Landis JR, Koch GG. (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159-74.
31. Viera AJ, Garrett JM. (2005) Understanding interobserver agreement: the kappa statistic. Fam Med 37(5): 360-3.
32. Engers, R. (2007) Reproducibility and reliability of tumor grading in urological neoplasms. World J Urol 25(6): 595-605.
33. Allsbrook WC Jr, Mangold KA, Johnson MH, Lane RB, Lane CG, Amin MB, et al. (2001) Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. Hum Pathol 32(1): 74-80.
34. Romanoff AM, Cohen A, Schmidt H, Weltz CR, Jaffer SM, Nagi CS, et al. (2014) Breast pathology review: does it make a difference? Ann Surg Oncol 21(11): 3504-8.
35. Troxel DB. (2005) An insurer's perspective on error and loss in pathology. Arch Pathol Lab Med 129(10): 1234-6.

36. Nakhleh RE, Nosé V, Colasacco C, Fatheree LA, Lillemoe TJ, McCrory DC, et al. (2016) Interpretive Diagnostic Error Reduction in Surgical Pathology and Cytology: Guideline From the College of American Pathologists Pathology and Laboratory Quality Center and the Association of Directors of Anatomic and Surgical Pathology. Arch Pathol Lab Med 140(1): 29-40.
37. Federal Policy for the Protection of Human Subjects. Available from: http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/ucm118862.htm.

Improved cytodiagnostics and quality of patient care through double reading of selected cases by an expert cytopathologist

Chantal C.H.J. Kuijpers
Mike Visser
Daisy M.D.S. Sie-Go
Henk de Leeuw
Mathilda J. de Rooij
Paul J. van Diest
Mehdi Jiwa

Published in Virchows Archiv 2015; 466(6): 617-624

Abstract

Introduction

Double reading may be a valuable tool for improving the quality of patient care by restoring diagnostic errors before final sign out, but standard double reading would significantly increase costs of pathology. We aimed to assess the added value of routine double reading of defined categories of clinical cytology specimens by specialized cytopathologists.

Methods

Specialized cytopathologists routinely re-diagnosed, blinded, defined categories of clinical cytology specimens that had been signed out by routine pathologists from January 2012 up to December 2013. Major and minor discordance rates between initial and expert diagnoses were determined, and both diagnoses were validated by comparison with same-site histological follow up.

Results

Initial and expert diagnoses were concordant in 131/218 specimens (60.1%). Major and minor discordance were present in 28 (12.8%) and 59 (27.1%) specimens, respectively. Pleural fluid, thyroid and urine specimens showed the highest major discordance rates (19.4%, 19.2% and 16.7%, respectively). Histological follow up (where possible) supported the expert diagnosis in 95.5% of specimens.

Conclusion

Our implemented double reading strategy of defined categories of cytology specimens showed major discordance in 12.8% of specimens. The expert diagnosis was supported in 95.5% of discordant cases where histological follow up was available. This indicates that this double reading strategy is worthwhile and contributes to better cytodiagnostics and quality of patient care, especially for suspicious pleural fluid, thyroid and urine specimens. Our results emphasize that cytopathology is a subspecialization of pathology and requires specialized cytopathologists.

Introduction

There is a growing awareness that pathology diagnosis is not infallible and that diagnostic errors may lead to under- or overtreatment, and thereby compromise patient safety. Double reading is a potentially valuable tool for reducing diagnostic errors and thereby improving the quality of patient care. It may reveal inaccurate diagnoses that otherwise might have led to improper or unnecessary patient management or treatment. In response to the Institute of Medicine report 'To err is human; building a safer health system' from 1999 [1], the American Society of Clinical Pathologists (ASCP) recognized second opinion as a key aspect in the assurance of patient safety for histological and cytological diagnoses [2]. They recommended to consider second opinion in several situations, including highly critical or significant cases, problem-prone cases and cases suggested for review by clinicians [2].

Many studies focused on second opinion in diagnostic surgical pathology, and reported major diagnostic disagreement rates of 2% to 28%, mainly depending on the organ system studied [3-24]. A smaller number of studies focused on second opinion in cytopathology, of which the majority reported disagreement rates of specific organs or organ systems, predominantly the thyroid [25-31]. Few studies, however, assessed the impact of double reading on patient care for the whole subset of cytological specimens. These studies reported major disagreement rates ranging from 7.4% to 9.3% [32-34], and second opinion diagnoses were better supported by histological follow up than the initial diagnoses [32,33].

Therefore, we implemented intradepartmental double reading by expert cytopathologists on January 1st 2012. Since routine double reading of all specimens would significantly increase costs, we predefined selected categories of cytology cases where yield of double reading was expected to be highest. In this study we assessed the added value of this expert double reading strategy. To this end, we retrospectively determined the rates of concordance, and major and minor discordance between initial and second opinion diagnoses of all cytology cases reviewed by the expert cytopathologists. Furthermore, we validated both diagnoses by comparison with same-site histological follow up.

Methods

Routine intradepartmental second review

Figure 1 demonstrates the routine cytology diagnostics process at Symbiant's three pathology laboratories (Alkmaar Medical Centre, Westfriesgasthuis Hoorn and Zaandam Medical Centre). All cytological specimens were routinely prescreened for abnormalities by 1 or 2 cytotechnicians. Subsequently, the prescreened specimens are examined by either a general pathologist (in the Alkmaar and Hoorn laboratories) and both general pathologists and expert cytopathologists in the Zaandam laboratory.

Starting from January 1st 2012, a cytopathology expert team in the Zaandam laboratory began reviewing defined categories of clinical cytology specimens from the Alkmaar and Hoorn pathology laboratories, resulting in a consensus diagnosis. The team consisted of two expert cytopathologists (DSG and MV). Second review was performed blinded to the initial diagnoses. The following types of specimens were routinely sent for intradepartmental second review: difficult or suspicious cases, and cases with a discrepancy between the general pathologists' diagnoses, with the clinical presentation, or with immunohistochemical stains.

The cytopathologists were either consulted before case sign out, when the initial pathologist was unable to offer a preliminary diagnosis, or asked for a second opinion after preliminary sign-out. For the purpose of this study, the cases where the expert cytopathologists were consulted pre-sign out were excluded from analysis. In the remaining cases, the initial diagnoses and the expert diagnoses were recorded and compared.

Assessment of concordance between initial and expert diagnoses

We thereby retrospectively assessed all clinical cytopathology cases of 2012 and 2013 that had been reviewed by the expert cytopathologists, and determined concordance between initial and expert diagnoses. We applied the same definitions for minor and major discordances as described by Lueck *et al* and Bomeisl *et al* [32,33]. Minor discordance was defined as a 1-step deviation on the scale of "non-diagnostic, benign, atypical, suspicious, and malignant" without an effect on treatment or prognosis. Major discordance was defined as either a deviation of ≥ 2 steps on this scale or a discordance with effect on patient management or prognosis.

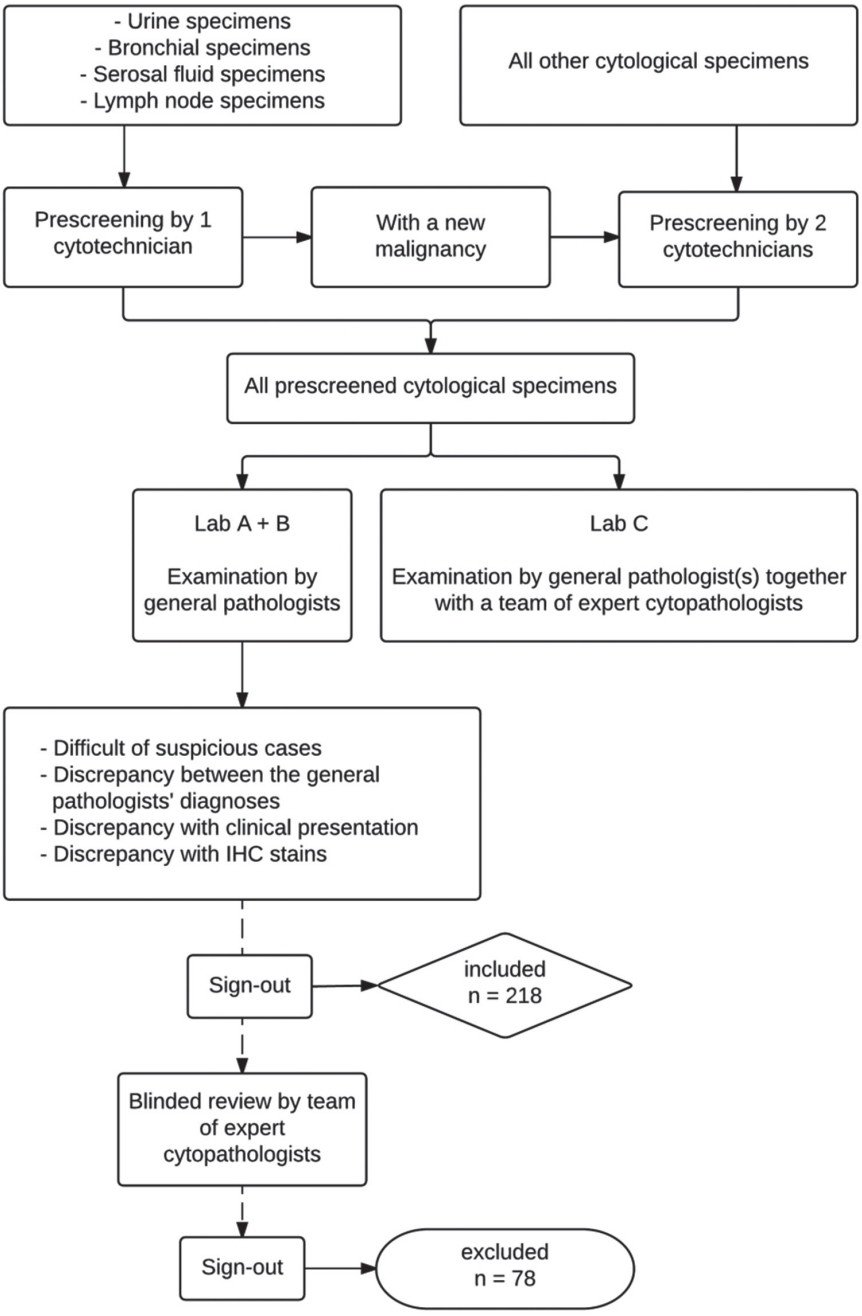


Figure 1. The routine cytology diagnostics process at Symbiant's three pathology laboratories. Lab A = Alkmaar Medical Centre, Lab B = Westfriesgasthuis Hoorn, Lab C = Zaandam Medical Centre.

Validation of diagnoses by comparison with histological follow up

We validated initial and expert diagnoses by comparison with same-site histological follow up diagnoses. The process of follow up identification is explained below. The diagnosis closest to the follow up diagnosis was deemed correct. Non-diagnostic specimens, which had insufficient diagnostic material or were non-representative, were not validated by histological follow up.

Identification of cytohistologically discordant cases

As follow up identification is a very time-consuming activity, a cytotechnician at Symbiant (HdL) developed the follow up tool Follow Up application SYMbiant (FUSYM), which provides histological follow up for cytology specimens by automating several steps in the process. FUSYM has been developed in Microsoft Visual Studio 2010 professional and was written in Microsoft Visual Basic.

The Netherlands employs a unique system whereby all reports from the Dutch pathology laboratories are stored in a central database (PALGA) via a local server. All histological examinations subsequent to the cytological examination were routinely extracted from PALGA and loaded into FUSYM. The following was coded while loading data: tissue type, organ, sampling region, sampling method and side. Furthermore, diagnoses were classified at three levels, known as diagnostic group (unknown, non-diagnostic, benign, atypia or malignant), main diagnosis (benign was subdivided into no abnormalities, benign lesion and benign neoplasm, and malignant was subdivided into suspicious and malignant neoplasm) and specific diagnosis. Subsequently, the actual follow up examination from the same-site as the cytology specimen (i.e. same organ and sampling region) was determined manually for every cytology specimen.

Finally, cytology and histology follow up diagnoses were compared at the level of diagnostic group to determine the cytohistologic concordance rate and the number of 'false-negative' and 'false-positive' cytology diagnoses. Suspicious as well as malignant cytology diagnoses with a malignant histological follow up were deemed concordant. Non-diagnostic cytology or histology specimens or specimens with a diagnosis of atypia were excluded from analysis.

Retrospective double reading of a sample of cytohistologically discordant cases initially not undergoing double reading

To assess the quality of cytodiagnostics of the cases from 2012 and 2013 that

Table 1. Summary of tissue types and acquisition methods of 218 clinical cytology specimens undergoing double reading by expert cytopathologists.

Tissue type	Number	Percentage
Thyroid FNA	52	23.9%
Lymph node FNA	40	18.3%
Pleural fluid	31	14.2%
Salivary gland FNA	22	10.1%
Bile duct brush	13	6.0%
Urine	12	5.5%
Bronchial FNA/brush/lavage	11	5.0%
Breast FNA/nipple discharge	11	5.0%
Ascitic fluid	8	3.7%
Adrenal gland FNA	6	2.8%
Liver FNA/brush	3	1.4%
Pancreas FNA	2	0.9%
Pericardial fluid	2	0.9%
Cerebrospinal fluid	1	0.5%
Peritoneal FNA	1	0.5%
Esophagal FNA	1	0.5%
Scrotal FNA	1	0.5%
Retro-auricular FNA	1	0.5%

FNA=fine needle aspiration

Table 2. Types of major discordances (n=28) and minor discordances (n = 59) for clinical cytology specimens undergoing double reading by expert cytopathologists.

Type of major discordance	Number	Percentage	Type of minor discordance	Number	Percentage
Underestimated	14	50.0%	**Underestimated**	31	52.6%
Benign ⟶ suspicious	4	14.3%	Benign ⟶ atypia	1	1.7%
Benign ⟶ malignant	3	10.7%	Atypia ⟶ suspicious	7	11.9%
Atypia ⟶ malignant	7	25.0%	Suspicious ⟶ malignant	23	39.0%
Overestimated	13	46.4%	**Overestimated**	24	40.7%
Malignant ⟶ atypia	1	3.6%	Malignant ⟶ suspicious	2	3.4%
Malignant ⟶ benign	2	7.1%	Suspicious ⟶ atypia	7	11.9%
Suspicious ⟶ benign	10	35.7%	Atypia ⟶ benign	15	25.4%
Other	1	3.6%	**Other**	4	6.8%

did no undergo routine double reading, concordance with histological follow up was determined. Of the cytohistologically discordant cases, we randomly selected a sample of 100 specimens from the Alkmaar pathology laboratory to be retrospectively reviewed by the expert cytopathologists, blinded to the original diagnosis, and recorded whether the expert diagnosis was concordant or discordant with the initial diagnosis. A discordant diagnosis was subdivided into 1. a change from benign to malignant, or vice versa, resulting in expert cytohistologic concordance, 2. a change from benign or malignant into atypia, or 3. a change into non-diagnostic.

Statistics

Statistical analysis was performed using SPSS 20. Initial diagnoses and expert diagnoses were compared using the unweighted Cohen's Kappa (K) coefficient. A Kappa-value of 0.00–0.20 indicates a slight agreement, 0.21–0.40 a fair agreement, 0.41–0.60 a moderate agreement, 0.61–0.80 a substantial agreement, and 0.81–1 a perfect agreement. Furthermore, the percentages of concordance and discordance with their 95% confidence intervals (CI) were calculated. P-values < 0.05 were considered statistically significant.

Results

Specimens routinely undergoing double reading

During the study period, 296 clinical cytology specimens underwent routine double reading by the expert cytopathologists. We excluded 78 cases where the initial diagnoses were not recorded, because pathologists consulted the expert team before case sign-out, leaving 218 specimens. From 12 patients multiple cytology specimens were included as separate cases (11 patients with 2 specimens and 1 patient with 3 specimens). Table 1 summarizes the tissue types and sampling methods of the 218 specimens.

Both diagnoses were concordant in 131 specimens (60.1%; 95% CI: 0.535-0.666, Kappa 0.489, $p < 0.0001$). Major discordance between the initial and the expert diagnosis was seen in 28 specimens (12.8%; 95% CI: 0.084-0.173) and minor discordance in 59 specimens (27.1%; 95% CI: 0.211-0.330).

Table 2 summarizes the types of major and minor discordances. Of all discordant specimens, the initial diagnosis was underestimated 45 times (51.7%) and overestimated 37 times (42.5%). Twice, a benign diagnosis was changed into non-diagnostic, and 1 specimen was changed from non-diagnostic

Table 3. Frequencies of discordances subdivided by tissue type. Tissue types with ≥ 10 cytology specimens reviewed by the expert cytopathologists were compared.

Tissue type	# specimens	Total discordant expert diagnoses	Major discordance	Minor discordance
Pleural fluid	31	18 (58.1%)	6 (19.4%)	12 (38.7%)
Urine	12	6 (50.0%)	2 (16.7%)	4 (33.3%)
Bile duct brush	13	6 (46.2%)	-	6 (46.2%)
Bronchial FNA/brush/ lavage	11	5 (45.5%)	1 (9.1%)	4 (36.4%)
Thyroid FNA	52	23 (44.2%)	10 (19.2%)	13 (25.0%)
Lymph node FNA	40	11 (27.5%)	6 (15.0%)	5 (12.5%)
Salivary gland FNA	22	6 (27.3%)	-	6 (27.3%)
Breast FNA/nipple discharge	11	1 (9.1%)	-	1 (9.1%)
Total	192	76 (39.6%)	25 (13.0%)	51 (26.6%)

FNA= fine needle aspiration.

into malignant. Furthermore, for 1 specimen an unspecific benign diagnosis (no malignancy) was specified into a Warthin tumor, and in another case a diagnosis of metastatic squamous cell carcinoma was changed into metastatic adenocarcinoma.

Table 3 shows the frequencies of discordant second opinion diagnoses subdivided by the eight tissue types with ≥ 10 specimens reviewed by the expert cytopathologists. Pleural fluid-, urine- and bile duct brush specimens showed the highest overall discordance rates (58.1%, 50.0% and 46.2%, respectively). Major discordances were most commonly observed in pleural fluid-, thyroid- and urine specimens with 19.4%, 19.2% and 16.7%, respectively. Minor discordances were most commonly observed in bile duct brush-, pleural fluid- and bronchial specimens with 46.2%, 38.7% and 36.4%, respectively. Breast cytology specimens showed the lowest discordance rate, with 1 minor discordant expert diagnosis (9.1%). The total major and minor discordance percentages of this subset of specimens (tissue types with ≥ 10 specimens) were similar to that of the whole study selection (all tissue types).

Validation by comparison with histological follow up

Same-site histological follow up was available for 25 of the 87 discordant specimens, but was non-diagnostic for 3 specimens. Hence, we validated the initial and expert diagnoses of 22 cytology specimens by comparison with

histological follow up. The expert diagnosis was supported by the histology diagnosis in 21/22 specimens (95.5%; 95% CI: 0.860-1.049). The case that was not supported by histology revealed a malignant mesothelioma, which was not diagnosed in the pleural fluid cytology by both the initial pathologist as well as the expert cytopathologist.

Retrospective second review of cytohistologically discordant cases

In order to get an impression of the quality of cytology diagnostics in those specimens that did not undergo double reading in routine diagnostics, we determined the rates of concordance and discordance with histological follow up in these specimens. Same-site histological follow up was available for 1,613 cases, of which we excluded 338 cases, because either cytology, histology or both were non-diagnostic or had a diagnosis of atypia. Furthermore, we excluded 24 cases with a time period between cytological and histological examination longer than 6 months and 17 cases with only a few malignant cells on histology, leaving 1,234 cases. Cytohistological concordance was found for 943 cytology specimens (76.4%; 95% CI: 0.740-0.788)) and 291 cytology specimens (23.6%) had a discordant histological follow up.

For the random sample of 100 cytohistologically discordant cases from the Alkmaar pathology laboratory, the expert diagnosis was consistent with the initial diagnosis in 57% (95% CI: 0.471-0.669). The cytopathologists changed the diagnosis in 43% of cases: in 17 cases the diagnosis was changed from benign to malignant (10 cases) or vice versa (7 cases), resulting in cytohistological concordance, a benign or malignant diagnosis was changed into atypia in 8 cases, and a diagnosis was changed into non-diagnostic 18 times.

Discussion

This study assessed the added value of our implemented intradepartmental double reading strategy of defined categories of clinical cytology specimens by a team of expert cytopathologists. We demonstrated a 60.1% concordance rate, a 12.8% major discordance rate and a 27.1% minor discordance rate between initial and expert diagnoses. The highest major discordance rates were observed in pleural fluid-, thyroid- and urine specimens. Validation by comparison with same-site histological follow up confirmed that expert diagnoses were correct in 95.5% (95% CI: 0.860-1.049). These findings emphasize the importance of double reading of selected specimens by expert cytopathologists.

Previous studies on cytopathology double reading demonstrated somewhat lower major discordance rates (7.4% to 9.3%) [32-34], probably due to differences in specimen selection. At our institution, defined categories of clinical cytology specimens were reviewed, whereas others described second review of all referred cytopathology material before definitive treatment. Furthermore, we specifically assessed the added value of double reading by expert cytopathologists. In these studies as well, high major disagreement rates of thyroid FNA specimens were observed (16.2% to 24.3%), and in the study of Lueck *et al* [32], major discrepancies in urine specimens were the third most common (16.2%).

The high discordance rates in urine and pleural fluid specimens might be partly explained by the lack of standard terminology and the use of inadequate terms, especially for atypical lesions [35]. Implementation of the Paris System for Urinary Cytopathology, which is currently being developed, might improve urine cytology diagnostics [36]. This explanation does, however, not hold true for thyroid cytology specimens, because of well-defined terminology in the Bethesda System for Reporting Thyroid Cytopathology (BSRTC) [37]. We therefore suppose that most discrepancies were a result of inadequate interpretation instead of inadequate terms used. The majority of thyroid cytology discrepancies were caused by initial overestimation of benign and atypical specimens.

Initial underestimation occurred in slightly more discordant cytology specimens than overestimation did (51.7% and 42.5%, respectively). This difference mainly represented minor discordant specimens, of which malignancies being underestimated as suspicious were most commonly observed, indicating reluctance among general pathologists to label cases as malignant. Among the major discordant specimens, the proportions of underestimated and overestimated diagnoses by general pathologists were evenly distributed.

A limitation of this study is the relatively small availability of same-site histological follow up (in 25/87 discordant specimens) to validate expert diagnoses, which may lead to partial verification bias. Reasons for the absence of same-site histological follow up were assessed. They included the presence of a benign cytology or benign follow up cytology diagnosis (n=18), the presence of histological follow up obtained from another related site (n=10) or radiological follow up (n=3). For 17 patients histological follow up would

have been superfluous, because they already suffered from incurable metastatic malignancies. Furthermore, 12 patients were treated in an academic hospital or another local hospital, of which patient charts were unavailable to us, and 1 patient died very shortly after cytological examination. Finally, for 1 patient with an atypical thyroid cytology specimen, an intended hemithyroidectomy was probably cancelled for an, to us, unknown reason.

Our double reading strategy reveals major discordant diagnoses in a substantial number of cytology cases. Although standard double reading of all cytology specimens would be ideal in terms of patient safety, it would significantly increase workload and costs of pathology. Alternatively, all cytology could be signed out by expert cytopathologists, but this is in general pathology practice difficult to realize. In line with the present results, Raab *et al* [38] demonstrated that focused review of diagnostically challenging areas of surgical pathology was more time- and cost-effective than 5% random review and detected a significantly higher frequency of discrepancies. In order to get an impression of the quality of cytology diagnostics in those specimens that did not undergo double reading in routine diagnostics, we determined the concordance rate with histological follow up, which appeared to be 76.4%. Retrospective double reading of a random sample of 100 cytohistologically discordant specimens changed the diagnosis in 43 cases, with urine- and lymph node specimens most commonly adapted. In these cases, the sign-out pathologist probably had been sufficiently confident of the diagnosis and therefore had not consulted the expert cytopathologists. This argues for investigating which further specimen types are problematic as well and would probably also benefit from initial double reading, in order to refine the double reading strategy.

Patient safety is of utmost importance, and, in our opinion, cytopathology is a subspecialization of pathology that requires specialized cytopathologists, because the discordance rates are unacceptably high. The Dutch thyroid carcinoma guideline [39] states that "thyroid FNAs should be assessed by a pathologist with interest and experience in thyroid cytology and histology, who can recommend management or treatment based on the cytology results. If an experienced pathologist is not available locally, the sample should be sent to a pathologist that does have expertise in this field." Also the Dutch bladder carcinoma guideline [40] states that "reliability of urine cytology evaluation is dependent on the expertise of the (cyto)pathologist." The Board of Pathology of the European Union of Medical Specialists (UEMS) published requirements for recognition of postgraduate training in pathology, and stated that

cytopathology is an integral part of pathology, and well-trained pathologists must be able to cover basic cytological diagnosis [41]. Pathologists can obtain post-graduate 'advanced level of competence' certificates in cytopathology [41]. We agree with Anshu *et al* [42] who stated that "European and international guidelines for training and accreditation in cytopathology should be developed with some urgency". Guidelines should include an annual minimum number of specimens that a (cyto)pathologist must view and recommendations for further education and examination.

Conclusion: Our implemented double reading strategy of defined categories of cytology specimens showed major discordance in 12.8% of specimens. The expert review was supported in 95.5% of discordant cases where histological follow up was available. This indicates that this double reading strategy is worthwhile and contributes to better cytodiagnostics and quality of patient care, especially for suspicious pleural fluid, thyroid and urine specimens. Although it is currently not reimbursed and formal cost-effectiveness studies are lacking, we believe that selected second review may prevent overtreatment of a subgroup of patients in a cost-effective way, and also in the light of the upcoming claim culture in Europe, should therefore be considered for regular reimbursement. Our results emphasize that cytopathology is a subspecialization of pathology and requires specialized cytopathologists.

References

1. Kohn LT, Corrigna JM, Donaldson MS. (1999) To Err is Human; Building a Safer Health System, Washington DC: Institute of Medicine, National Academy Press
2. Tomaszewski JE, Bear HD, Connally JA, Epstein JI, Feldman M, Foucar K, et al. (2000) Consensus conference on second opinions in diagnostic anatomic pathology. Who, What, and When. Am J Clin Pathol 114(3): 329-35.
3. Swapp RE, Aubry MC, Salomao DR, Cheville JC. (2013) Outside case review of surgical pathology for referred patients: the impact on patient care. Arch Pathol Lab Med 137(2): 233-40.
4. Gaudi S, Zarandona JM, Raab SS, English JC 3rd, Jukic DM. (2013) Discrepancies in dermatopathology diagnoses: the role of second review policies and dermatopathology fellowship training. J Am Acad Dermatol 68(1): 119-28.
5. Matasar MJ, Shi W, Silberstien J, Lin O, Busam KJ, Teruya-Feldstein J, et al. (2012) Expert second-opinion pathology review of lymphoma in the era of the World Health Organization classification. Ann Oncol 23(1): 159-66.
6. LaCasce AS, Kho ME, Friedberg JW, Niland JC, Abel GA, Rodriguez MA, et al. (2008) Comparison of referring and final pathology for patients with non-Hodgkin's lymphoma in the National Comprehensive Cancer Network. J Clin Oncol 26(31): 5107-12.
7. Brimo F, Schultz L, Epstein JI. (2010) The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. J Urol 184(1): 126-30.
8. Nguyen PL, Schultz D, Renshaw AA, Vollmer RT, Welch WR, Cote K, et al. (2004) The impact of pathology review on treatment recommendations for patients with adenocarcinoma of the prostate. Urol Oncol 22(4): 295-9.
9. Wurzer JC, Al Saleem TI, Hanlon AL, Freedman GM, Patchefsky A, Hanks GE. (1998) Histopathologic review of prostate biopsies from patients referred to a comprehensive cancer center: correlation of pathologic findings, analysis of cost, and impact on treatment. Cancer 83(4): 753-9.
10. Lee MC, Levin HS, Jones JS. (2010) The role of pathology review of transurethral bladder tumor resection specimens in the modern era. J Urol 183(3): 921-7.
11. Jones K, Jordan RC. (2010) Patterns of second-opinion diagnosis in oral and maxillofacial pathology. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 109(6): 865-9.
12. Zembowicz A, Ahmad A, Lyle SR. (2011) A comprehensive analysis of a web-based dermatopathology second opinion consultation practice. Arch Pathol Lab Med 135(3): 379-83.
13. Lurkin A, Ducimetiere F, Vince DR, Decouvelaere AV, Cellier D, Gilly FN, et al. (2010) Epidemiological evaluation of concordance between initial diagnosis and central pathology review in a comprehensive and prospective series of sarcoma patients in the Rhone-Alpes region. BMC Cancer 10: 150.
14. Arbiser ZK, Folpe AL, Weiss SW. (2001) Consultative (expert) second opinions in soft tissue pathology. Analysis of problem-prone diagnostic situations. Am J Clin Pathol 116(4): 473-6.
15. Hahm GK, Niemann TH, Lucas JG, Frankel WL. (2001) The value of second opinion in gastrointestinal and liver pathology. Arch Pathol Lab Med 125(6): 736-9.
16. Chafe S, Honore L, Pearcey R, Capstick V. (2000) An analysis of the impact of pathology review in gynecologic cancer. Int J Radiat.Oncol Biol Phys 48(5): 1433-8.

17. Khalifa MA, Dodge J, Covens A, Osborne R, Ackerman I. (2003) Slide review in gynecologic oncology ensures completeness of reporting and diagnostic accuracy. Gynecol Oncol 90(2): 425-30.
18. Bruner JM, Inouye L, Fuller GN, Langford LA. (1997) Diagnostic discrepancies and their clinical impact in a neuropathology referral practice. Cancer 79(4): 796-803.
19. Wayment RO, Bourne A, Kay P, Tarter TH. (2011) Second opinion pathology in tertiary care of patients with urologic malignancies. Urol Oncol 29(2): 194-8.
20. Santoso JT, Coleman RL, Voet RL, Bernstein SG, Lifshitz S, Miller D. (1998) Pathology slide review in gynecologic oncology. Obstet Gynecol 91(5 Pt 1): 730-4.
21. Scott CB, Nelson JS, Farnan NC, Curran WJ Jr, Murray KJ, Fischbach AJ, et al. (1995) Central pathology review in clinical trials for patients with malignant glioma. A Report of Radiation Therapy Oncology Group 83-02. Cancer 76(2): 307-13.
22. Aldape K, Simmons ML, Davis RL, Miike R, Wiencke J, Barger G, et al. (2000) Discrepancies in diagnoses of neuroepithelial neoplasms: the San Francisco Bay Area Adult Glioma Study. Cancer 88(10): 2342-9.
23. Romanoff AM, Cohen A, Schmidt H, Weltz CR, Jaffer SM, Nagi CS, et al. (2014) Breast pathology review: does it make a difference? Ann Surg Oncol 21(11): 3504-8.
24. Marco V, Muntal T, García-Hernandez F, Cortes J, Gonzalez B, Rubio IT. (2014) Changes in breast cancer reports after pathology second opinion. Breast J 20(3): 295-301
25. Baloch ZW, Hendreen S, Gupta PK, LiVolsi VA, Mandel SJ, Weber R, et al. (2001) Interinstitutional review of thyroid fine-needle aspirations: impact on clinical management of thyroid nodules. Diagn Cytopathol 25(4): 231-4.
26. Tan YY, Kebebew E, Reiff E, Caron NR, Ogilvie JB, Duh QY, et al. (2007) Does routine consultation of thyroid fine-needle aspiration cytology change surgical management? J Am Coll Surg 205(1): 8-12.
27. Davidov T, Trooskin SZ, Shanker BA, Yip D, Eng O, Crystal J, et al. (2010) Routine second-opinion cytopathology review of thyroid fine needle aspiration biopsies reduces diagnostic thyroidectomy. Surgery 148(6): 1294-9.
28. Park JH, Kim HK, Kang SW, Jeong JJ, Nam KH, Chung WY, et al. (2012) Second opinion in thyroid fine-needle aspiration biopsy by the Bethesda system. Endocr J 59(3): 205-12.
29. Bajaj J, Morgenstern N, Sugrue C, Wasserman J, Wasserman P. (2012) Clinical impact of second opinion in thyroid fine needle aspiration cytology (FNAC): A study of 922 interinstitutional consultations. Diagn Cytopathol 40(5): 422-9.
30. Olson MT, Boonyaarunnate T, Aragon Han P, Umbricht CB, Ali SZ, Zeiger MA. (2013) A tertiary center's experience with second review of 3885 thyroid cytopathology specimens. J Clin Endocrinol Metab 98(4): 1450-7.
31. Kronz JD, Westra WH. (2005) The role of second opinion pathology in the management of lesions of the head and neck. Curr Opin Otolaryngol.Head Neck Surg 13(2): 81-4.
32. Lueck N, Jensen C, Cohen MB, Weydert JA. (2009) Mandatory second opinion in cytopathology. Cancer 117(2): 82-91.
33. Bomeisl PE Jr, Alam S, Wakely PE Jr. (2009) Interinstitutional consultation in fine-needle aspiration cytopathology: a study of 742 cases. Cancer 117(4): 237-46.
34. Layfield LJ, Jones C, Rowe L, Gopez EV. (2002) Institutional review of outside cytology materials: a retrospective analysis of two institutions' experiences. Diagn Cytopathol 26(1): 45-8.

35. Owens CL, Vandenbussche CJ, Burroughs FH, Rosenthal DL. (2013) A review of reporting systems and terminology for urine cytology. Cancer Cytopathol 121(1): 9-14.
36. The Paris Reporting System for Urinary Cytopathology. Available from: https://paris.soc.wisc.edu/index.htm.
37. Cibas ES, Ali SZ. (2009) The Bethesda System for Reporting Thyroid Cytopathology. Thyroid 19(11): 1159-65.
38. Raab SS, Grzybicki DM, Mahood LK, Parwani AV, Kuan SF, Rao UN. (2008) Effectiveness of random and focused review in detecting surgical pathology error. Am J Clin Pathol 130(6): 905-12.
39. Dutch guideline for thyroid cancer. Available from: http://www.oncoline.nl/index.php?pagina=/richtlijn/item/pagina.php&richtlijn_id=554
40. Dutch guideline for bladder carcinoma. Available from: http://www.oncoline.nl/bladdercarcinoma.
41. Tötsch M, Cuvelier C, Vass L, Fassina A; UEMS Section/Board of Pathology meeting in Paris 2012. (2012) The UEMS section/board of pathology, chapter 6: requirement for recognition of postgraduate training in pathology: a presentation of the paris document. Cytopathology 23(5): 295-9.
42. Anshu, Herbert A, Cochand-Priollet B, Cross P, Desai M, Dina R, et al. (2010) Survey of medical training in cytopathology carried out by the journal Cytopathology. Cytopathology 21(3): 147-56.

Added value of HER-2 amplification testing by Multiplex Ligation-dependent Probe Amplification in invasive breast cancer

Chantal C.H.J. Kuijpers
Cathy B. Moelans
Henk-Jan van Slooten
Anja Horstman
John W.J. Hinrichs
Shaimaa Al-Janabi
Paul J. van Diest
Mehdi Jiwa

Published in PLoS One 2013;8(12):e82018

Abstract

Introduction

HER-2 is a prognostic and predictive marker, but as yet no technique is perfectly able to identify patients likely to benefit from HER-2 targeted therapies. We aimed to prospectively assess the added value of first-line co-testing by IHC, and multiplex ligation-dependent probe amplification (MLPA) and chromogenic *in situ* hybridization (CISH).

Methods

As local validation, HER-2 MLPA and CISH were compared in 99 breast cancers. Next, we reviewed 937 invasive breast cancers, from 4 Dutch pathology laboratories, that were prospectively assessed for HER-2 by IHC and MLPA (and CISH in selected cases).

Results

The validation study demonstrated 100% concordance between CISH and MLPA, if both methods were assessable and conclusive (81.8% of cases). Significant variation regarding percentages IHC 0/1+ and 2+ cases was observed between the laboratories ($p<0.0001$). Overall concordance between IHC and MLPA/CISH was 98.1% (575/586) (Kappa=0.94). Of the IHC 3+ cases, 6.7% failed to reveal gene amplification, whereas 0.8% of the IHC 0/1+ cases demonstrated gene amplification. Results remained discordant after retrospective review in 3/11 discordant cases. In the remaining 8 cases the original IHC score was incorrect or adapted after repeated IHC staining.

Conclusions

MLPA is a low-cost and quantitative high-throughput technique with near perfect concordance with CISH. The use of MLPA in routinely co-testing all breast cancers may reduce HER-2 testing variation between laboratories, may serve as quality control for IHC, will reveal IHC 0/1+ patients with gene amplification, likely responsive to trastuzumab, and identify IHC 3+ cases without gene amplification that may respond less well.

Introduction

In breast cancer patients, the HER-2 oncogene is both a prognostic and a predictive marker, but as yet no technique is perfectly able to identify patients likely to benefit from HER-2 targeted therapies. The HER-2 oncogene is amplified and/or overexpressed in approximately 10-15% of human breast cancers [1]. Overexpression of the HER-2 oncogene is associated with poor prognosis and resistance to chemotherapy and hormonal therapy [2]. More importantly, HER-2 status identifies patients likely to benefit from treatment with the recombinant humanized monoclonal antibody trastuzumab and the small molecule tyrosine kinase inhibitor lapatinib [3,4]. As both these therapies are expensive, and trastuzumab is associated with serious, particularly cardiotoxic, side-effects, the ASCO/CAP guidelines [5] stipulate that trastuzumab therapy is only applicable for patients who strongly overexpress the HER-2 protein (3+) and those who present with equivocal HER-2 protein levels (2+) with confirmed gene amplification.

Accurate HER-2 assessment is required, and an accurate, robust and reproducible assay is essential. The most commonly used method to assess HER-2 status is immunohistochemistry (IHC), probably because it is widely available and relatively inexpensive. However, despite of a standardized testing protocol, IHC has proven to be sensitive to pre-analytical variation and interobserver and intraobserver variability [6-8].

Assessment of HER-2 gene amplification status, usually limited to the IHC equivocal cases, is most commonly performed by fluorescence (FISH) and chromogenic *in situ* hybridization (CISH). These two techniques, demonstrating excellent concordance [9], are less sensitive to pre-analytical variation than IHC, but they are labor-intensive, expensive and somewhat difficult to interpret.

A cost-effective, easy-to-perform and quantitative high-throughput technique for routinely assessing HER-2 gene amplification status may be the PCR-based multiplex ligation-dependent probe amplification (MLPA) technique. Moerland *et al* [10] first described MLPA for the use of HER-2 gene amplification detection in 2006. In their study and in other studies, MLPA has proven to be a reliable, less expensive and high-throughput alternative, highly concordant with ISH [11-14].

HER-2 protein expression and gene amplification are not in complete accordance. Previous studies have demonstrated IHC 0/1+ breast cancers with HER-2 gene amplification as well as IHC 3+ cases without gene amplification [7,10-12,15-19]. Therefore, in April 2011 we at Symbiant started co-testing for HER-2 with IHC and MLPA in every invasive breast cancer case, instead of only the equivocal cases. In the University Medical Centre (UMC) Utrecht co-testing of every invasive breast cancer case was first applied in 2004.

The aim of our study was to prospectively assess the added value of first-line co-testing for HER-2 using a combination of IHC and MLPA in routine pathology practice.

Methods

Ethics statement

Since we used archival pathology material which does not interfere with patient care and does not involve the physical involvement of the patient, no ethical approval is required according to Dutch legislation [the Medical Research Involving Human Subjects Act (Wet medisch-wetenschappelijk onderzoek met mensen, WMO [20])]. Use of anonymous or coded left over material for scientific purposes is part of the standard treatment contract with patients, and therefore informed consent procedure was not required according to our institutional medical ethical review board. This has also been described by van Diest *et al* [21]. We assume that our study is subjected to exemption from the Federal regulation as has been suggested below: Exemption 4 includes research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects [22]. Also, this is based on the Dutch guidelines for research [23].

MLPA validation study

We conducted a validation study prior to implementing MLPA for assessing HER-2 amplification status in invasive breast cancer into the diagnostics of the 3 Symbiant laboratories (Alkmaar Medical Centre, Zaandam Medical Centre and Westfriesgasthuis Hoorn, The Netherlands) in April 2011. We triple-tested 99 invasive breast cancer cases by IHC, MLPA and CISH. The aim of this validation study was to locally determine the concordance rates of these techniques.

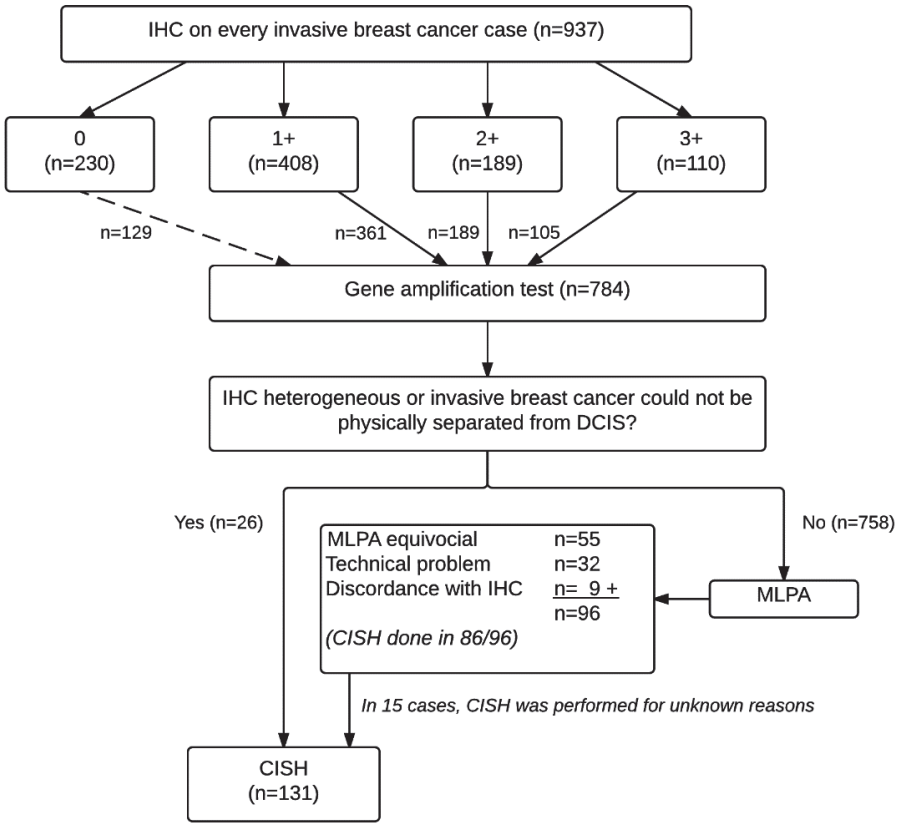


Figure 1. HER-2 co-testing protocol. IHC was performed on every case of invasive breast cancer. 129/230 cases with IHC score 0, 361/409 IHC 1+ cases, all 189 IHC 2+ cases and 105/110 IHC 3+ cases were tested by either MLPA, CISH or both. MLPA was performed in 758 cases and reflex CISH was performed in 86/96 cases with an equivocal MLPA result, a technical problem or a discordant result from IHC. The remaining 10/96 cases did not undergo reflex CISH either due to an insufficient amount of tumor tissue or an unknown reason. Furthermore, in 16 cases, CISH was performed beyond the protocol. Finally, CISH was performed instead of MLPA in 26 cases (4 were immunohistochemically heterogeneous, in 4 cases invasive tumor could not be physically separated from DCIS, in 12 cases tumor cell percentage was low, in 3 cases no MLPA was requested accidentally and in 3 cases the reason why no MLPA was performed could not be elucidated).

Patient cohorts

We reviewed the pathology reports of consecutive cohorts of 920 invasive breast cancer patients from 4 laboratories in The Netherlands (Figure 1). For 720 patients, their HER-2 status was determined in one of the 3 laboratories of Symbiant between April 2011 and February 2012. For the other 200 patients,

their HER-2 status was determined in the pathology laboratory of the UMC Utrecht between January 2010 and December 2011. Eighteen patients had multiple tumors that were included as separate cases. We excluded 2 cases which had only a very small micro-invasive component. In total 937 tumors were thereby analyzed in this study. Of these cases, 78.7% were resections, 19.9% were biopsies, and 3.6% were metastases.

Immunohistochemistry

IHC staining was performed fully automated using the Bond-III or the Bond-Max™ immunostainers (Leica, Wetzlar, Germany), according to the manufacturers' instructions on 4 μm thick tissue paraffin sections. In addition, separate paraffin sections with 4 human breast cancer cell lines (0, 1+, 2+ and 3+ intensity, provided by Leica), were taken along as controls to validate staining runs. Moreover, a small control tissue array containing a 0, 1+ and 3+ breast tumor sample was mounted on the same slide as the tissue section of the tumor sample to be analyzed to serve as negative and positive controls.

Antigen retrieval was performed for 20 minutes with EDTA (100 °C), and slides were incubated for 25 minutes with HER-2 rabbit monoclonal antibody (clone SP3, Neomarkers, Fremont, CA, USA; dilution 1:50). Peroxidase blocking was performed during 5 minutes, and slides were subsequently incubated with poly-HRP (ready-to-use) for 8 minutes, with DAB for 10 minutes and counterstained with hematoxylin for 5 minutes to visualize cell nuclei.

HER-2 protein expression was scored by consensus of 2 observers as negative (0/1+), equivocal (2+) or positive (3+), according to the 2007 ASCO/CAP guidelines [5]. First, either a general or a breast pathologist assessed and scored the staining in his/her daily routine. Thereafter, every invasive breast cancer case was reviewed by a breast pathologist before being discussed at the multidisciplinary meeting. In case of a discrepancy, a second breast pathologist was consulted, resulting in a definite consensus score. Samples that were originally reported to comprise a range of scores (i.e. 1-2+) were reviewed for the sake of this study by a breast pathologist (HS) in order to provide an unambiguous score. DCIS areas were excluded from the evaluation, and cytoplasmic staining was ignored.

Multiplex Ligation-Dependent Probe Amplification (MLPA)

Invasive tumor areas as identified on serial H&E sections were harvested from 1-2 whole 4 μm thick paraffin sections (corresponding to approximately 1-2

square cm tumor tissue) using a scalpel as before [19,24]. Paraffin sections containing normal breast and blood (UMC Utrecht) and normal lymph nodes (Symbiant) were taken along for normalization. DNA was isolated from these tissue fragments by a direct lysis method with proteinase K (Roche, Almere, The Netherlands). After centrifugation, DNA was used in the MLPA analysis, according the manufacturers' instructions, using the P078-B1 probemix (MRC Holland, Amsterdam, The Netherlands) [25]. All tests were performed in duplicate in an ABI 9700 PCR machine (Applied Biosystems, Foster City, CA, USA) or a Tprofessional thermocycler (Biometra, Goettingen, Germany). The PCR products were analyzed on an ABI3730 or ABI310 capillary sequencer (Applied Biosystems). Normal breast tissue was used as a control.

Gene copy numbers were analyzed using Genemapper (Applied Biosystems) and Coffalyser (version 7.0) software (MRC-Holland). To confirm validity of the HER-2 amplification results, copy number ratio results of 11 internal reference probes, included in the P078-B1 probemix, were checked. If ≥ 2 reference probes were aberrant, test results were considered invalid or inconclusive.

Copy number ratio for every probe (including HER-2 probes) was obtained by dividing the relative peak height for each probe in the tumor tissue by the relative value of the same peak for the reference DNA samples. To make the normalization robust the algorithm makes use of every MLPA probe signal, set as a reference probe for normalization to produce an independent relative ratio. All ratios were finally normalized by setting the median of the tumor to reference DNA copy number ratios of the reference probes in the probe mixture to 1.0 [26].

The mean of all 4 HER-2 probe copy number peaks in duplicate was calculated. If the MLPA copy number ratio was < 1.3 HER-2 status was defined as normal, a value between 1.3 and 2.0 as equivocal and values > 2.0 as amplified [27].

Chromogenic *in situ* hybridization (CISH)

CISH was performed in case of technical problems with MLPA (aberrant copy numbers of reference probes, unreliable duplicates, poor quality of the DNA) or if the MLPA result was equivocal (MLPA value between 1.3 and 2.0) or discordant with the IHC score. In these cases, the CISH result determined the definite amplification status.

Moreover, because MLPA is a non-morphologic technique, CISH was performed as a primary amplification test if invasive tumor could not be separated from DCIS or if the tumor had heterogeneous protein expression by IHC (Figure 1).

In the selected cases, CISH analysis was performed using the ZytoDot SPEC HER-2 Probe Kit (ZytoVision, Bremerhaven, Germany) or the FDA-approved SPoT-Light HER-2 CISH kit (Invitrogen, Carlsbad, CA, USA), according to the manufacturers' instructions. A positive control was included in each CISH run and consisted of paraffin sections of a case known to be HER-2 amplified by CISH. Normal cells on the same slide, containing 2 copies, served as a "negative" control.

Per sample, nuclei of 30 randomly selected invasive breast cancer cells were counted. Samples with an average of < 4 spots / nucleus were considered unamplified and samples with > 6 spots / nucleus were considered amplified. If the average copy number was 4-6 spots / nucleus, nuclei of 30 additional cells were counted. An average of < 5 spots / nucleus was in these cases considered unamplified and an average of > 5 spots / nucleus was considered amplified [5,28].

The centromeric region of chromosome 17 (CEP-17) was used as an internal control. However, because polysomy 17 is believed to be a very rare event in breast cancer, and its clinical relevance is deemed questionable, we did not use the HER-2/CEP-17 ratio for result interpretation [29,30]. Moreover, in our experience, HER-2/CEP-17 ratio did not correlate as well with MLPA scores as the average HER-2 copy number did. We analyzed CEP-17 copy number gain in all 30 cases for which CEP-17 copy number was mentioned in the pathology report.

Statistical analysis

The Chi-squared test statistic was used to compare the frequencies of 0/1+, 2+ and 3+ IHC scores between the 4 laboratories. Furthermore, the frequencies of HER-2 gene amplification were compared between the 4 laboratories using Chi-squared.

The concordance between protein expression determined by IHC and gene amplification assessed either by CISH, MLPA or both was assessed using the unweighted Cohen's Kappa, not taking the IHC equivocal cases into account.

Table 1. Concordance between MLPA and CISH in the validation study (n=99).

MLPA	CISH No amplification	Amplification	na	Total
No amplification	66 (98.5%)	-	1 (1.5%)	67 (67.7%)
Amplification	-	15 (100%)	-	15 (15.1%)
na / equivocal	11 (64.7%)	-	6 (35.3%)	17 (17.2%)
Total	77 (77.8%)	15 (15.1%)	7 (7.1%)	99

na: not assessable

Table 2. Concordance between MLPA/CISH and IHC in the validation study (n=99).

IHC	MLPA/CISH No amplification	Amplification	na	Total
Negative (0-1+)	53 (94.6%)[a]	-	3 (5.4%)	56 (56.6%)
Equivocal (2+)	24 (77.4%)[b]	4 (12.9%)	3 (9.7%)	31 (31.3%)
Positive (3+)	1 (8.3%)	11 (91.7%)	-	12 (12.1%)
Total	78 (78.8%)	15 (15.1%)	6 (6.1%)	99

na: not assessable

[a] 6/53 (11.3%) lacked amplification by CISH following an equivocal or na MLPA result.

[b] 5/24 (20.8%) lacked amplification by CISH following an equivocal or na MLPA result.

A Kappa value of 0.00-0.20 indicates a slight agreement, 0.21-0.40 suggests a fair agreement, 0.41-0.60 suggests a moderate agreement, 0.61-0.80 implies a substantial agreement, and finally a Kappa value 0.81-1 indicates a perfect agreement. P-values < 0.05 were considered statistically significant. All reported p-values are 2-sided.

Results

MLPA validation study with CISH and IHC

We observed 100% concordance between MLPA and CISH in our validation study, if both methods were assessable and conclusive (81.8% of cases) (Table 1). Concordance between MLPA/CISH and IHC was 94.1% (64/68) with a Kappa value of 0.95 (95% CI 0.84-1.00), not taking the IHC equivocal cases into account (Table 2).

Interlaboratory differences in IHC scores

Table 3 illustrates the percentages of cases scored as 0/1+, 2+ or 3+ by IHC in the 4 pathology laboratories. The percentages of cases scored as 0/1+ and 2+ varied significantly between the laboratories ($p < 0.0001$ for both scores). On the other hand, the percentages of cases scored as positive (3+) were similar ($p = 0.974$).

MLPA and CISH for HER-2 amplification

The gene amplification test protocol presented in this study, comprising MLPA (and CISH in selected cases), gave 98.6% (773/784) conclusive results. First-line CISH was performed in 26 cases and was conclusive in 25 cases (96.2%). MLPA was used as a primary test to determine HER-2 gene amplification status in 758 cases and resulted in a conclusive and unequivocal result in 671 cases (88.5%). In the remaining 87 cases (11.5%) MLPA was either not assessable or equivocal. In 81/87 of these cases reflex CISH was performed, and in the remainder CISH could not be performed due to an insufficient amount of tumor tissue. CISH gave conclusive results in 77/81 cases (95.1%). In the remaining 4 cases CISH was not assessable due to poor tissue quality by bone decalcification or poor fixation. Of the 49 cases that had an equivocal MLPA result, reflex CISH demonstrated no amplification in 34 cases (69.4%) and amplification in 15 cases (30.6%) (Figure 2).

We reviewed 30 cases for which the CEP-17 copy number was included in the pathology report. Three of them showed CEP-17 polysomy. In 1/3 use of the HER-2/CEP-17 ratio would have resulted in no amplification, whereas both HER-2 copy number and MLPA copy number ratio showed amplification.

Concordance between IHC and MLPA/CISH

The concordance rates between IHC and MLPA/CISH were determined for a subset of 784/937 cases (83.7%) of which HER-2 gene amplification status was determined (Table 4). Of the 490 IHC negative cases, 478 (97.6%) were not amplified, 4 (0.8%) presented with amplification, and 8 (1.6%) were not assessable by either MLPA or CISH. Of the 189 IHC equivocal cases, 171 (90.5%) were not amplified, 17 (9.0%) were amplified, and 1 (0.5%) was not assessable. Of the 105 IHC positive cases, 7 (6.7%) failed to demonstrate amplification, 97 (92.3%) were amplified, and 1 case (1.0%) was not assessable. Eleven cases (1.9%) in total demonstrated discordances between IHC and MLPA/CISH. The overall concordance between IHC and MLPA/CISH was

Table 3. Comparison of HER-2 IHC scoring percentages between 4 Dutch pathology laboratories.

Laboratory	A (n=345)	B (n=255)	C (n=122)	D (n=215)	Total (n=937)
Negative (0-1+)	200 (58.0%)	183 (71.7%)	87 (71.3%)	168 (78.1%)	638 (68.1%) *
Equivocal (2+)	106 (30.7%)	42 (16.5%)	21 (17.2%)	20 (9.3%)	189 (20.2%) *
Positive (3+)	39 (11.3%)	30 (11.8%)	14 (11.5%)	27 (12.6%)	110 (11.7%)

* statistically significant ($p < 0.05$)

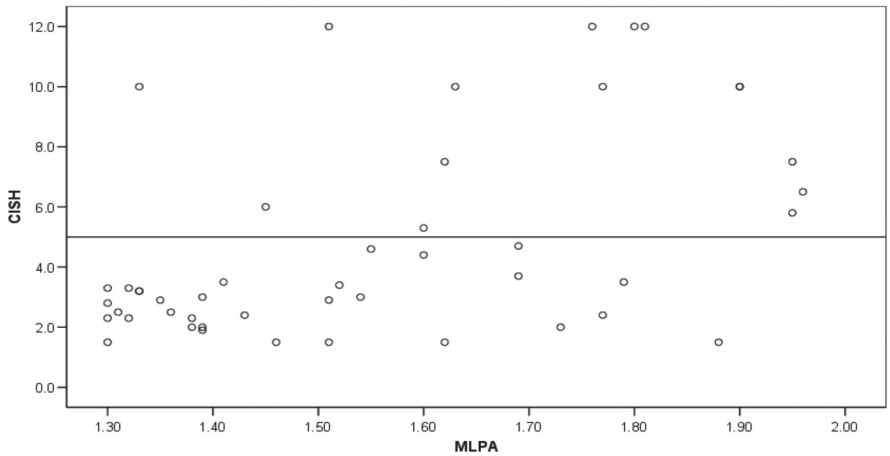


Figure 2. Scatterplot demonstrating MLPA and reflex CISH scores of 47/49 MLPA equivocal cases. Exact CISH copy numbers of 2 cases could not be retrieved. Reflex CISH demonstrated amplification in 15/49 (30.6%) equivocal cases.

98.1% (575/586) with a Kappa value of 0.94 (95% CI 0.90-0.97), not taking the IHC 2+ cases into account.

In total, 15.1% (118/784) of the cases tested revealed HER-2 gene amplification and were thus eligible for trastuzumab. The observed percentages of cases presenting with gene amplification did not vary significantly between the 4 laboratories (13.3%, 19.2%, 15.5% and 13.5% in hospitals A, B, C and D respectively; p=0.287).

Trastuzumab would have been considered in 13.5% (127/937) of the patients if the 2007 ASCO/CAP guidelines [5] were applied. These 127 patients comprised 110 patients with protein overexpression at the 3+ level (merely 105 of these

Table 4. Concordance between HER-2 protein expression determined by IHC and amplification determined by MLPA and/or CISH (n=784).

IHC	MLPA/CISH No amplification	Amplification	na	Total
Negative (0-1+)	478 (97.6%)	4 (0.8%)	8 (1.6%)	490
Equivocal (2+)	171 (90.5%)	17 (9.0%)	1 (0.5%)	189
Positive (3+)	7 (6.7%)	97 (92.3%)	1 (1.0%)	105
Total	656 (83.7%)	118 (15.1%)	10 (1.2%)	784

na: not assessable

Table 5. Retrospective review of 11 discordant cases.

Case number	IHC	MLPA (copy number ratio)	CISH (copy number)	Review IHC	Review CISH	Repeated IHC staining
D2 [a]	1+	A (2.47)	nt	0	na	0 [b]
A1 [a]	1+	Eq (1.60)	A (5.3)	1+	A (5.4)	1+ [b]
D1	1+	A (3.55)	A (>10.0)	2+	A (>10.0)	
B1	1+	A (2.31)	nt	3+	A (12.0)	
D4	3+	Eq (1.62)	NA (1.0-2.0, hg 7.0-8.0 (~20%))	1+, hg 2+	NA (2.8)	
A3	3+	NA (1.24)	NA (3.5)	2+	NA (3.0)	
D3	3+	Eq (1.79)	NA (3.5)	2+	NA (3.5)	
A2	3+	NA (1.08)	nt	2+	NA (2.7)	
C1	3+	NA (0.99)	nt	2+	NA (2.0)	
B2 [a]	3+	Eq (1.77)	NA (2.4)	3+	NA (2.4)	3+ [b]
B3 [a]	3+	Eq (1.52)	NA (3.4)	3+	NA (4.98)	2+

A, amplification; Eq, Equivocal; hg, heterogeneous; NA, no amplification; na, not assessable; nt, not tested.

[a] Amplification status would not have been determined properly when applying the ASCO/CAP protocol.

[b] True discordant cases.

were co-tested by MLPA/CISH) and 17 cases with IHC overexpression at the 2+ level with confirmed gene amplification.

Review of IHC/amplification discordant cases

The IHC sections of the 11 discordant cases were retrospectively reviewed by a breast pathologist (HS) and the CISH sections by a molecular technician (AH)

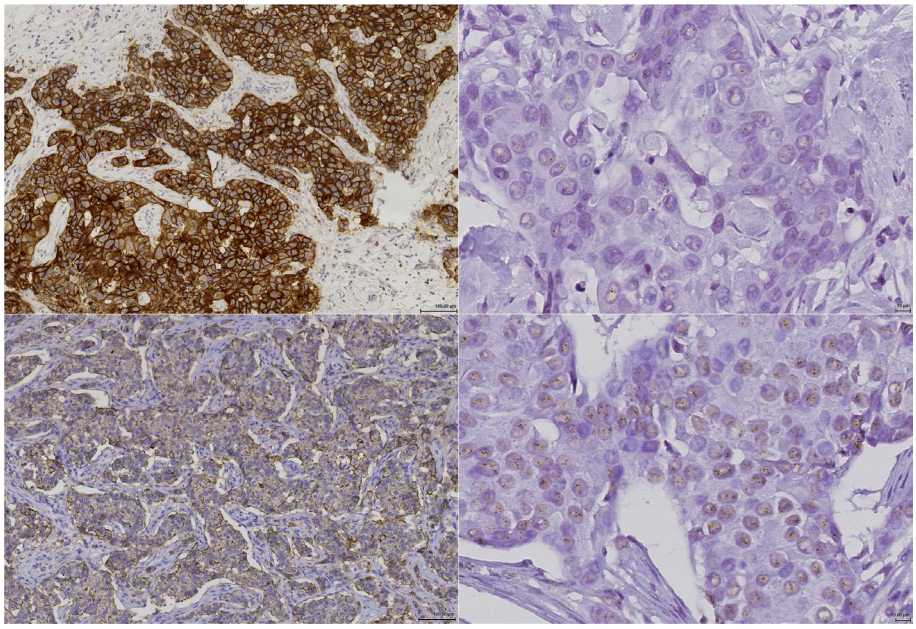


Figure 3. Representative images of 2 discordant cases (B2 and A1). Upper left: IHC of case B2 scored as 3+. Upper right: no amplification with CISH in case B2. Lower left: IHC of case A1 scored as 1+. Lower right: amplification with CISH in case A1. IHC images: 10x magnification. CISH images: 40x magnification.

to ensure or deny real discordance between IHC and MLPA/CISH (Table 5). In 2/4 (50%) of the amplified cases originally scored as IHC 0/1+, review of the IHC score led to an adjustment to 2+ or 3+. Also, in 5/7 (71%) of the non-amplified cases originally scored as IHC 3+, IHC score was altered to 2+ or 1+ on review. CISH score was not adjusted in any of the initially discordant cases. In 4/11 discordant cases (D2, A1, B2 and B3) results remained discordant after retrospective review. As a final confirmation of the 4 discordant cases, we repeated IHC staining at a single location (laboratory A). Review of these IHC sections by a breast pathologist (HS) eventually led to the adaption of one more non-amplified 3+ case into a 2+ score (case B3), leaving 3 true discordant cases. Representative images of discordant cases A1 and B2 are presented in Figure 3.

Discussion

Although IHC is the most commonly used first-line test for examining HER-2 status, HER-2 assessment by IHC can be problematic. First, IHC scores

vary depending on the antibodies used [9]. Second, IHC is sensitive to pre-analytical variation, including tissue fixation [31]. This problem may especially be troublesome in laboratories serving several hospitals (like Symbiant), with inevitably suboptimal or delayed fixation during tissue transportation. This problem is expected to be exacerbated by the ongoing centralization of pathology services. Finally, IHC scoring is subjective and thus shows interobserver and intraobserver variability [6-8]. Gene amplification tests may be less sensitive to these sources of error. As a second line test in case of an equivocal IHC result, ISH is now commonly applied, but it is relatively laborious and expensive. The quantitative MLPA technique has proven to be a reliable, less expensive and high-throughput test, and can therefore be a cost-effective alternative to ISH [11-14].

Our study aimed to prospectively assess the added value of co-testing every invasive breast cancer for HER-2 using a combination of IHC and MLPA in routine pathology practice. We analyzed data from consecutive cohorts of 937 invasive breast cancers derived from 4 Dutch pathology laboratories.

Despite standardized IHC protocols and double-review, our data reveal significant interlaboratory variation in the percentages of cases scored as IHC 0/1+ and 2+. This variation clearly shows the pre-analytical, tissue processing and interpretation problems associated with IHC. The percentages of IHC 3+ cases were, however, quite comparable between the laboratories. Among biopsies, high rates of false-positive IHC scores were reported in several publications [32,33]. In our cohorts, we observed a significantly higher percentage of IHC 3+ cases among the biopsies than among the resections (17.7% vs. 10.2%; $p = 0.004$). No further effort was done to determine whether these 3+ cases in biopsies contained more false-positives compared to these in resections.

MLPA as a single test produced conclusive results in 88.5% of cases. Our combined MLPA/CISH strategy, where the result of reflex CISH determined the definite amplification status, produced conclusive results in 98.6% of cases. Reflex CISH failed in 10/87 (11.5%) of cases, because of insufficient tumor material, bone decalcification or poor tissue fixation. Reflex CISH revealed an absence of amplification in 69.4% of the cases with equivocal MLPA results, and it was thereby a useful addition to MLPA, especially in the borderline cases.

We did not use the HER-2/CEP-17 ratio for CISH result interpretation. Several studies state that polysomy 17 is a very rare event in breast cancer, and its clinical relevance is deemed questionable [29,30]. Moelans *et al* showed that CEP-17 copy number was unrelated to the gains and losses of 17 genes located on chromosome 17, as determined by MLPA [30]. In our 3 cohorts from Symbiant, only 3/30 cases showed CEP-17 polysomy, and in 1 case use of the HER-2/CEP-17 ratio would have resulted in no amplification, whereas both HER-2 copy number and MLPA copy number ratio showed amplification. Unfortunately, the CEP-17 copy number was counted and reported in 30 cases only.

The percentage of HER-2 amplified cases in our study (15.1%) was similar to that in previous studies [12,34]. Choritz *et al* [34] demonstrated an average of 14.6% (ranging from 7.6% to 31.6%), calculated with the results from 42 pathology laboratories. Moelans *et al* [12] detected HER-2 amplification in 14% of 518 invasive breast cancers. Moreover, in our study, the percentages of HER-2 amplified cases were similar between the 4 laboratories.

The 2007 ASCO/CAP guidelines [5] state that at most 5% of cases without protein overexpression should present with HER-2 gene amplification, and the percentage of cases with highly overexpressed protein lacking HER-2 gene amplification should not exceed 10%. The data presented in our study are in compliance with these guidelines, as 0.8% of cases scored as IHC 0/1+ revealed HER-2 gene amplification, whereas 6.7% of cases scored as IHC 3+ lacked gene amplification. Review of the 11 discordant cases revealed true discordance in only 3 cases, while in the remainder the original IHC score was incorrect or adapted after repeated IHC staining. Discordance between protein expression and gene amplification described by others ranged from 0% to 11.5% IHC negative / amplified tumors and from 0% to 21.9% IHC positive / non-amplified tumors [7,10-12,15-18].

The central question in view of the discrepancies between IHC and gene amplification tests is whether IHC 3+ patients without HER-2 gene amplification and IHC 0/1+ patients with gene amplification respond to trastuzumab. Thus far, there is no consensus on the gold standard for HER-2 testing, because no technique is perfectly able to identify patients likely to benefit from trastuzumab therapy [5]. Although there are some indications that gene amplification may be a better predictor for response to trastuzumab than protein overexpression [35-39], there are also indications that amplified tumors

particularly respond to trastuzumab if they present with protein overexpression at the 3+ level [37,40,41]. The final answer to this question could probably only be provided by a meta-analysis combining response data for such IHC / amplification discrepant patients from different clinical trastuzumab trials, but so far this has not been executed. Such an analysis is eagerly awaited, in view of the high costs of trastuzumab and its potential side effects.

The question whether laboratories should switch to the use of a first line gene amplification test for all cases is subject of an ongoing debate. Some laboratories already perform frontline ISH testing [42,43]. In Australia, an HER-2 ISH test is required for all early breast cancer patients [44], and patients demonstrating HER-2 gene amplification are eligible for trastuzumab, regardless of their IHC score. The decision to stratify patients according to ISH results was based on guidelines [39] preferring ISH over IHC because of its greater test accuracy, objectivity and reproducibility.

Performing a reflex gene amplification test routinely in all patients has several advantages. First, gene amplification testing seems to be less sensitive to pre-analytical factors, is more quantitative and easier to interpret, and may therefore reduce variation between labs. Second, it may serve as quality control for IHC scoring as illustrated in our study, where IHC score was, on review, adjusted in over half of the discrepant cases after gene amplification testing. Third, it will reveal gene amplified cases among the IHC 0/1+ patients that will likely respond to trastuzumab, and who would possibly be denied an effective therapy based on IHC alone. Fourth, it may identify IHC 3+ patients without gene amplification, who may respond less well to trastuzumab, although further evidence is required here as discussed above.

In order to control costs of HER-2 co-testing, a low-cost high-throughput gene amplification test like MLPA is demanded. MLPA reagent costs are approximately 30% lower than CISH reagent costs. Although the hands-on time to perform the tests is similar for MLPA and CISH, MLPA analysis time is approximately 5 times shorter than CISH analysis time. Furthermore, in contrast to CISH, which needs to be double-reviewed by a technician and a pathologist, MLPA produces a directly interpretable quantitative value and does not involve analysis by a pathologist. A qualified pathologist should, however, mark areas of invasive breast cancer to be included in the MLPA analysis. MLPA is thereby particularly cost-effective for the analysis of large sample sizes and is accordingly an ideal technique for HER-2 co-testing.

In conclusion, assessment of HER-2 gene amplification by MLPA is a low-cost, quantitative and high-throughput technique with near perfect concordance with CISH. When routinely applied in all breast cancer patients, it may improve the quality of HER-2 testing by reducing variation in HER-2 testing between labs, serving as a quality control for IHC, identifying gene amplified cases among the IHC 0/1+ patients that will likely respond to trastuzumab as well as IHC 3+ cases without gene amplification that may respond less well to trastuzumab.

References

1. Ross JS, Slodkowska EA, Symmans WF, Pusztai L, Ravdin PM, Hortobagyi GN. (2009) The HER-2 receptor and breast cancer: ten years of targeted anti-HER-2 therapy and personalized medicine. Oncologist 14(4): 320-68.
2. Jukkola A, Bloigu R, Soini Y, Savolainen ER, Holli K, Blanco G. (2001) c-erbB-2 positivity is a factor for poor prognosis in breast cancer and poor response to hormonal or chemotherapy treatment in advanced disease. Eur J Cancer 37(3): 347-54.
3. Dahabreh IJ, Linardou H, Siannis F, Fountzilas G, Murray S. (2008) Trastuzumab in the adjuvant treatment of early-stage breast cancer: a systematic review and meta-analysis of randomized controlled trials. Oncologist 13(6): 620-30.
4. Ahn ER, Vogel CL. (2012) Dual HER2-targeted approaches in HER2-positive breast cancer. Breast Cancer Res Treat 131(2): 371-83.
5. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. (2007) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol 25(1): 118-45.
6. Hoang MP, Sahin AA, Ordonez NG, Sneige N. (2000) HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. Am J Clin Pathol 113(6): 852-9.
7. Press MF, Sauter G, Bernstein L, Villalobos IE, Mirlacher M, Zhou JY, et al. (2005) Diagnostic evaluation of HER-2 as a molecular target: an assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. Clin Cancer Res 11(18): 6598-607.
8. Perez EA, Suman VJ, Davidson NE, Martino S, Kaufman PA, Lingle WL, et al. (2006) HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial. J Clin Oncol 24(19): 3032-38.
9. Rosa FE, Santos RM, Rogatto SR, Domingues MA. (2013) Chromogenic in situ hybridization compared with other approaches to evaluate HER2/neu status in breast carcinomas. Braz.J Med Biol Res 46(3): 207-16
10. Moerland E, van Hezik RL, van der Aa TC, van Beek MW, van den Brule AJ. (2006) Detection of HER2 amplification in breast carcinomas: comparison of Multiplex Ligation-dependent Probe Amplification (MLPA) and Fluorescence In Situ Hybridization (FISH) combined with automated spot counting. Cell Oncol 28(4): 151-9.
11. Purnomosari D, Aryandono T, Setiaji K, Nugraha SB, Pals G, van Diest PJ. (2006) Comparison of multiplex ligation dependent probe amplification to immunohistochemistry for assessing HER-2/neu amplification in invasive breast cancer. Biotech.Histochem. 81(2-3): 79-85.
12. Moelans CB, de Weger RA, van Blokland MT, Ezendam C, Elshof S, Tilanus MG, et al. (2009) HER-2/neu amplification testing in breast cancer by multiplex ligation-dependent probe amplification in comparison with immunohistochemistry and in situ hybridization. Cell Oncol 31(1): 1-10.
13. Farshid G, Cheetham G, Davies R, Moore S, Rudzki Z. (2011) Validation of the multiplex ligation-dependent probe amplification (MLPA) technique for the determination of HER2 gene amplification in breast cancer. Diagn Mol.Pathol 20(1): 11-7.
14. Bravaccini S, Rengucci C, Medri L, Zoli W, Silvestrini R, Amadori D. (2012) Detection of HER2 and Topo 2 in breast cancers: comparison between MLPA and FISH approaches. J Clin Pathol 65(2): 183-5.

15. Owens MA, Horten BC, Da Silva MM. (2004) HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. Clin Breast Cancer 5(1): 63-9.
16. Dybdal N, Leiberman G, Anderson S, McCune B, Bajamonde A, Cohen RL, et al. (2005) Determination of HER2 gene amplification by fluorescence in situ hybridization and concordance with the clinical trials immunohistochemical assay in women with metastatic breast cancer evaluated for treatment with trastuzumab. Breast Cancer Res Treat 93(1): 3-11.
17. Ellis CM, Dyson MJ, Stephenson TJ, Maltby EL. (2005) HER2 amplification status in breast cancer: a comparison between immunohistochemical staining and fluorescence in situ hybridisation using manual and automated quantitative image analysis scoring techniques. J Clin Pathol 58(7): 710-4.
18. Prati R, Apple SK, He J, Gornbein JA, Chang HR. (2005) Histopathologic characteristics predicting HER-2/neu amplification in breast cancer. Breast J 11(6): 433-9.
19. Moelans CB, Kibbelaar RE, van den Heuvel MC, Castigliego D, de Weger RA, van Diest PJ. (2010) Validation of a fully automated HER2 staining kit in breast cancer. Cell Oncol 32(1-2): 149-55.
20. Central Committee on Research involving Human Subjects (Centrale Commissie Mensgebonden Onderzoek) Available from: http://www.ccmo.nl.
21. van Diest PJ. (2002) No consent should be needed for using leftover body material for scientific purposes. For. BMJ 325(7365): 648-51.
22. Federal Policy for the Protection of Human Subjects. Available from: http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/ucm118862.htm
23. Gedragscode verantwoord omgaan met lichaamsmateriaal ten behoeve van wetenschappelijk onderzoek. Available from: http://www.kncv.nl/Uploads/2011/9/pdf-Code-Goed-Gebruik-met-logo-Federa-Coreon.pdf.
24. Moelans CB, de Weger RA, Ezendam C, van Diest PJ. (2009) HER-2/neu amplification testing in breast cancer by Multiplex Ligation-dependent Probe Amplification: influence of manual- and laser microdissection. BMC Cancer 9: 4.
25. Product description SALSA MLPA KIT P078-B1 Breast Tumour
26. Coffa J, van den Berg J. Analysis of MLPA Data Using Novel Software Coffalyser.NET by MRC-Holland. 2011; Available from: http://www.intechopen.com/books/modern-approaches-to-quality-control/analysis-of-mlpa-data-using-novel-software-coffalyser-net-by-mrc-holland.
27. Moelans CB, de Weger RA, van Diest PJ. (2010) Multiplex ligation-dependent probe amplification to detect HER2 amplification in breast cancer: new insights in optimal cut-off value. Cell Oncol 32(4): 311-2.
28. van de Vijver M, Bilous M, Hanna W, Hofmann M, Kristel P, Penault-Llorca F, et al. (2007) Chromogenic in situ hybridisation for the assessment of HER2 status in breast cancer: an international validation ring study. Breast Cancer Res 9(5): R68.
29. Yeh IT, Martin MA, Robetorye RS, Bolla AR, McCaskill C, Shah RK, et al. (2009) Clinical validation of an array CGH test for HER2 status in breast cancer reveals that polysomy 17 is a rare event. Mod.Pathol 22(9): 1169-75.
30. Moelans CB, de Weger RA, van Diest PJ. (2010) Absence of chromosome 17 polysomy in breast cancer: analysis by CEP17 chromogenic in situ hybridization and multiplex ligation-dependent probe amplification. Breast Cancer Res Treat 120(1): 1-7.

31. Middleton LP, Price KM, Puig P, Heydon LJ, Tarco E, Sneige N, et al. (2009) Implementation of American Society of Clinical Oncology/College of American Pathologists HER2 Guideline Recommendations in a tertiary care facility increases HER2 immunohistochemistry and fluorescence in situ hybridization concordance and decreases the number of inconclusive cases. Arch Pathol Lab Med 133(5): 775-80.
32. Taucher S, Rudas M, Mader RM, Gnant M, Dubsky P, Roka S, et al. (2004) Prognostic markers in breast cancer: the reliability of HER2/neu status in core needle biopsy of 325 patients with primary breast cancer. Wien.Klin Wochenschr 116(1-2): 26-31.
33. Seferina SC, Nap M, van den Berkmortel F, Wals J, Voogd AC, Tjan-Heijnen VC. (2013) Reliability of receptor assessment on core needle biopsy in breast cancer patients. Tumour.Biol 34(2): 987-94.
34. Choritz H, Busche G, Kreipe H, Study Group HER2 Monitor. (2011) Quality assessment of HER2 testing by monitoring of positivity rates. Virchows Arch 459(3): 283-9.
35. Pauletti G, Dandekar S, Rong H, Ramos L, Peng H, Seshadri R, et al. (2000) Assessment of methods for tissue-based detection of the HER-2/neu alteration in human breast cancer: a direct comparison of fluorescence in situ hybridization and immunohistochemistry. J Clin Oncol 18(21): 3651-64.
36. Seidman AD, Fornier MN, Esteva FJ, Tan L, Kaptain S, Bach A, et al. (2001) Weekly trastuzumab and paclitaxel therapy for metastatic breast cancer with analysis of efficacy by HER2 immunophenotype and gene amplification. J Clin Oncol 19(10): 2587-95.
37. Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, et al. (2002) Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol 20(3): 719-26.
38. Sauter G, Lee J, Bartlett JM, Slamon DJ, Press MF. (2009) Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations. J Clin Oncol 27(8): 1323-33.
39. Mass RD, Press MF, Anderson S, Cobleigh MA, Vogel CL, Dybdal, et al. (2005) Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. Clin Breast Cancer 6(3): 240-6.
40. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344(11): 783-92.
41. Burstein HJ, Harris LN, Gelman R, Lester SC, Nunes RA, Kaelin CM, et al. (2003) Preoperative therapy with trastuzumab and paclitaxel followed by sequential adjuvant doxorubicin/cyclophosphamide for HER2 overexpressing stage II or III breast cancer: a pilot study. J Clin Oncol 21(1): 46-53.
42. Blank PR, Schwenkglenks M, Moch H, Szucs TD. (2010) Human epidermal growth factor receptor 2 expression in early breast cancer patients: a Swiss cost-effectiveness analysis of different predictive assay strategies. Breast Cancer Res Treat 124(2): 497-507.
43. Clay MR, Iberri DJ, Bangs CD, Cherry A, Jensen KC. (2013) Clinicopathologic characteristics of HER2 FISH-ambiguous breast cancer at a single institution. Am J Surg Pathol 37(1): 120-7.
44. Bilous M, Morey AL, Armes JE, Bell R, Button PH, Cummings MC, et al. (2012) Assessing HER2 amplification in breast cancer: findings from the Australian In Situ Hybridization Program. Breast Cancer Res Treat 134(2): 617-24.

The value of autopsies in the era of high-tech medicine: discrepant findings persist

Chantal C.H.J. Kuijpers*
Judith Fronczek*
Frank R.W. van de Goot
Hans W.M. Niessen
Paul J. van Diest
Mehdi Jiwa

*These authors equally contributed to the manuscript

Published in J Clin Pathol 2014; 67(6): 512-519

Abstract

Introduction

Although the autopsy is still the gold standard for quality assessment of clinical diagnoses, autopsy rates have been declining over the last decades to < 10%. The aim of this study was to investigate the value of autopsies in the high-tech medicine era by determining the frequency of discrepancies between clinical and autopsy diagnoses.

Methods

We classified all adult autopsy cases (n=460), performed at Symbiant, Pathology Expert Centre, in 2007 and 2012/13, as having major-, or minor discrepancy or total concordance. The roles of possible contributory factors were analyzed. Finally, we assessed the role of microscopic examination in identifying cause of death.

Results

Major and minor discrepancies were found in 23.5% and 32.6% of the classifiable autopsies, respectively. Most commonly observed major discrepancies were myocardial infarction, pulmonary embolism and pneumonia. Improper imaging and discontinuation of active treatment were significantly associated with a higher and a lower frequency of major discrepancies, respectively. Comparing 2007 and 2012/13, the frequency of minor discrepancies significantly increased from 26.8% to 39.3%. Final admission length of > 2 days was significantly associated with a lower frequency of class III minor discrepancies. Microscopic examination contributed to establishing cause of death in 19.6% of the cases.

Conclusions

Discrepant findings persist at autopsy, even in the era of high-tech medicine. Therefore, autopsies still should serve as a very important part of quality control in clinical diagnosis and treatment. Learning from both individual and system-related diagnostic errors can aid in improving patient safety.

Introduction

The autopsy is for long been regarded as the "gold standard" as the most important tool for retrospective quality assessment of clinical diagnoses as well as a key educational tool [1]. This is evident from previous studies comparing clinical diagnoses and autopsy findings, which revealed major discrepancies in approximately 25% of the deceased patients that underwent postmortem examination [2,3].

However, throughout the world, autopsy rates have been declining over the past few decades [4-6]. Reasons for this decline include the non-reimbursement of autopsies, clinicians' fear of medicolegal problems, and advances in laboratory testing and imaging techniques that often results in the belief among clinicians that the autopsy had become redundant.

We assessed the value of autopsies by determining the major and minor discrepancy rates in a total of 460 consecutive autopsy cases, divided over two time periods. In the most recent time period, the majority of autopsies was performed by a specialized autopsy pathologist. Furthermore, we analyzed the influence of several factors, including age, sex, length of final admission, and the use of imaging techniques on the frequency of major and minor discrepancies. Finally, we determined the role of microscopic examination in identifying the cause of death (COD).

Methods

Cases and data extraction

We retrospectively reviewed all consecutive adult (> 18 years) autopsy cases, performed at the 3 locations of Symbiant, Pathology Expert Centre (Alkmaar Medical Centre, Zaandam Medical Centre and Westfriesgasthuis Hoorn) from 2007 and from 2012 on up to July 2013. Partial autopsies restricted to certain parts of the body (e.g. brain, thorax) were excluded, as well as autopsies from other local hospitals whose patient charts were not available to us. Autopsies requested by general practitioners or other primary care providers were included, but analyzed separately as 'external autopsies'. In 2007, all autopsies were performed by general pathologists. Starting from April 2011, 3 specialized autopsy pathologists performed the majority of the autopsies.

All clinical and post-mortem diagnoses were recorded. Clinical diagnoses were extracted from the clinical information written on the autopsy request form, and from patient charts including clinicians' letters directed to the general practitioner, the medical history and radiology results. Post-mortem diagnoses were extracted from macroscopic and microscopic autopsy findings described in the autopsy report.

From every case, the following data were recorded: age, sex, length of final admission, whether imaging techniques (MRI, CT, PET, ultrasound and X-ray) were applied during life not more than one month before death, whether active treatment was discontinued, and the last admission unit. Furthermore, we recorded which pathologist performed the autopsy (autopsy pathologist versus general pathologist), whether the autopsy also included the brain, the post-mortem time, and the time until completion of the preliminary and the final autopsy report.

Imaging

We assessed all cases of patients who underwent imaging in the hospital of final admission not more than one month before death. For these patients we determined whether it was possible to visualize the COD with imaging. If so, we determined whether imaging was applied to the proper part of the body (brain, thorax, abdomen, neck) needed to diagnose the COD, and whether the proper imaging modality was used. For example, an X-ray in case of a pulmonary embolism was considered an improper imaging modality, as the proper imaging modality to diagnose a pulmonary embolism is a computed tomographic pulmonary angiography (CTPA) [7].

Classification of discrepancies

We classified the discrepancies between clinical and post-mortem diagnoses according to the Goldman classification system [8], modified by Battle *et al* [9], as described by Schwanda-Burger *et al* [10]. Major discrepancies (classes I and II) are missed diagnoses related to the COD. Knowledge before death would have changed management of care and could have prolonged survival or cured the patient (class I), or probably would not have changed the outcome (class II). Minor discrepancies (classes III or IV) are not directly related to the COD. Class III includes diseases with symptoms that should have been treated or that would ultimately have affected the prognosis. Class IV includes minor non-diagnosable diseases or events with possible epidemiological or genetic importance. Full concordance was classified as class V, and non-classifiable

cases were assigned class VI. In case of two or more discrepant findings, the case was classified according to the most severe Goldman class. All cases were classified by one specialized autopsy pathologist (JF). For the equivocal cases a senior autopsy pathologist (FG) was consulted. In case of insufficient clinical information, "discrepant" findings were appointed non-classifiable (class VI). In cases where active treatment was withdrawn, we only classified discrepant diseases that certainly or most probably developed before active treatment discontinuation (for example liver cirrhosis or neoplasms). Cases were designated class VI if the time point of origination of the discrepant disease was doubted (for example pneumonia or myocardial infarction).

Role of microscopy

We analyzed the role microscopic examination, of both histochemical and immunohistochemical stainings, played in identifying COD. We determined whether histology contributed to establishing COD (i.e. provided COD, changed COD or added to COD made by macroscopical examination), confirmed COD, or played no role in determining COD. The same classification was used by Fronczek *et al* in their study determining the role of histology in forensic autopsies [11]. Cases were non-classifiable if there was no clearly defined COD reported, if the report lacked either the diagnosis made at macroscopical or at microscopical examination, or if diagnoses made at macroscopical and microscopical examination were not reported separately.

Statistical analysis

Statistical analysis was performed with the SPSS statistics program (Windows version 20). Chi-squared analysis was used to compare the frequencies of discrepancies between the two time periods. Furthermore, we performed logistic regression (OR, 95% CI and p-value) for univariable (UV) and multivariable (MV) analysis. To make sure not to miss any possible contributory factor, all factors with a p-value < 0.2 in UV analysis were included in MV analysis. In MV analysis, p-values < 0.05 were considered statistically significant. A non-parametric median test was used to compare median times to autopsy report completion. All p-values reported are two-sided.

Results

Numbers

A total of 740 autopsies were performed. Autopsy rates decreased from 13.2% in 2007 to 6.6% in 2012/13. Eventually, 460 autopsies were included in this

study. The 280 excluded cases comprised 163 patients under the age of 18 (including foetuses), 108 patients from other local hospitals, 6 partial autopsies (3 brain autopsies, 2 thoracic autopsies and 1 liver autopsy), and 3 cases that were not signed out by the end of the inclusion period.

The included autopsies were divided into two groups, clinical and external autopsies, and analyzed separately. Table 1 summarizes the patient characteristics. The 'clinical autopsies' included 362 patients that were hospitalized or stayed at least 1 hour at the emergency department. The 'external autopsies' included 98 cases submitted by a general practitioner, a nursing home physician or a forensic physician, or patients who had stayed at the emergency department for less than 1 hour.

Discrepant autopsy findings

Table 2 illustrates the frequencies and percentages of Goldman classes in all autopsies (n=460), separately analyzed for the two time periods. Overall, major discrepancies were observed in 18.1% of cases, minor discrepancies in 26.6% of cases, and full concordance was observed in 37.8%. Comparing 2007 and 2012/13, the frequency of major discrepancies decreased (from 20.1% to 16.0%; $p = 0.256$), and the frequency of minor discrepancies significantly increased (from 21.8% to 31.2%; $p = 0.023$). Furthermore, in total 17.6% of cases were non-classifiable, mostly due to insufficient clinical information, which was predominantly seen in the 'external autopsies' (57.1%). Cases where no clear COD had been found or one had not been specified in the report, or where active treatment was withdrawn also qualified as non-classifiable.

In the subgroup of clinical autopsies, 25/362 (6.9%) were non-classifiable. Table 3 shows the percentages of discrepancies in all 337 classifiable clinical autopsy cases (classes I-V), separately analyzed for the two time periods. Overall, major discrepancies were found in 23.5%, minor discrepancies in 32.6%, and full concordance was observed in 43.9%. Comparing 2007 and 2012/13, the frequency of major discrepancies decreased (from 25.2% to 21.6%; $p = 0.434$), and the frequency of minor discrepancies significantly increased (from 26.8% to 39.3%; $p = 0.015$).

Tables 4 and 5 summarize clinical diagnosis (including differential diagnoses) and autopsy diagnoses of all class I and class II discrepant cases, respectively. The most commonly observed major discrepancies were myocardial infarction (n=18), pulmonary embolism (n=15), and pneumonia (n=11). Other common

Table 1. Characteristics of the included autopsies.

	Clinical autopsies		External autopsies	
	2007 (n=195)	2012/13 (n=167)	2007 (n=34)	2012/13 (n=64)
Age				
Average (SD)	71.8 (13.5)	72.2 (11.2)	57.8 (17.7)	59.5 (13.6)
Range	20 – 96	35 – 94	20 - 89	25 - 89
Sex; n (%)				
Male	104 (53.3%)	95 (56.9%)	22 (64.7%)	48 (75.0%)
Female	91 (46.7%)	72 (43.1%)	12 (35.3%)	16 (25.0%)
Length of final admission (days)				
Average (SD)	9.4 (11.1)	9.3 (9.8)	na	na
Range	0 – 60	0 – 44	na	na
Imaging (within 1 month before death); n (%)				
Yes	176 (90.3%)	144 (86.2%)	1 (2.9%)	7 (10.9%)
No	19 (9.7%)	23 (13.8%)	7 (20.6%)	20 (31.1%)
Unknown	-	-	26 (76.5%)	37 (57.8%)
Active treatment discontinuation; n (%)				
Yes	66 (33.8%)	80 (47.9%)	-	-
No	114 (58.5%)	71 (42.5%)	34 (100%)	64 (100%)
Unknown	15 (7.7%)	16 (9.6%)	-	-
Post-mortem time (days)				
Average (SD)	1.1 (0.9)	0.9 (0.8)	1.28 (1.3)	1.39 (1.4)
Range	0 – 4	0 – 3	0 – 4	0 – 7
Pathologist; n (%)				
General pathologist	195 (100%)	77 (46.1%)	34 (100%)	24 (37.5%)
Autopsy pathologist	-	90 (53.9%)	-	40 (62.5%)
Brain autopsy performed; n (%)				
Yes	26 (13.3%)	38 (22.8%)	8 (23.5%)	21 (32.8%)
No	169 (86.7%)	129 (77.2%)	26 (76.5%)	43 (67.2%)
Last admission unit / origin of the patient; n (%)				
Internal medicine	44 (22.6%)	51 (30.5%)	-	-
Intensive care	35 (17.9%)	38 (22.8%)	-	-
Surgery	33 (16.9%)	18 (10.8%)	-	-
Cardiology	25 (12.8%)	9 (5.4%)	-	-
Lung	21 (10.8%)	16 (9.6%)	-	-
Emergency	21 (10.8%)	12 (7.2%)	7 (20.6%)	22 (34.4%)
Geriatrics	9 (4.6%)	5 (3.0%)	-	-
Neurology	5 (2.6%)	9 (5.4%)	-	-
Gastrointestinal	1 (0.5%)	7 (4.2%)	-	-
Orthopedics	1 (0.5%)	1 (0.6%)	-	-
Plastic surgery	-	1 (0.6%)	-	-
General practitioner	-	-	25 (73.5%)	33 (51.6%)
Nursing home	-	-	2 (5.9%)	6 (9.4%)
Forensic physician	-	-	-	3 (4.7%)

na: not applicable

Table 2. Goldman classification for the full group of autopsy cases evaluated for discrepancies between clinical and autopsy diagnoses (n=460), separately analyzed for 2007 and 2012/13.

	Class	Total (%)	2007 Frequency	Percentage		2012/13 Frequency	Percentage	
Major	I	18.1%	26	11.4%	20.1%	18	7.8%	16.0%
	II		20	8.7%		19	8.2%	
Minor	III	26.6%	17	7.4%	21.8%	21	9.1%	31.2%
	IV		33	14.4%		51	22.1%	
	V	37.8%	98	42.8%		76	32.9%	
	VI	17.6%	35	15.3%		46	19.9%	
Total			229			231		

Table 3. Goldman classification for the subgroup of classifiable (classes I-V) clinical autopsy cases (n=337) evaluated for discrepancies between clinical and autopsy diagnoses, separately analyzed for 2007 and 2012/13.

	Class	Total (%)	2007 Frequency	Percentage		2012/13 Frequency	Percentage	
Major	I	23.5%	25	14.0%	25.2%	17	10.8%	21.6%
	II		20	11.2%		17	10.8%	
Minor	III	32.6%	16	8.9%	26.8%	17	10.8%	39.3%
	IV		32	17.9%		45	28.5%	
	V	43.9%	86	48.0%		62	39.2%	
Total			179			158		

major discrepancies were malignancy (n=7), fungal infection (n=6), ruptured aneurysm, aorta dissection or aorta-oesophageal fistula (n=6), acute pancreatitis (n=5), and gastrointestinal perforation, severe bleeding or both (n=5).

The most commonly observed minor discrepancies were benign tumors (n=23), polyps (n=18), cysts (n=16), malignancies that were not contributory to the COD (n=15), gallbladder-/kidney-/prostate stones (n=11), diverticulosis (n=10), liver cirrhosis (n=9), and multinodular goitre (n=6).

Table 4. Comparison of clinical diagnoses (including differential diagnoses) and autopsy diagnoses of all class I major discrepant cases (n = 42).

Clinical (differential) diagnoses	Autopsy diagnoses
1. Gastrointestinal haemorrhage	1. Ruptured aortic aneurysm
2. Cerebrovascular accident	2. Ruptured aortic aneurysm
3. Gastroenteritis	3. Gastrointestinal haemorrhage, ulcus duodeni
4. Sepsis	4. Endocarditis
5. Myocardial infarction, alcohol withdrawal syndrome	5. Acute necrotizing pancreatitis
6. Subdural haematoma, myocardial infarction	6. Subdural haematoma, pulmonary embolism
7. Liver cirrhosis	7. Liver cirrhosis, myocardial infarction, pneumonia
8. Pancreas or liver malignancy, cholangitis, peritonitis	8. Perforated stomach laesion, gastrointestinal haemorrhage
9. Perforated duodenum	9. Pulmonary embolism
10. Non-hodgkin lymphoma, pneumonia	10. Pulmonary embolism
11. Metastatic breast carcinoma	11. Metastatic breast carcinoma, bilateral pneumonia
12. Liver cirrhosis, esophageal varices, gastrointestinal haemorrhage	12. Liver fibrosis, esophageal varices, gastrointestinal haemorrhage, bilateral pneumonia
13. Hypokalaemia-induced arrhythmia	13. Pneumonia
14. Malignancy, pulmonary embolism, cardiac decompensation	14. Cardiac tamponade, uremic pericarditis
15. Mors subita after toe surgery for osteomyelitis	15. Mechanical obstruction aortic valve, esophageal carcinoma
16. Pneumonia, sepsis, diffuse intravasal coagulation	16. Diffuse intravasal coagulation, Aspergillus pneumonia
17. Metastatic frontal sinus carcinoma, pneumonia, pulmonary embolism	17. Metastatic undifferentiated carcinoma, Candida albicans pneumonia
18. Cardiac pathology, malignancy, parasitic infection	18. Pancreatitis, peritonitis
19. Metastatic breast carcinoma, myelodysplastic syndrome, ischaemic bowels, perforated diverticulitis	19. Metastatic breast carcinoma, pulmonary embolism
20. Ischaemic bowels	20. Ischaemic bowels, pulmonary embolism
21. Cerebrovascular accident	21. Myocardial infarction, pleural empyema
22. Pneumonia, myocardial infarction, sepsis	22. Sepsis, pulmonary embolism
23. Blood loss after hip surgery	23. Ruptured aorta
24. Sepsis, cholecystitis, cardiac decompensation	24. Exsanguination from the wound bed of the gall bladder
25. Pneumonia, enterocolitis	25. Pneumonia, sepsis, acute cholecystitis with perforation
26. Sepsis, stomach and duodenum ulcera	26. Pulmonary embolism, Aspergillus pneumonia, ischaemic bowels
27. Mediastinal undifferentiated tumor, sepsis, pleural empyema, arrhythmia, myocardial infarction	27. Hilar undifferentiated carcinoma, pulmonary embolism
28. Pancreatic carcinoma	28. Metastatic pancreatic carcinoma, pneumonia, pulmonary embolism
29. Sepsis, diverticulitis, endocarditis	29. Acute necrotizing pancreatitis
30. Metastatic esophageal carcinoma, pulmonary embolism, myocardial infarction, bowel perforation	30. Pneumonia
31. Acute coronary syndrome, myocardial infarction	31. Ruptured aortic aneurysm
32. Pulmonary embolism after breast lipofilling	32. Fat embolism
33. Sepsis, pneumonia, pleural empyema, pulmonary embolism	33. Metastatic lung carcinoma, pyogenic pericarditis
34. Salmonella sepsis, endocarditis	34. Endocarditis, colitis, acute pancreatitis, Aspergillus pneumonia

Clinical (differential) diagnoses	Autopsy diagnoses
35. Sepsis	35. Pulmonary embolism
36. Blood loss of unknown origin	36. Exsanguination from aortoesophageal fistula
37. Urinary tract infection	37. Pulmonary embolism
38. Cardiac tamponade	38. Aortic dissection, myocardial infarction
39. Candida esophagitis	39. Pneumonia, sepsis
40. Arrhythmia	40. Lymphocytic myocarditis
41. Pneumonia, space occupying lesion lung	41. Lung adenocarcinoma, myocardial infarction, pulmonary embolism
42. Unexplained dyspnea	42. Aspergillus pneumonia, pulmonary embolism, metastatic adenocarcinoma lung

Imaging

Imaging was performed not more than one month before death in 300/337 classifiable clinical autopsy cases (89.0%). In 29.7% of the cases, COD could not have been observed with imaging. The proper imaging modality for the body part needed to determine the COD was applied in 50.7%. Imaging was performed on another body part or with a different imaging modality than needed in 5.3% and 10.3%, respectively.

Table 5. Comparison of clinical diagnoses (including differential diagnoses) and autopsy diagnoses of all class II major discrepant cases (n = 37).

Clinical (differential) diagnoses	Autopsy diagnoses
1. Cardiac decompensation, pneumonia, unspecified infection	1. Myocardial infarction, pulmonary embolism, acute respiratory distress syndrome
2. Pneumonia	2. Malignant mesothelioma
3. Mors subita after hip replacement	3. Myocardial infarction
4. Metastatic lung carcinoma	4. Metastatic lung carcinoma, pneumonia
5. Pneumonia, amyloidosis	5. Pneumonia, lung adenocarcinoma
6. Mors subita after resection of sigmoid carcinoma. Metastases?	6. Myocardial infarction
7. Pneumonia, space occupying lesion intra-abdominal	7. Pneumonia, myocardial infarction
8. Unspecified infection, acute respiratory distress syndrome	8. Pneumonia, endocarditis
9. Cardiac decompensation	9. Myocardial infarction
10. Metastatic non-hodgkin lymphoma	10. Metastatic non-hodgkin lymphoma, myocardial infarction
11. Abdominal hematoma, endocarditis, morbus Kahler	11. Retroperitoneal hematoma, necrotizing adenocarcinoma of the coecum with abcess formation
12. Resection double tumor colon, post-operative intestinal necrosis	12. Ischaemic small intestines, mesenteric vessel thrombosis
13. Acute pancreatitis, sepsis	13. Acute necrotizing pancreatitis, sepsis, myocardial infarction
14. Pneumonia	14. Diffuse alveolar damage
15. Cardiac decompensation, urosepsis	15. Myocardial infarction
16. Mors subita after hip surgery	16. Myocardial infarction
17. Sepsis, myocardial infarction	17. Sepsis, bleeding from stomach ulcera
18. Lung carcinoma, retroperitoneal hematoma	18. Metastatic renal cell carcinoma

Clinical (differential) diagnoses	Autopsy diagnoses
19. Meningitis	19. Subacute meningitis, pulmonary embolism
20. Sepsis, diffuse intravasal coagulation, lung haemorrhage	20. Sepsis, lung haemorrhage, bleeding from esophageal varices
21. Sepsis	21. Peritonitis, perforation stomach ulcus
22. Pancreatic carcinoma, cholangitis, sepsis	22. Pancreatic carcinoma, bile duct adenocarcinoma
23. Ulcus duodeni, cardiac decompensation, cardiac arrest	23. Cardiac decompensation, vascular amyloidosis
24. Intestinal ischaemia, ruptured abdominal aortic aneurysm	24. Intestinal ischaemia, abdominal aortic aneurysm, hepatic infarction, arterial thrombosis (mesenteric, hepatic and pulmonary)
25. Space occupying lesion / malignancy right transsphenoidal orbit, cerebral infarction	25. Meningioma, cerebral infarction, thrombosis carotid artery with Aspergillus infection
28. Cardiac decompensation, malignancy	28. Cardiac decompensation, airway infection
29. Metastatic adenocarcinoma of unknown origin, pneumonia	29. Metastatic non-small cell carcinoma lung, pneumonia, acute respiratory distress syndrome, herpes esophagitis
30. Ischaemic bowel resection, sepsis	30. Sepsis, Aspergillus pneumonia
31. Malignancy, peritonitis	31. Metastatic tumor of unknown origin, peritonitis, myocardial infarction
32. Cardiac decompensation, malignancy	32. Coronary artery thrombosis, myocardial infarction
33. Pneumonia, malignancy	33. Pneumonia, coronary artery thrombosis, myocardial infarction
34. Cardiac decompensation, endocarditis	34. Myocardial infarction
35. Most subita (myocardial infarction?) after ischaemic bowel resection	35. Ischaemic bowels, cardiac ischaemia, peritonitis, pneumonia
36. Sepsis, arrhythmia, bleeding abdominal aortic aneurysm	36. Ruptured abdominal aortic aneurysm, myocardial infarction
37. (Metastatic) lung carcinoma	37. Metastatic non-small cell lung carcinoma, pneumonia

Factors contributing to discrepancies

Tables 6 and 7 show analyses of possible contributory factors to major and minor discrepancies, respectively. The following factors contributed to major discrepancies in UV analysis: advanced age, sex (female > male), length of final admission > 2 days, the use of an improper imaging modality or imaging of a different body part, no active treatment discontinuation, and no brain autopsy included. Factors that contributed to minor discrepancies were active treatment discontinuation, autopsy performed in 2012 or 2013, and autopsy performed by an autopsy pathologist. Because the factors time period and type of pathologist are statistically related, only the most significant factor (time period) was included in the MV analysis.

MV analysis showed that the use of an improper imaging modality or imaging of an improper body part was significantly associated with a higher percentage of major discrepancies. Furthermore, active treatment withdrawal significantly contributed to a lower frequency of major discrepancies and a higher frequency

Table 6. Analysis of possible contributory factors to major discrepancies between clinical and final pathology diagnosis at autopsies.

	UV analysis Odds ratio	95% CI	p-value	MV analysis Odds ratio	95% CI	p-value
Age (years)	1.033	1.009 – 1.057	0.006 *	1.022	0.996 – 1.048	0.098
Sex	1.607	0.968 – 2.668	0.067 *	1.494	0.818 – 2.730	0.192
Length of final admission >2 days	1.497	0.850 – 2.637	0.162 *	1.737	0.857 – 3.517	0.125
Imaging (y/n)	1.662	0.667 – 4.140	0.276			
Imaging						
Proper imaging	1			1		
No imaging	0.726	0.279 – 1.890	0.512	0.739	0.213 – 2.557	0.633
Improper imaging	3.594	1.799 – 7.181	< 0.0001 *	2.851	1.299 – 6.255	0.009 †
Not imagable	0.951	0.497 – 1.817	0.878	0.929	0.448 – 1.930	0.844
Active treatment discontinuation	0.479	0.265 – 0.866	0.015 *	0.458	0.239 – 0.878	0.019 †
Admission unit ‡						
First aid	1		0.466			
Cardiology	1.833	0.576 – 5.831				
Surgery	1.061	0.345 – 3.269				
IC	0.917	0.313 – 2.689				
Internal medicine	1.419	0.513 – 3.929				
Lung	0.632	0.171 – 2.343				
Year 2012/13	0.816	0.491 – 1.357	0.434			
Autopsy pathologist	1.123	0.638 – 1.977	0.688			
Brain autopsy included	0.469	0.212 – 1.036	0.061 *	0.507	0.194 – 1.329	0.167

* p<0.2 were regarded as factors contributing to major discrepancies in UV analysis and included in MV analysis.

† p<0.05 were regarded as factors contributing to major discrepancies in MV analysis.

‡ Departments with >20 autopsy cases were included in this analysis.

MV, multivariable; UV, univariable.

Table 7. Analysis of possible contributory factors to minor discrepancies between clinical and final pathology diagnosis at autopsies.

	UV analysis Odds ratio	95% CI	p-value	MV analysis Odds ratio	95% CI	p-value
Age (years)	1.012	0.993 – 1.031	0.214			
Sex	0.944	0.598 – 1.491	0.805			
Length of final admission >2 days	0.764	0.473 – 1.235	0.272			
Imaging (y/n)	0.680	0.338 – 1.369	0.370			
Active treatment discontinuation	1.991	1.233 – 3.216	0.005 *	1.832	1.124 – 2.986	0.015 †
Admission unit ‡						
First aid	1		0.875			
Cardiology	0.792	0.263 – 2.385				
Surgery	1.337	0.502 – 3.560				
IC	0.968	0.378 – 2.477				
Internal medicine	1.190	0.480 – 2.947				
Lung	1.307	0.456 – 3.743				
Year 2012/13	1.763	1.113 – 2.791	0.016 *	1.608	0.987 – 2.620	0.057
Autopsy pathologist	1.532	0.924 – 2.540	0.098			
Brain autopsy included	1.326	0.737 – 2.387	0.346			

* p<0.2 were regarded as factors contributing to major discrepancies in UV analysis and included in MV analysis.

† p<0.05 were regarded as factors contributing to major discrepancies in MV analysis.

‡ Departments with >20 autopsy cases were included in this analysis.

MV, multivariable; UV, univariable.

Table 8. The role of microscopic examination in identifying COD at autopsies (n=460).

Role of microscopic examination	Frequency	Percentage
Provided/changed added to COD	90	19.6%
Confirmed macroscopic findings	220	47.8%
No role	76	16.5%
Non-classifiable	74	16.1%

of minor discrepancies (based on adjusted OR). Additionally, longer admission length (> 2 days) was significantly associated with a lower frequency of class III discrepancies (OR = 0.433 (95% CI 0.197 – 0.948; p = 0.036).

Role of microscopy in identifying cause of death

Table 8 shows that microscopic examination contributed to establishing COD in 19.6% of the cases, it confirmed macroscopical diagnoses in 47.8%, played no role in identifying COD in 16.5%, and 16.1% of the cases were non-classifiable. Microscopic examination most commonly played a role in diagnosing pneumonia (n=28), myocardial infarction (n=11) and lymphocytic or catecholamine-induced myocarditis (n=10) as COD.

Time to completion of the autopsy report

We observed a significant reduction in both the median time to the preliminary and final autopsy report from 11 days in 2007 to 3 days in 2012/13 (p = 0.001), and from 91 days in 2007 to 54 days in 2012/13 (p < 0.001), respectively. Specialized autopsy pathologists had finished their preliminary report in a median of 2 days vs. general pathologists in 7 days (p = 0.003), and their final report in a median of 52 days vs. general pathologists in 85 days (p < 0.001).

Discussion

This study comparing clinical diagnoses and post-mortem diagnoses demonstrates a 23.5% major discrepancy rate and a 32.6% minor discrepancy rate in 337 classifiable clinical autopsy cases. This is in line with recent literature, in which major discrepancy rates ranged from 7% to 50%, mainly depending on patient populations studied [3,10,12-18]. The 23.5% major discrepancy rate is identical to that presented in a review by Shojania *et al* using the results from 42 studies [2].

A reason for the persistently high discrepancy rates may be selection bias, because clinicians are thought to request autopsies mainly for the clinically challenging cases [19]. Nevertheless, several groups have shown that clinicians were not able to predict, based on their clinical certainty, cases that would uncover discrepant autopsy findings [20-22]. Berner *et al* described clinicians' overconfidence in their diagnoses as a contributing cause of diagnostic errors [23]. Moreover, Combes *et al* demonstrated that percentages of major diagnostic discrepancies were similar between patients that had undergone modern diagnostic techniques

and patients that had not, emphasizing the value of the autopsy, even in the era of modern diagnostic techniques [24].

The most commonly observed major discrepancies found in this study were myocardial infarction, pulmonary embolism and pneumonia. This is in agreement with those found by others [25-29], and is comprehensible as myocardial infarction, pulmonary embolism and pneumonia can present atypically or even asymptomatically [30-32]. In addition, Winters *et al* reported aspergillosis, which was the fifth leading major discrepancy in our study, to be a frequently missed class I disease [28].

Surprisingly, we found a higher percentage of major discrepancies when imaging was applied during life. Further analysis revealed that this was mainly due to imaging of an improper body part or with an improper imaging modality thereby failing to identify the actual COD, which was the case in 15.6%.

Similar to previous studies, we demonstrated that microscopic examination has a major impact on macroscopical diagnoses made during clinical autopsies [33-35]. In our study, microscopic examination contributed to the final COD in 19.6% of cases, especially for diagnosing pneumonia, myocardial infarction and myocarditis. In accordance, Hunt *et al* showed a substantial discrepancy rate between macroscopical and microscopically confirmed diagnoses of pneumonia [36].

In these times of fewer monetary resources, quality of care is a critical point. Identification of problematic disease categories can help to reduce the number of unnecessary deaths [37,38]. Autopsies are crucial to determine potential diagnostic errors underlying these high mortality rates, and offer clinicians the opportunity to receive feedback from which lessons can be learned. Furthermore, frequent discrepant diagnoses revealed at autopsy should make health care organizations aware of the incidence of system-related errors, and make them search for interventions on the system-level, such as introducing double readings for certain diagnostic tests and offering clinical decision support opportunities [1].

In previous studies, a longer length of admission at the ICU, of > 2 days and > 10 days, respectively, was significantly associated with more major discrepancies [39,40]. Contrarily, Tavora *et al* found that a shorter length of hospital stay significantly contributed to majors discrepant findings [41].

Although in our study length of final admission did not influence the frequency of major discrepancies, an admission length of > 2 days significantly reduced the frequency of class III minor discrepancies. Longer admission length may influence both mortality and morbidity.

Alternate non-invasive ways of post-mortem examination are being explored. Virtual autopsies by means of CT and MRI have already been used in forensic medicine, and although they seemed promising in clinical medicine, there certainly are drawbacks. In several studies [42-44], a substantial number of diagnoses were missed on virtual autopsy, and the most commonly missed ones were exactly those discrepancies most frequently described in literature as well as in our study.

Due to technical and practical limitations, routine toxicology tests were not included in our clinical autopsy protocols, in line with most other pathology labs. However, routine toxicology testing may reveal otherwise undetected CODs, including death from fatal adverse drug reactions to properly prescribed and administered drugs. These adverse drug reactions have been described to be between the fourth and sixth leading COD in the United States [45]. In future studies we would like to analyze the value of routine toxicology testing.

In The Netherlands, relatives have to give separate permission for body and brain autopsy, leading to a relatively low number of the latter. This is another limitation of this study, since intracranial pathology in cases without brain autopsy cannot be excluded. As a complete autopsy includes the brain, efforts should be made by clinicians to obtain relatives' consent. Furthermore, pathologists should make clinicians more aware of the importance of a complete autopsy.

Regarding the autopsy report, we make several recommendations, based both on literature and our own experiences. The preliminary report should preferably be distributed within 24 hours. It has been proven effective to start with the main findings (COD and major discrepancies), and to describe further findings point by point [46]. Immediate reporting will be most effective as clinicians can directly reflect on their diagnoses [47]. The final report should be distributed within 1 month, since reports received after 1 month are much less useful to clinicians [48]. The timing of feedback is important. Immediate feedback is more effective than delayed feedback [49]. Although in our subset of cases, the median time to completion of the autopsy report was longer than 1 month, we observed a significant reduction in the median time to both the preliminary and

final autopsy report over the study period, mainly ascribed to the deployment of specialized autopsy pathologists who are apparently more dedicated to completing the final reports.

Conclusion: Major discrepancies remain persistent at autopsy, even in the era of high-tech medicine. Therefore, they still serve as a very important part of quality control in clinical diagnosis and treatment. Learning from both individual and system-related diagnostic errors can aid in improving patient safety.

References

1. Graber M. (2009) Diagnostic errors in medicine: a case of neglect. Jt.Comm J Qual Patient Saf 31(2): 106-13.
2. Shojania KG, Burton EC, McDonald KM, Goldman L. (2003) Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. JAMA 289(21): 2849-56.
3. Wittschieber D, Klauschen F, Kimmritz AC, von Winterfeld M, Kamphues C, Scholman HJ, et al. (2012) Who is at risk for diagnostic discrepancies? Comparison of pre- and postmortal diagnoses in 1800 patients of 3 medical decades in East and West Berlin. PLoS One 7(5): e37460.
4. Shojania KG, Burton EC. (2008) The vanishing nonforensic autopsy. N Engl J Med 358(9): 873-5.
5. Chariot P, Witt K, Pautot V, Porcher R, Thomas G, Zafrani ES, et al. (2000) Declining autopsy rate in a French hospital: physician's attitudes to the autopsy and use of autopsy material in research publications. Arch Pathol Lab Med 124(5): 739-45.
6. Loughrey MB, McCluggage WG, Toner PG. (2000) The declining autopsy rate and clinicians' attitudes. Ulster Med J 69(2): 83-9.
7. Klok FA, Mos IC, Kroft LJ, de Roos A, Huisman MV. (2011) Computed tomography pulmonary angiography as a single imaging test to rule out pulmonary embolism. Curr Opin Pulm Med 17(5): 380-6.
8. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. (1983) The value of the autopsy in three medical eras. N Engl J Med 308(17): 1000-5.
9. Battle RM, Pathak D, Humble CG, Key CR, Vanatta PR, Hill RB, et al. (1987) Factors influencing discrepancies between premortem and postmortem diagnoses. JAMA 258(3): 339-44.
10. Schwanda-Burger S, Moch H, Muntwyler J, Salomon F. (2012) Diagnostic errors in the new millennium: a follow-up autopsy study. Mod Pathol 25(6): 777-83.
11. Fronczek J, Hollingbury F, Biggs M, Rutty G. (2014) The role of histology in forensic autopsies. Is a histological examination always necessary to determine a cause of death? Forensic Sci Med Pathol 10(1): 39-43
12. Kotovicz F, Mauad T, Saldiva PH. (2008) Clinico-pathological discrepancies in a general university hospital in Sao Paulo, Brazil. Clinics (Sao Paulo) 63(5): 581-8.
13. Scordi-Bello IA, Kalb TH, Lento PA. (2010) Clinical setting and extent of premortem evaluation do not predict autopsy discrepancy rates. Mod Pathol 23(9): 1225-30.
14. Burgesser MV, Camps D, Calafat P, Diller A. (2011) Discrepancies between clinical diagnoses and autopsy findings. Medicina (B Aires) 71(2): 135-8.
15. Fares AF, Cury PM, Lobo SM. (2011) Clinical-pathological discrepancies in critically ill patients with difficult premortem diagnoses. Rev Bras Ter Intensiva 23(4): 442-7.
16. Podbregar M, Kralj E, Čičak R, Pavlinjek A. (2011) A triad algorithm for analysing individual ante- and post-mortem findings to improve the quality of intensive care. Anaesth Intensive Care 39(6): 1086-92.
17. Tejerina E, Esteban A, Fernández-Segoviano P, María Rodríguez-Barbero J, Gordo F, Frutos-Vivar F, et al. (2012) Clinical diagnoses and autopsy findings: discrepancies in critically ill patients*. Crit Care Med 40(3): 842-6.
18. Frohlich S, Ryan O, Murphy N, McCauley N, Crotty T, Ryan D. (2014) Are Autopsy Findings Still Relevant to the Management of Critically Ill Patients in the Modern Era? Crit Care Med 42(2): 336-43.

19. De Vlieger GY, Mahieu EM, Meersseman W. (2010) Clinical review: What is the role for autopsy in the ICU? Crit Care 14(2): 221.
20. Cameron HM, McGoogan E, Watson H. (1980) Necropsy: a yardstick for clinical diagnoses. Br Med J 281(6246): 985-8.
21. Landefeld CS, Chren MM, Myers A, Geller R, Robbins S, Goldman L. (1988) Diagnostic yield of the autopsy in a university hospital and a community hospital. N Engl J Med 318(19): 1249-54.
22. Podbregar M, Voga G, Krivec B, Skale R, Pareznik R, Gabrscek L. (2001) Should we confirm our clinical diagnostic certainty by autopsies? Intensive Care Med 27(11): 1750-5.
23. Berner ES, Graber ML. (2008) Overconfidence as a cause of diagnostic error in medicine. Am J Med 121(5 Suppl): S2-23.
24. Combes A, Mokhtari M, Couvelard A, Trouillet JL, Baudot J, Hénin D, et al. (2004) Clinical and autopsy diagnoses in the intensive care unit: a prospective study. Arch Intern Med 164(4): 389-92.
25. Perkins GD, McAuley DF, Davies S, Gao F. (2003) Discrepancies between clinical and postmortem diagnoses in critically ill patients: an observational study. Crit Care 7(6): R129-R132.
26. Vadillo M, Pujol RP, Corbella X, Gorriz T, Rabasa P, Bernat R. (2006) Improvements in clinical diagnostic accuracy after a 5-year systematic analysis of clinical and autopsy discrepancies. Arch Pathol Lab Med 130(9): 1261-2.
27. Thurnheer R, Hoess C, Doenecke C, Moll C, Muntwyler J, Krause M. (2009) Diagnostic performance in a primary referral hospital assessed by autopsy: evolution over a ten-year period. Eur J Intern Med 20(8): 784-7.
28. Winters B, Custer J, Galvagno SM Jr, Colantuoni E, Kapoor SG, Lee H, et al. (2012) Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. BMJ Qual Saf 21(11): 894-902.
29. Gibson TN, Shirley SE, Escoffery CT, Reid M. (2004) Discrepancies between clinical and postmortem diagnoses in Jamaica: a study from the University Hospital of the West Indies. J Clin Pathol 57(9): 980-5.
30. Zarling EJ, Sexton H, Milnor P Jr. (1983) Failure to diagnose acute myocardial infarction. The clinicopathologic experience at a large community hospital. JAMA 250(9): 1177-81.
31. Kannel WB, Abbott RD. (1984) Incidence and prognosis of unrecognized myocardial infarction. An update on the Framingham study. N Engl J Med 311(18): 1144-7.
32. Ryu JH, Olson EJ, Pellikka PA. (1998) Clinical recognition of pulmonary embolism: problem of unrecognized and asymptomatic cases. Mayo Clin Proc 73(9): 873-9.
33. Roulson J, Benbow EW, Hasleton PS. (2005) Discrepancies between clinical and autopsy diagnosis and the value of post mortem histology; a meta-analysis and review. Histopathology 47(6): 551-9.
34. Zaitoun AM, Fernandez C. (1998) The value of histological examination in the audit of hospital autopsies: a quantitative approach. Pathology 30(2): 100-4.
35. Bernardi FD, Saldiva PH, Mauad T. (2005) Histological examination has a major impact on macroscopic necropsy diagnoses. J Clin Pathol 58(12): 1261-4.
36. Hunt CR, Benbow EW, Knox WF, McMahon RF, McWilliam LJ. (1995) Can histopathologists diagnose bronchopneumonia? J Clin Pathol 48(2): 120-3.
37. Jarman B, Bottle A, Aylin P, Browne M. (2005) Monitoring changes in hospital standardised mortality ratios. BMJ 330(7487): 329.

38. Jarman B, Pieter D, van der Veen AA, Kool RB, Aylin P, Bottle A, et al. (2010) The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? Qual Saf Health Care 19(1): 9-13.
39. Mort TC, Yeston NS. (1999) The relationship of pre mortem diagnoses and post mortem findings in a surgical intensive care unit. Crit Care Med 27(2): 299-303.
40. Maris C, Martin B, Creteur J, Remmelink M, Piagnerelli M, Salmon I, et al. (2007) Comparison of clinical and post-mortem findings in intensive care unit patients. Virchows Arch 450(3): 329-33.
41. Tavora F, Crowder CD, Sun CC, Burke AP. (2008) Discrepancies between clinical and autopsy diagnoses: a comparison of university, community, and private autopsy practices. Am J Clin Pathol 129(1): 102-9.
42. Wichmann D, Obbelode F, Vogel H, Hoepker WW, Nierhaus A, Braune S, et al. (2012) Virtual autopsy as an alternative to traditional medical autopsy in the intensive care unit: a prospective cohort study. Ann Intern Med 156(2): 123-30.
43. Roberts IS, Benamore RE, Benbow EW, Lee SH, Harris JN, Jackson A, et al. (2012) Post-mortem imaging as an alternative to autopsy in the diagnosis of adult deaths: a validation study. Lancet 379(9811): 136-42.
44. Weustink AC, Hunink MG, van Dijke CF, Renken NS, Krestin GP, Oosterhuis JW. (2009) Minimally invasive autopsy: an alternative to conventional autopsy? Radiology 250(3): 897-904.
45. Lazarou J, Pomeranz BH, Corey PN. (1998) Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 279(15): 1200-5.
46. Valenstein PN. (2008) Formatting pathology reports: applying four design principles to improve communication and patient safety. Arch Pathol Lab Med 132(1): 84-94.
47. van den Tweel JG. (2008) Autopsy pathology should become a recognised subspecialty. Virchows Arch 452(5): 585-7.
48. Waters BL. (2009) Ensuring quality in the hospital autopsy. In: Waters BL, ed. Handbook of autopsy practice. Humana Press: 3-8
49. Siebert JR. (2009) Increasing the efficiency of autopsy reporting. Arch Pathol Lab Med 133(12): 1932-7.

Addendum

Obducties

Hoofdstuk 30.4 in “handboek patiëntveiligheid”, onder redactie van Rob Dillmann e.a., Tijdstroom Uitgeverij 2016

Chantal C.H.J. Kuijpers, J. Fronczek, F.R.W. van de Goot, N.M. Jiwa

Een belangrijke tool voor kwaliteitscontrole en verbetering van de medische zorg is de obductie, ook wel sectie of autopsie genoemd. Door na overlijden te kijken welke ziektes, aandoeningen of afwijkingen bij de patiënt aanwezig zijn en dit te vergelijken met bij leven gestelde diagnosen of verwachtingen kan het medisch handelen direct worden geëvalueerd. De obductie wordt dan ook gezien als de gouden standaard voor evaluatie van medisch handelen. Terugkoppeling van bij obductie gestelde diagnosen helpt clinici in het vervolg bij het maken van beslissingen in soortgelijke gevallen, hetgeen de patiëntveiligheid en kwaliteit van de zorg ten goede komt. In dit hoofdstuk komt de procedure van de obductie aan bod, het belang van obducties, de zorgwekkende dalende trend in het aantal uitgevoerde obducties en tenslotte een aantal aanbevelingen om de neerwaartse trend te keren. Daarnaast wordt een tweetal illustratieve voorbeeldcasus besproken die het belang van obductie benadrukken.

De procedure

Een goed uitgevoerde obductie zal volgens een vast stramien worden verricht. Als eerste zal er een aanmelding zijn. Bij een klinische obductie of een huisartsen inbreng zal de patholoog zo veel mogelijk informatie willen hebben. Het is dan van groot belang om een aanvraag bij te voegen waarin de medische voorgeschiedenis van de patiënt beschreven staat, daar tegenwoordig door gestage progressie van wetenschappelijk inzicht steeds meer syndromen bekend worden. Het min of meer gelijktijdig optreden van colon carcinoom en glioblastoom is natuurlijk illustratief voor een dergelijke bewering. Na het inlezen zal de patholoog overgaan tot het uitwendig inspecteren van het lichaam. Dit natuurlijk omdat, net als bij een lichamelijk onderzoek bij levenden, ook bij de doden aan de buitenzijde een schat aan informatie te zien is. Intrekkingen van borsten bij verborgen mammacarcinoom, melanoom op de rug, horlogeglasnagels bij aanhoudende hypoxie dan wel faciale hyperplasie bij Pierre Marie Bamberg.

Na de uitwendige schouw zal de patholoog het lichaam openen door middel van de klassieke Y snede, lopende van de beide schouders naar centraal op de

borst en vandaaruit naar de onderbuik. Op geleide van bijzondere vragen kan van deze snede worden afgeweken. Het is vanzelfsprekend dat bijvoorbeeld bij tumoren van de ledematen ook daar onderzoek dient plaats te vinden.

Na de eerste opening zal de patholoog alle relevante organen uit het lichaam nemen om op een aangrenzende tafel deze verder te onderzoeken. Van alle relevante organen worden kleine stukjes bewaard op formaline. Het komt zeer sporadisch voor dat een heel orgaan wordt bewaard. Dit is vrijwel alleen het geval als er een expliciete vraagstelling is waarbij aanvullende academische consultatie noodzakelijk wordt geacht. Indien een orgaan wordt bewaard zal dit altijd met bekendmaking en toestemming van de nabestaanden zijn. Na de lichaamsobductie wordt dan indien gewenst een schedellichting uitgevoerd. Hierbij wordt de huid van het achterhoofd losgemaakt op een dermate wijze dat na reconstructie deze handeling tijdens aansluitende opbaring onzichtbaar blijft.

Na het openen van de schedel zullen dan de hersenen onderzocht worden. Ofschoon het niet in ieder ziekenhuis protocol is, zullen de hersenen na onderzoek aan de patiënt worden geretourneerd tenzij er een relevante neuropathologische vraagstelling is. Te denken is dan aan zeldzame neurodegeneratieve afwijkingen. Na afsluiten van het onderzoek worden alle organen aan de patiënt geretourneerd waarna de obductie assistent het lichaam kan sluiten. Het sluiten gebeurt door het afhechten van de voorste snede en het nadien afplakken zodat een patiënt netjes voor opbaring kan worden afgelegd. Na het voltooien van de macro-sectie volgt dan voor de patholoog nog het beoordelen van het bewaarde materiaal (de kleine stukjes). Dit wordt verwerkt tot coupes voor onder de microscoop. Daar zullen ziektebeelden verder worden onderzocht en benoemd. Het materiaal voor microscopie blijft bij het laboratorium voor pathologie vele jaren bewaard en kan ook jaren na datum nog steeds worden gebruikt om bijvoorbeeld genetisch onderzoek te doen mocht later in de familie van een patiënt onverwacht een ziektebeeld de kop op steken.

Het belang van obducties

Wanneer men aan een obductie denkt wordt vaak het eerste gedacht aan het achterhalen van de oorzaak van overlijden. Wanneer een patiënt plotseling overlijdt en niet duidelijk is waaraan deze is overleden, is postmortaal onderzoek uiteraard zeer nuttig.

Echter, vaak wordt ten onrechte gedacht dat het achterhalen van de doodsoorzaak het enige doel van een obductie is, en dat een obductie dan ook alleen nuttig zou zijn wanneer de oorzaak van overlijden niet bekend of onduidelijk is. Een obductie is echter nuttig bij alle overleden patiënten. Uit wetenschappelijk onderzoek blijkt dat bij postmortaal onderzoek bij een groot aantal patiënten (~25%) belangrijke onverwachte bevindingen worden gedaan, ook bij patiënten bij wie de doodsoorzaak bekend was of bekend leek [1-3]. Tien tot 15% van de bij obductie gevonden diagnoses zouden, wanneer zij bij leven zouden zijn gesteld, invloed hebben gehad op behandeling en mogelijk zelfs op overleving. Hartinfarct, longembolie en pneumonie blijken vaak gemiste diagnosen te zijn. Veel ziektebeelden kunnen atypisch zijn, bijvoorbeeld van hartinfarcten is bekend dat slechts een klein deel de klassieke symptomen geeft. Ook infectieziekten kunnen met name bij ouderen een aspecifiek beloop hebben, waardoor deze pas laat of helemaal niet worden opgemerkt. Daarnaast wordt mogelijk teveel waarde toegekend aan ja/nee-uitslagen van aanvullende onderzoeken, zoals een al dan niet verhoogde waarde van een bepaald eiwit bij bloedonderzoek.

Ook bij obducties op neonaten en kinderen worden vaak onverwachte bevindingen gedaan. Een recente systematische literatuur review liet zien dat dit in ~20% van de secties op neonaten of kinderen die overleden op een intensive care afdeling het geval was [4].

Deze bevindingen zijn van belang voor evaluatie van medisch handelen. Door tijdige terugkoppeling aan de behandelend artsen kunnen zij kritisch naar de door hun gestelde diagnosen en gegeven behandeling kijken. Dit kan hen in het vervolg helpen bij het nemen van beslissingen bij soortgelijke casus, hetgeen de patiëntveiligheid en de kwaliteit van de zorg ten goede komt. Daarnaast kunnen er ook bevindingen worden gedaan die van belang zijn voor de nabestaanden, zoals tekenen van een besmettelijke ziekte of een erfelijke aandoening. Met name dit laatste wordt onderschat. Uit diverse studies komt het belang naar voren voor het doen van gericht onderzoek bij nabestaanden na het acuut overlijden van relatief jonge personen [5]. Substantiële hartafwijkingen en dan met name het soort afwijkingen waarvan bekend is dat dit ritmestoornissen kan veroorzaken blijken vaker voor te komen dan tot nu toe in de literatuur gemeld wordt. Het is hierbij tevens van belang om aan te geven dat bij een gewone obductie dit soort afwijkingen niet worden opgemerkt daar het hier om genetisch bepaalde pathofysiologie gaat en niet zozeer om macropathologie of histopathologie.

Het leidt geen twijfel dat het van groot belang is om dit soort afwijkingen in kaart te brengen bij verkeersongevallen. Momenteel wordt slechts een minimaal percentage van alle verkeersslachtoffers onderzocht terwijl met name ook bij verzekeringskwesties obducties doorslaggevend kunnen zijn voor het bepalen van een toedracht. Daarnaast zijn ook de epidemiologische, wetenschappelijke en educatieve belangen van de obductie niet onbelangrijk. Zo zal een toename in het aantal obducties leiden tot meer betrouwbare statistieken. Tenslotte kan het nabestaanden helpen bij het verwerkingsproces wanneer er postmortaal onderzoek plaatsvindt.

Daling in het aantal obducties

Ondanks het belang van postmortaal onderzoek, is het percentage obducties al jaren dalende. In Nederland sloegen pathologen in 2014 de noodklok over het dalende percentage obducties, dat met maar liefst 60% daalde in 20 jaar [6]. Maar ook internationaal is een neerwaartse trend te zien in het aantal secties [7]. Voor deze daling zijn een aantal redenen te noemen:

1. De houding van artsen ten opzichte van de obductie.
   a. Door het toenemende vertrouwen in moderne beeldvormende diagnostische technieken wordt de obductie in toenemende mate overbodig geacht. Artsen denken dat zij door middel van deze technieken in staat zijn alles te diagnosticeren. In principe is het natuurlijk ook mogelijk om (bijna) alles hiermee te diagnosticeren, maar uit onderzoek [3] blijkt dat wanneer bij obductie iets gevonden werd dat bij leven gemist bleek te zijn, dit vaak kwam doordat ofwel voor het verkeerde lichaamsdeel (bijvoorbeeld abdomen in plaats van thorax) beeldvorming gebruikt was of doordat een verkeerde/suboptimale beeldvormende techniek gebruikt was voor het stellen van die bepaalde diagnose. Er wordt bij obductie bijvoorbeeld een colon carcinoom gevonden, maar bij leven is er alleen een röntgenfoto van de thorax gemaakt. Of obductie laat een longembolie zien die bij leven gemist was, ondanks beeldvorming van de thorax. Echter, er was een röntgenfoto van de thorax gemaakt, terwijl een longembolie gediagnosticeerd dient te worden met een CT-angiografie. Onderzoek heeft ook aangetoond dat de percentages onverwachte bevindingen bij obductie gelijk waren bij patiënten met en zonder beeldvorming bij leven [8].
   b. Daarnaast zijn artsen vaak terughoudend in het vragen van toestemming voor obductie aan de nabestaanden. Ze vinden het lastig om over de mogelijkheid van obductie te beginnen wanneer mensen zojuist hun geliefde hebben verloren. Ook heerst er de angst onder artsen voor een eventuele tuchtzaak wanneer blijkt dat zij iets "fout" gedaan hebben of

gemist hebben. Een open patiëntveiligheidscultuur waarbij het leren van "fouten" in plaats van "fouten" af te straffen centraal staat is dus essentieel.

c. Tenslotte vragen artsen minder obducties aan doordat het obductierapport vaak lang op zich laat wachten en de communicatie met de patholoog niet altijd optimaal is [9].

2. Er wordt minder toestemming gegeven door nabestaanden voor het verrichten van een sectie.
   a. Dit kan deels verklaard worden door emotionele, culturele of religieuze bezwaren. Nabestaanden vinden vaak dat hun geliefde al genoeg geleden heeft en nu rust verdient. Daarnaast strookt de obductie met culturen of religies waarbij de overledene zo snel mogelijk begraven dient te worden.
   b. Ook is er in de media een en ander naar voren gekomen over het achterhouden van organen na obductie, waardoor nabestaanden mogelijk afgeschrikt worden. In sommige gevallen is het nodig om (delen van) organen te bewaren voor verder onderzoek. Het is een geaccepteerd protocol dat hersenen 6 weken fixeren voordat ze onderzocht kunnen worden. Voor veel nabestaanden is het niet direct retourneren van de hersenen een wezenlijke belemmering voor het toestaan van obductie. Met name dit laatste aspect zou feitelijk geen rol hoeven te spelen daar het zondermeer mogelijk is in een groot deel van de gevallen om ook op vers materiaal neuropathologisch onderzoek te doen. Slechts in uitzonderlijke gevallen kan het zijn dat de gemelde fixatie nodig is. Het is in die gevallen van belang dat artsen dit voorafgaand aan de obductie met de familieleden bespreken.
3. Ook onder de pathologen is de obductiepathologie niet geliefd. Slechts een kleine minderheid van de pathologen heeft affiniteit met het doen van obducties. Met name in (perifere) pathologie afdelingen met een hoge werkdruk wordt de obductie als een vervelende bijkomstigheid ervaren. Dit heeft uiteraard invloed op de kwaliteit van de obducties en het aantal obductie aanvragen. Vanuit de Nederlandse Vereniging voor Pathologie echter wordt het belang van obductie zondermeer onderstreept.
4. Tenslotte wordt een obductie niet vergoed door de zorgverzekeraars, waardoor de kosten volledig voor het ziekenhuis zijn. Door de toenemende bezuinigingen in de zorg komt de obductie onder druk te staan.

Aanbevelingen

De trend in de daling van obducties heeft een negatieve invloed op de kwaliteit van de zorg en de patiëntveiligheid. De neerwaartse trend zal dan ook moeten worden gekeerd, zodat de gezondheidszorg weer optimaal kan profiteren

van dit belangrijke kwaliteitsinstrument. Hiervoor zijn een aantal zaken van belang.

Kliniek / behandelend artsen

Door onderwijs en voorlichting moeten artsen en medisch studenten meer bewust worden van het belang van obducties. Wanneer (toekomstige) artsen meer inzicht hebben in het nut van obducties zullen zij eerder geneigd zijn een obductie aan te vragen. Wat hiermee verband houdt is de wijze waarop door een arts toestemming aan nabestaanden wordt gevraagd voor het doen van een obductie. Ook hierin dienen artsen opgeleid te worden.

Pathologie

De kwaliteit van het obductieverslag moet beter en de resultaten moeten sneller beschikbaar zijn voor de kliniek. Directe feedback is namelijk effectiever dan vertraagde feedback. Idealiter worden obducties uitgevoerd door toegewijde obductiepathologen. Ervaring leert ons dat obductieverslagen door toegewijde obductiepathologen vaak completer zijn dan die gemaakt door generalisten. Daarnaast zagen wij dat de tijd tot het afronden van het rapport significant korter was bij toegewijde obductiepathologen dan bij algemene pathologen [3]. De Nederlandse Vereniging voor Pathologie die verantwoordelijk is voor het opleiden van pathologen zal in de opleiding meer aandacht moeten besteden aan het obductieonderwijs.

Ziekenhuis

Elk ziekenhuis zou een goede informatiefolder voor nabestaanden moeten hebben over wat een obductie inhoudt. Hierin zou beschreven moeten staan hoe er te werk wordt gegaan en wat er met de organen gebeurt, zodat er geen misverstanden ontstaan. Ziekenhuizen zouden obducties moeten stimuleren in een veilige cultuur waarbij artsen niet bang hoeven te zijn "gestraft" te worden voor "fouten", om zo de kwaliteit van de zorg te verbeteren. Ook necrologiebesprekingen waarbij het klinische beloop en de bevindingen bij obductie worden besproken, moeten worden gestimuleerd vanuit de organisatie. Dit is uitermate belangrijk voor de behandelend artsen van de te bespreken patiënten, daar zij directe terugkoppeling krijgen van de patholoog en deze zo nodig ook direct vragen kunnen stellen Daarnaast zijn necrologiebesprekingen ook erg leerzaam voor overige artsen, assistenten, coassistenten en verplegend personeel.

Een initiatief uit het verenigd Koninkrijk is wellicht als overweging aan te dragen. Hierbij werd een Nurse Practitioner aangesteld die specifiek belast is met het vragen van toestemming aan nabestaanden en het coördineren van transport, afhandeling en uiteindelijk de multidisciplinaire besprekingen.

Zorgverzekeraars

Op dit moment wordt postmortaal onderzoek niet vergoed door zorgverzekeraars, omdat de patiënt al overleden is en dus zelf geen baat meer heeft bij de obductie. Echter, het belang voor de samenleving moet worden meegewogen in de beslissing tot het vergoeden van obducties. Betere zorg zal in de toekomst de kosten voor de zorgverzekeraars omlaag brengen. Uit een Amerikaanse studie is reeds naar voren gekomen dat er een wezenlijke groep patiënten is die komt te overlijden door of mede door medicijngebruik. Hierbij gaat het zowel om onjuiste medicatie maar opmerkelijk genoeg ook om juist geïndiceerde en juist gedoceerde medicatie.

Het belang van obductie is zondermeer te onderstrepen bij de huidige verandering in de zorg waarbij schaalvergroting tot kwaliteitsverbetering moet leiden. Het is van essentieel belang dat vanuit dat oogpunt ook daadwerkelijk gecontroleerd wordt of complicaties bij en overlijden van patiënten daadwerkelijk iets te maken hebben met de gecentraliseerde behandeling. Het vergelijken van sterftecijfers van ziekenhuizen onderling als maatstaf voor kwaliteit met een dermate laag percentage obducties zoals op dit moment in Nederland voorhanden is volstrekt zinloos.

Geneeskundeopleiding

De grootste winst valt te behalen helemaal aan het begin van de keten, bij het opleiden van nieuwe artsen. Op dit moment wordt er tijdens de opleiding geneeskunde in Nederland te weinig aandacht besteed aan zowel pathologie onderwijs in het algemeen als onderwijs in obductiepathologie. Artsen in spe moeten echter al tijdens de opleiding het belang van de obductie meekrijgen. Dit zal hun denken over en handelen ten opzichte van obducties positief beïnvloeden.

Het is in dit kader van wezenlijk belang dat de exacte vakken van de geneeskunde zoals de fysiologie, anatomie en pathologie herkenbaar worden aangeboden aan de nieuwe generatie dokteren. Er steekt nog steeds veel wijsheid in de opmerking: De doden leren de levenden.

Voorbeeldcasussen

Casus 1

Klinisch verloop

Een vrouw van 42 jaren oud met ernstig overgewicht (156 kg bij 170 cm) geeft aan dat ze forse buikpijn heeft. Een urinetest geeft een indicatie voor een blaasontsteking waarna antibiotische behandeling wordt ingezet. Het beeld lijkt hier iets op te verbeteren echter de buikpijn blijft aanhouden. Tevens wordt er obstipatie gemeld. In eerste instantie wordt een laxeermiddel gegeven echter als na een week de obstipatie en de buikpijn blijft aanhouden wordt een buikoverzichtsfoto gemaakt. Alhier wordt duidelijke coprostase (= ophoping van ontlasting) gezien echter door de BMI is het beeld moeilijk te beoordelen. Op zondagochtend wordt de vrouw in slechte conditie aangetroffen. Ze is amper aanspreekbaar. Er volgt een spoedopname waar overtuigende ontstekingsparameters worden aangetoond. De urine is nu schoon en een primair focus laat zich niet zondermeer aanwijzen. De vrouw blijft septisch, begint te decompenseren en heeft ondersteunende medicatie voor het hart nodig. Op de dag van opname ontstaat een reanimatiesetting welke vruchteloos zal verlopen.

Er wordt klinische obductie gevraagd om na te gaan waar het septisch focus heeft gezeten.

Obductie

Bij sectie wordt een groot abces gevonden, links in de buikwand. Er blijkt een doorgebroken divertikel in het sigmoid te zijn waarbij door verklevingen de darminhoud door het peritoneum in het vetweefsel terecht is gekomen. Het abces breidt zich uit richting de heup, richting de wervelkolom en richting de linker ribbenboog. De totale afmeting is dan circa 20 cm in diameter.
Terugkijkende op de buik overzichtsfoto is de diverticulose te zien, zelfs de aanliggende verkleefde divertikel is te zien en met wat voorkennis lijkt ook de perforatie al aanwezig.

Casus 2

Klinisch verloop

Een man van 47 jaren oud presenteert zich op een eerste hulp na acuut onwel worden met trekkingen. Er bestaat een verdenking op een primaire hersenafwijking. CT en MRI geven rechts fronto-temporaal (voor in de grote hersenen) een grillig proces waar de radioloog een sterke verdenking

op een glioblastoom durft af te geven. Een glioblastoom is een hooggradige hersentumor met een slechte prognose. Momenteel worden gecombineerde radio-/chemokuren gegeven teneinde tijd te winnen. Er wordt een biopt genomen hetgeen de diagnose bevestigt. Het proces groeit echter gestaag en de functionele toestand van de man daalt sterk. Spoedoverleg en hernieuwde scans geven aan dat het proces snel uitbreidt onder ingezette behandeling en de man komt nog tijdens de behandelingen in een reanimatiesetting terecht waarbij in overleg met de familie geen handelingen meer worden verricht en hij komt te overlijden. De familie wil graag obductie op verzoek van de neurologen, omdat deze tumor zo snel en agressief groeide.

Obductie

Bij sectie wordt een zeer groot necrotiserend proces rechts in de hersenen gezien. Op dit moment steekt het proces via de hersenbalk reeds over naar links. Er is nu sprake van een grote bloeding rond de tumor met tekenen van inklemming. Het microscopisch beeld is conform de biopten en het radiologisch beeld. Opvallend alleen is de aanwezigheid van vele colonpoliepen en in het sigmoid een laesie die naar zou blijken reeds kanker (coloncarcinoom) is. De combinatie glioblastoom en coloncarcinoom is reden genoeg voor genetisch onderzoek waar zondermeer de diagnose Lynch syndroom naar voren komt. Helaas met verstrekkende gevolgen voor de familie van deze man.

Referenties

1. Shojania, K.G., et al., Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. JAMA, 2003. 289(21): p. 2849-56.
2. Wittschieber, D., et al., Who is at risk for diagnostic discrepancies? Comparison of pre- and postmortal diagnoses in 1800 patients of 3 medical decades in East and West Berlin. PLoS One, 2012. 7(5): p. e37460.
3. Kuijpers, C.C., et al., The value of autopsies in the era of high-tech medicine: discrepant findings persist. J Clin Pathol, 2014. 67(6): p. 512-9.
4. Custer, J.W., et al., Diagnostic errors in the pediatric and neonatal ICU: a systematic review. Pediatr Crit Care Med, 2015. 16(1): p. 29-36.
5. Hendrix, A., et al., Cardiogenetic screening of first-degree relatives after sudden cardiac death in the young: a population-based approach. Europace, 2011. 13(5): p. 716-22.
6. Persbericht NVVP en PALGA. 2014; http://www.palga.nl/assets/uploads/Publicaties/Persbericht%20NVVP_steeds%20minder%20obducties%20in%20nederland(12%20april2014).pdf
7. Burton, J.L. and J. Underwood, Clinical, educational, and epidemiological value of autopsy. Lancet, 2007. 369(9571): p. 1471-80.
8. Combes, A., et al., Clinical and autopsy diagnoses in the intensive care unit: a prospective study. Arch Intern Med, 2004. 164(4): p. 389-92.
9. Hagestuen, P.O. and S. Aase, The organisation and value of autopsies. Tidsskr Nor Laegeforen, 2012. 132(2): p. 152-4.

General discussion and future perspectives

The pathology report often contains prognostic information of the disease at hand and is often the basis for treatment decisions, especially in oncology. An adequate diagnosis and pathology report are therefore of paramount importance, and should be accurate, timely, and complete [1]. Not only should the main diagnosis be correct, also critical features that are related to diagnosis, prognosis and treatment selection, such as tumor size, histological subtype, grade and stage, must be determined and reported accurately. The pathology report is the end product of all steps performed within the diagnostic process of the pathology department, from arrival of the material, gross examination, microscopic examination, additional tests (e.g. immunohistochemistry (IHC) and molecular assays) up until the final sign out of the report. Unfortunately, it is inevitable that in all steps of this process errors may and do ultimately arise. It is the task of all involved (e.g. technicians, secretaries, residents, pathologists, and directors) to minimize the error rate.

In this thesis, we aimed to assess and improve the quality of some aspects of pathology practice, and thereby improve quality of health care and patient safety. To this end, several diagnostic processes in pathology practice were assessed, and the added value of multiple interventions or strategies, with a focus on oncology, was investigated. We addressed several problems that may lead to an inaccurate diagnosis and pathology report. Firstly, we addressed the problem of high pathologist workload. The second very common problem we addressed was interobserver variation. In addition, we determined the role of autopsies in quality improvement of health care. This chapter summarizes the main findings of this thesis, and discusses future perspectives.

Main findings of this thesis

The results of this thesis clearly show that quality of pathology practice and patient care can be considerably improved by several relatively simple quality assurance interventions:

1. Task redistribution by the employment of pathologists' assistants (PAs) to take over a selection of routine tasks from pathologists and by subspecialization of pathologists.
2. Incorporation of redundancy by double reading of a selection of histopathology and cytopathology specimens (by specialized pathologists) and by routinely co-testing all invasive breast cancers for HER-2 status by IHC and a gene amplification test (e.g. multiplex ligation-dependent probe amplification (MLPA)).

3. Feedback on performance to pathology laboratories, individual pathologists and clinicians by comparing laboratory performance on a nationwide basis, by feedback on discordant double reading diagnoses, and by feedback of autopsy findings to clinicians.

Discussion and future perspectives

Reducing pathologist workload

The first problem we addressed was high pathologist workload, which is expected to compromise quality of pathology [2]. In the upcoming decades, pathologist workload is expected to increase even more, due to an aging population (increase in malignant and chronic diseases), the increasing requirement of in-depth knowledge of new diagnostic methods for personalized medicine [3,4], and by the introduction of new screening programs.

Task redistribution

Pathologist workload can be decreased by the employment of PAs to take over certain routine activities, such as dissecting relatively simple specimens (e.g. gall bladder, appendix, and skin biopsies) and oncological resection specimens (e.g. breast, colon), and harvesting lymph nodes (LNs), from pathologists. Prior to the introduction of the PA training program in The Netherlands in 2008, there was a discussion whether deployment of PAs, who lack the 5-year training in pathology, would hamper the quality of patient care or it would lead to quality improvement. We investigated this question in one specific aspect, namely the harvest of LNs from colorectal cancer resection specimens (**Chapter 2**).

We showed that PAs actually contributed to quality improvement, at least in this specific aspect, as they gathered, on average, significantly more LNs and in a significantly higher proportion of resection specimens an adequate number of LNs (≥ 10 according to the Dutch guideline [5]) than pathologists did. This resulted in more accurate tumor staging and better treatment selection, as the number of colon cancer patients eligible for adjuvant chemotherapy due to inadequate LN sampling alone decreased when PAs harvested LNs.

Further studies are needed to assess the added value of PAs in other aspects of gross examination, such as tissue resubmission rates. Based on the experience in their laboratory, Goldstein *et al* [6] reported that a very small number of tissue blocks resected by PAs needed recutting and laboratory reprocessing due to overly thick blocks. Galvis *et al* [7] reported a lower number of cases

requiring extra tissue blocks for microscopic examination when PAs instead of pathology residents performed the gross examination.

Currently, in The Netherlands, 83 PAs are employed and another twelve technicians are in training for PA. The Dutch Society of Pathology (NVVP) recognizes that PAs can very well take over specific tasks from pathologists, provided that the following quality standards are met: 1) a proper education, 2) clear supervision with regular evaluation, 3) clear agreement on tasks and responsibilities, and 4) the presence of up-to-date protocols [8].

As workload of pathologists further increases with the introduction of screening programs, PAs may, in addition to certain routine gross examination tasks, take over some microscopy tasks from pathologists as well, including the microscopic pre-screening of colorectal polyps received as part of the national bowel cancer screening program, which is already performed by PAs in certain laboratories, or skin tumors, and preparing a report for these tumors that the pathologist can sign out after quality control [9]. The quality of microscopic prescreening performed by PAs is constantly assessed by reviewing all cases together with a pathologist, a setting comparable with that of pathology residents.

Subspecialization of pathologists is another way of task redistribution to improve quality of care [10] and decrease workload, due to increased expertise and subsequent reduced work complexity. In **Chapter 6**, we showed improved quality of cytology diagnostics by double reading of a selection of cytology specimens by a team of specialized cytopathologists. In addition, in **Chapter 8**, we reported a significant reduction in time to autopsy report completion by specialized autopsy pathologists. Also, the autopsy reports of autopsy pathologists were more comprehensive than that of general pathologists. These results emphasize the need for subspecialization in pathology practice.

Reducing and overcoming interobserver variation

The other problem we addressed was interobserver variation. Variation in the diagnostic process, and thus in pathology practice is undesirable, and standardization is needed in case variation is present. The first step towards standardization is providing insight into the presence of this variation.

In **Chapters 3 and 4**, we provided insight into the variation in daily practice between Dutch pathology laboratories and individual pathologists with regard

to the grading of colorectal pre-malignancies (adenomas) and malignancies (adenocarcinomas). Using data from PALGA (the Dutch Pathology Registry), considerable variation in grading of colorectal adenomas (n = 32,391) and adenocarcinomas (n = 11,719) was observed, with a few laboratories diagnosing significantly aberrant proportions of adenomas with high grade dysplasia and/or poorly differentiated adenocarcinomas, even after adjusting for differences in case mix.

The results presented in these chapters indeed illustrate that variation is present and that better standardization of histological grading is needed, e.g. by education. For example, with the aim of standardizing bowel cancer screening diagnostics, an e-learning program with concluding examination was developed by and for pathologists commissioned by the *Rijksinstituut voor Volksgezondheid en Milieu* (RIVM). Since October 2014, solely pathologists who passed the exam are allowed to perform bowel cancer screening diagnostics. The 2013 data on the grading of dysplasia in adenomas (**Chapter 3**) were compared with data on "non-screening" adenomas from October 2014 up to June 2015. In the latter period, decreased interlaboratory variation for the adenomas diagnosed on polypectomies, but increased variation for the adenomas diagnosed on biopsies was observed (SKMS research report '*Vermindering van de praktijkvariatie: Analyse van een mogelijke strategie bij colorectale adenomen*') [11]. Further analyses will need to determine the role of the e-learning program.

Future research needs to establish whether variation is also present in other aspects of the diagnostic process, such as the histological grading of other neoplasms (e.g. breast cancer), or other critical parameters (e.g. histological subtype). We presume that these will reveal comparable interlaboratory and intralaboratory variation.

Besides education of individual pathologists, other possible interventions to improve standardization are incorporating redundancy in the diagnostic process and giving feedback on performance. These interventions are explained in detail below.

Incorporation of redundancy

In this thesis we studied two manners of redundancy, namely by double reading of a selection of specimens (**Chapters 5 and 6**), and by co-testing for HER-2 status in invasive breast cancer patients with two techniques (**Chapter 7**).

Double reading

Chapter 5 concerned intradepartmental routine second review of histopathology specimens prior to discussion at a multidisciplinary meeting. In about 1% of the cases second review resulted in a major discordant diagnosis with potential clinical significance. Not only malignant, but also benign specimens encountered major discordances, indicating that double reading may also be useful for selected specimens with a benign diagnosis.

In **Chapter 6**, an even higher major discordance rate (13%) was observed for second review of cytology specimens by a team of expert cytopathologists. The expert diagnoses were supported in 95.5% of cases where histological follow up was available. Our results emphasize that cytopathology is a subspecialization of pathology that requires specialized cytopathologists.

As judged from our studies and from literature [12-20], double reading is a valuable tool to measure and reduce diagnostic errors. However, with the upscaling of pathology services through merging of pathology laboratories, long-distance double reading, is very time and labor-intensive, as glass slides are to be sent by mail. The use of digital whole slide images (WSI) is very helpful for easy and timely long-distance double reading and consultation [21]. For histology specimens, WSI are very well validated [21]. WSI are less widely used for cytological specimens due to several limitations, such as long scanning time and focus problems, because of the three-dimensional orientation of cells. However, due to improvements in technology, for example better focusing and Z-stack options, now better cytology WSI can be created [22].

Co-testing

Interobserver variation is also present in the evaluation of IHC stainings, which may affect diagnosis, prognosis and treatment choice. This is for example the case for evaluation of HER-2 overexpression [23], and has consequences for patient selection for treatment with anti-HER-2 therapies, such as trastuzumab and lapatinib. Furthermore, HER-2 IHC is affected by pre-analytical and tissue processing variation.

It is common practice to reflex test IHC equivocal (2+) cases with a gene amplification test, usually fluorescence (FISH) or chromogenic *in situ* hybridization (CISH) or MLPA [24,25]. However, IHC negative patients might demonstrate gene amplification and might respond to anti-HER-2 therapies, whereas IHC positive patients might lack gene amplification and thus not respond.

To overcome the problem of interobserver variation, routinely co-testing every invasive breast cancer case, instead of only the IHC equivocal ones, with an amplification test was proven to be useful (**Chapter 7**). At Symbiant and the University Medical Centre Utrecht, co-testing is performed with IHC and the quantitative PCR-based MLPA technique (and CISH in selected cases). We demonstrated significant interlaboratory variation with regard to frequencies of scoring IHC 0/1+ and 2+, indicating pre-analytical standardization difficulties and interpretational challenges of IHC. Co-testing revealed 11 discordant cases: gene amplification was observed in 4 out of 490 IHC 0/1+ cases (0.8%), and 7 out of 105 IHC 3+ cases (6.7%) lacked gene amplification. IHC score was adapted in 8/11 discordant cases after review or repeated IHC staining for study purposes, resulting in 3 true discordant cases. We concluded that routinely co-testing all invasive breast cancers with IHC and MLPA may improve the quality of HER-2 testing.

However, the most important question, whether patients with equivocal HER-2 status and patients with a discordance between HER-2 protein expression and gene amplification will respond or not to HER-2 targeted therapies, such as trastuzumab, remains unanswered. Unfortunately, this is a difficult question to address, because of the limited numbers of such patients. Currently, a clinical trial (NCT01275677) concerning "HER-2 low" patients, i.e. IHC 1+ or 2+ and/or unamplified with ISH, is ongoing. A meta-analysis, combining the results of clinical trials concerning equivocal and/or discordant patients, is needed to answer these questions. As long as the question remains unanswered we think that co-testing is a good solution to identify all patients that might respond to anti-HER-2 treatment, and to obtain a large group of patients with discordant IHC and gene amplification results to include in scientific research.

Feedback on performance

Feedback to pathologists and transparency of performance data may further improve diagnostic accuracy. In daily practice, pathologists receive feedback by means of consultation and discussing difficult cases with peers. Furthermore,

if routine double reading is performed, discordant results are preferably fed back to the initial pathologist. This gives pathologists an extra opportunity to learn from errors/inaccurate diagnoses that were probably not encountered without routine double reading, as the pathologist was sure enough about the diagnosis to sign out the case without consulting a peer.

In addition, feedback on individual laboratory performance in comparison to the other laboratories in The Netherlands can be given by assessing data from PALGA, as we did for grading of colorectal neoplasms. This enables laboratories with aberrant results (in this case aberrant proportions of adenomas with high-grade dysplasia and/or poorly differentiatedadenocarcinoma) to undertake action.

An increasing number of pathology reports in The Netherlands are reported synoptically instead of narratively [26,27]. The use of synoptic pathology reporting enables easy data extraction and analysis of large numbers of pathology reports, as all parameters are stored in PALGA as separate variables. This makes regular and timely feedback to pathology laboratories, but also to the clinic (e.g. the number of positive surgical margins), possible. Pathology laboratories can also easily determine individual pathologists' performance themselves. Currently, a synoptic reporting module for molecular data is being developed as well. Thus, in the near future, it will also be possible to compare molecular testing results between laboratories in a relatively easy and fast manner, which is very interesting for e.g. colorectal cancer, melanoma, and non-small cell lung cancer because of its many different (targetable) mutations and/or translocations. In addition to the expected indirect improvement of patient care by enabling easier feedback, the use of synoptic reporting will probably also directly contribute to better patient care by better standardization due to uniformity of reporting and more complete reports (review Sluijter *et al*, submitted).

The role of autopsies in quality improvement of health care

Since the autopsy is considered the gold standard for the evaluation of medical practice, proper and timely feedback of autopsy findings to clinicians is important to improve health care as well [28]. Feedback gives clinicians the opportunity to learn from their previous decisions and performance and assists them in making future decisions when a similar case is encountered. Immediate feedback is more effective than delayed feedback as clinicians have not yet 'closed' the case in his/her mind [29]. Furthermore, delayed feedback

of autopsy findings is reported as a reason for clinicians to request an autopsy less often [29,30]. A significant reduction in time to autopsy report completion by specialized autopsy pathologists was observed in **Chapter 8**.

Although some seem to think that the autopsy has become superfluous because of advances in medicine (such as modern imaging techniques), we showed in **Chapter 8** that major discrepancies persist in 23.5% of cases when clinical diagnoses and autopsy findings were compared using the Goldman classification. Major discrepant diagnoses are related to the cause of death and knowledge prior to death would have changed management of care and could have prolonged survival or cured the patient (class I) or probably would not have changed the outcome (class II). The proportion of major discrepancies is comparable with that found in literature [31].

In addition to evaluation of medical practice, autopsies can also reveal hereditary abnormalities that might have consequences for family members of the deceased person. Furthermore, it is a unique instrument for scientific research. For example to learn about the metastatic patterns of different cancer types [32,33] or the effect of therapies [34] to optimize treatment strategies.

Despite the advantages, autopsy rates have been declining over the last decades to less than 10%. In 1993, 9,811 autopsies were performed in The Netherlands, whereas in 2013 this number declined to only 3,785 autopsies, which was approximately 3% of all deaths in that year [35]. The NVVP considers this an alarming development, and refers to the high discrepancy rate that we observed in our study [35].

The decline in the number of autopsies has several causes, including the non-reimbursement of autopsies, clinicians' fear of medicolegal problems and reluctance in asking consent, reliance on laboratory tests and imaging results, and delayed feedback of autopsy results. We demonstrated that the problem of delayed feedback can be resolved by the employment of specialized autopsy pathologists, as time to report completion significantly decreased.

The results of **Chapter 8** will hopefully make clinicians and pathologists more aware of the persistent roles of the autopsy in improving quality of health care, even with the advances in medicine, and stimulate clinicians to request more autopsies. Furthermore, we hope to convince insurers to reconsider reimbursement of autopsies.

Cost-effectiveness

The quality interventions proposed in this thesis are accompanied by a slight increase in pathology costs. This increase in costs is probably the main reason that dissuades pathology laboratories from implementing these quality interventions into practice. However, we expect that by investing in good quality of pathology practice, costs can be saved elsewhere, e.g. costs of incorrect treatment [36], unnecessary hospital stay, or possible litigation costs in case of an incorrect diagnosis. No cost-effectiveness analyses were performed and also in literature these are lacking, except for one study [37], which reported that double reading of prostate biopsies prior to radical prostatectomy is cost-effective. Pathology laboratories, however, must be compensated for double reading and other quality interventions, by reallocation of budgets. More cost-effectiveness studies are eagerly awaited to convince pathologists and directors that they should not be discouraged by financial reasons to implement quality interventions.

Conclusions

The main conclusion of this thesis is that quality of pathology practice and patient care can be improved by incorporating relatively simple quality interventions. Although the interventions described will slighty increase costs of pathology, global costs of pathology will still be minimal, and improved diagnostic accuracy and subsequent treatment stratification probably will ultimately reduce health care costs elsewhere.

Most importantly, every patient has the right to receive optimal quality health care. The results of this thesis will certainly contribute to this and will hopefully stimulate laboratories/pathologists to incorporate one or more of the quality interventions into their own practice.

References

1. Nakhleh RE. (2006) What is quality in surgical pathology? J Clin Pathol 59(7) 669-72.
2. Raab SS, Grzybicki DM. (2006) Anatomic pathology workload and error. Am J Clin Pathol 125(6): 809-12.
3. Warth A, Stenzinger A, Andrulis M, Schlake W, Kempny G, Schirmacher P, et al. (2015) Individualized medicine and demographic change as determining workload factors in pathology: quo vadis? Virchows Arch (epub ahead of print).
4. Cheung CC, Torlakovic EE, Chow H, Snover DC, Asa SL. (2015) Modeling complexity in pathologist workload measurement: the Automatable Activity-Based Approach to Complexity Unit Scoring (AABACUS). Mod Pathol 28(3): 324-39.
5. Dutch guideline for colorectal cancer. Available from: http://www.oncoline.nl/index.php?pagina=/richtlijn/item/pagina.php&id=37094&richtlijn_id=933.
6. Goldstein NS. (2001) Recent pathology related advances in colorectal adenocarcinomas. Eur J Surg Oncol 27(5): 446-50.
7. Galvis CO, Raab SS, D'Amico F, Grzybicki DM. (2001) Pathologists' assistants practice: a measurement of performance. Am J Clin Pathol 116(6): 816-22.
8. Dutch Society for Pathology (NVVP): Standpunt taakherschikking binnen de pathologie. 2011; Available from: http://www.pathology.nl/images/Voorlichting/Definitief_standpunt_Taakherschikking_in_de_pathologie_11jan2012.pdf.
9. Blaauwgeers JL. (2013) Task differentiation, in pathology as well. Ned Tijdschr Geneeskd 157(19): A6187.
10. Sarewitz SJ. (2014) Subspecialization in community pathology practice. Arch Pathol Lab Med 138(7): 871-2.
11. Madani A, Kuijpers CCHJ, von der Thüsen J, Seegers PA, Grünberg K, Overbeek LIH, Nagtegaal ID. (2015) Onderzoeksrapport SKMS Vermindering van de praktijkvariatie: Analyse van een mogelijke strategie bij colorectale adenomen. Available from: http://pathology.nl/images/Onderzoeksrapport_SKMS_colorectale_adenomen_30112015_-_definitieve_versie.pdf.
12. Whitehead ME, Fitzwater JE, Lindley SK, Kern SB, Ulirsch RC, Winecoff WF 3rd. (1984) Quality assurance of histopathologic diagnoses: a prospective audit of three thousand cases. Am J Clin Pathol 81(4): 487-91.
13. Safrin RE, Bark CJ. (1993) Surgical pathology sign-out. Routine review of every case by a second pathologist. Am J Surg Pathol 17(11): 1190-2.
14. Lind AC, Bewtra C, Healy JC, Sims KL. (1995) Prospective peer review in surgical pathology. Am J Clin Pathol 104(5): 560-6.
15. Renshaw AA1, Pinnar NE, Jiroutek MR, Young ML. (2002) Quantifying the value of in-house consultation in surgical pathology. Am J Clin Pathol 117(5): 751-4.
16. Weydert JA, De Young BR, Cohen MB. (2005) A preliminary diagnosis service provides prospective blinded dual-review of all general surgical pathology cases in an academic practice. Am J Surg Pathol 29(6): 801-5.
17. Croce-Kleinmann S, Marcellin L, Neuville A, Onéa A, Lindner V, Casnedi S, et al. (2008) Internal quality control of histological diagnosis in pathology. A nine-year experience. Ann Pathol 28(1): 9-16.
18. Layfield LJ, Jones C, Rowe L, Gopez EV. (2002) Institutional review of outside cytology materials: a retrospective analysis of two institutions' experiences. Diagn Cytopathol 26(1): 45-8.
19. Lueck N, Jensen C, Cohen MB, Weydert JA. (2009) Mandatory second opinion in cytopathology. Cancer 117(2): 82-91.

20. Bomeisl PE Jr, Alam S, Wakely PE Jr. (2009) Interinstitutional consultation in fine-needle aspiration cytopathology: a study of 742 cases. Cancer 117(4): 237-46.
21. Al-Janabi S, Huisman A, Van Diest PJ. (2012) Digital pathology: current status and future perspectives. Histopathology 61(1): 1-9.
22. Bongaerts O, van Diest PJ, Pieters M, Nap M. (2015) Working toward consensus among professionals in the identification of classical cervical cytomorphological characteristics in whole slide images. J Pathol Inform 6: 52.
23. Hoang MP, Sahin AA, Ordòñez NG, Sneige N. (2000) HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. Am J Clin Pathol 113(6): 852-9.
24. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. (2014) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Arch Pathol Lab Med 138(2): 241-56.
25. Moelans CB, de Weger RA, van Blokland MT, Ezendam C, Elshof S, Tilanus MG, et al. (2009) HER-2/neu amplification testing in breast cancer by multiplex ligation-dependent probe amplification in comparison with immunohistochemistry and in situ hybridization. Cell Oncol 31(1): 1-10.
26. Srigley J, Lankshear S, Brierley J, McGowan T, Divaris D, Yurcan M, et al. (2013) Closing the quality loop: facilitating improvement in oncology practice through timely access to clinical performance indicators. J Oncol Pract 9(5): e255-61.
27. Williams CL, Bjugn R, Hassell LA. (2015) Current status of discrete data capture in synoptic surgical pathology and cancer reporting. Pathol Lab Med Int 7: 11-22.
28. van den Boogert J, den Hoed PT, Veen HF, Harteloh PP. (2003) The necrology meeting as an instrument for quality of care improvement; experiences at the Ikazia Hospital, 1990-2000. Ned Tijdschr Geneeskd 147(13): 599-603.
29. van den Tweel JG. (2008) Autopsy pathology should become a recognised subspecialty. Virchows Arch 452(5): 585-7.
30. Yawson AE, Tette E, Tettey Y. (2014) Through the lens of the clinician: autopsy services and utilization in a large teaching hospital in Ghana. BMC Res Notes 7: 943.
31. Shojania KG, Burton EC, McDonald KM, Goldman L. (2003) Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. JAMA 289(21): 2849-56.
32. Hugen N, van de Velde CJ, de Wilt JH, Nagtegaal ID. (2014) Metastatic pattern in colorectal cancer is strongly influenced by histological subtype. Ann Oncol 25(3): 651-7.
33. Cummings MC, Simpson PT, Reid LE, Jayanthan J, Skerman J, Song S, et al. (2014) Metastatic progression of breast cancer: insights from 50 years of autopsies. J Pathol 232(1): 23-31.
34. Morrissey C, Roudier MP, Dowell A, True LD, Ketchanji M, Welty C, et al. (2013) Effects of androgen deprivation therapy and bisphosphonate treatment on bone in patients with metastatic castration-resistant prostate cancer: results from the University of Washington Rapid Autopsy Series. J Bone Miner Res 28(2): 333-40.
35. Dutch Society for Pathology (NVVP). (2014) Pathologen: verontrustend weinig obducties. Available from: http://www.medischcontact.nl/Actueel/Nieuws/Nieuwsbericht/143689/Pathologen-verontrustend-weinig-obducties.htm.

36. Ho J, Ahlers SM, Stratman C, Aridor O, Pantanowitz L, Fine JL, et al. (2014) Can digital pathology result in cost savings? A financial projection for digital pathology implementation at a large integrated health care organization. J Pathol Inform 5: 33.
37. Epstein JI, Walsh PC, Sanfilippo F. (1996) Clinical and cost impact of second-opinion pathology. Review of prostate biopsies prior to radical prostatectomy. Am J Surg Pathol 20(7): 851-7.

Nederlandse samenvatting

10

Het pathologieverslag bevat vaak prognostische informatie over de gerapporteerde ziekte en is vaak de basis voor beslissingen omtrent de behandeling, met name in de oncologie. Daarom is het van het grootste belang dat de diagnose en het pathologieverslag juist, tijdig en volledig zijn [1]. De hoofddiagnose moet juist zijn, maar ook de kritische parameters die gerelateerd zijn aan de diagnose, prognose en behandelkeuze, zoals tumorgrootte, histologisch subtype, graad en stadium moeten correct bepaald en gerapporteerd worden. Het pathologieverslag is het eindproduct van alle stappen binnen het diagnostische proces op de afdeling pathologie, vanaf de aankomst van het materiaal, macroscopie, microscopie, aanvullende tests (zoals immunohistochemie (IHC) en moleculaire tests) tot en met het opstellen van het uiteindelijke verslag. Het is onvermijdelijk dat in alle stappen van dit proces uiteindelijk fouten ontstaan. Het is de taak van alle betrokkenen (waaronder analisten, secretaresses, pathologen (in opleiding) en bestuurders) om het foutenpercentage te minimaliseren.

Het doel van dit proefschrift was de kwaliteit van een aantal aspecten van de pathologie te analyseren en te verbeteren en daardoor verbetering van de gezondheidszorg en patiëntveiligheid. Hiertoe hebben wij verschillende diagnostische processen binnen de pathologie onderzocht en de toegevoegde waarde van een aantal kwaliteitsinterventies, met een focus op oncologie, bepaald. We hebben ons gericht op een aantal problemen die kunnen leiden tot een incorrecte diagnose en een onjuist pathologieverslag. Het eerste probleem waarop we ons hebben gericht was dat van hoge werkdruk onder pathologen. Het tweede veel voorkomende probleem dat wij hebben aangepakt was dat van variatie tussen beoordelaars (interobserver variatie). Daarnaast hebben we de rol van obducties in kwaliteitsverbetering van de gezondheidszorg bepaald. Dit hoofdstuk geeft een overzicht van de belangrijkste bevindingen van dit proefschrift en bespreekt toekomstperspectieven.

Belangrijkste bevindingen van dit proefschrift

De resultaten van dit proefschrift tonen duidelijk aan dat de kwaliteit van de pathologie en patiëntenzorg aanzienlijk kan worden verbeterd door het toepassen van een aantal relatief eenvoudige kwaliteitsinterventies:

1. Taakherschikking door het aanstellen van "pathologists' assistants" (PA's) om routine taken van pathologen over te nemen en door subspecialisatie van pathologen.

2. Het inbouwen van extra controles door het dubbel kijken (double reading) van een selectie van histo- en cytopathologische preparaten (door gespecialiseerde pathologen) en het routinematig co-testen van alle invasieve borstkankers voor HER-2-status met IHC en een gen-amplificatie test (bijvoorbeeld "multiplex ligation-dependent probe amplification" (MLPA)).
3. Het terugkoppelen van verrichtingen aan pathologielaboratoria, individuele pathologen en clinici door verrichtingen van laboratoria op landelijk niveau te vergelijken, door discordante diagnosen na double reading terug te koppelen en door bevindingen bij obductie terug te koppelen aan clinici.

Discussie en toekomstperspectieven

Vermindering werkdruk pathologen

Het eerste probleem waarop we ons hebben gericht was dat van hoge werkdruk onder pathologen, dat naar verwachting een negatieve invloed heeft op de kwaliteit van de pathologie [2]. In de komende decennia zal de werkdruk naar verwachting nog meer toenemen door vergrijzing van de bevolking (toename in het aantal maligne en chronische ziekten), door toename in geneeskunde op maat (personalized medicine) en daarmee samenhangend de behoefte aan diepgaande kennis over alle nieuwe diagnostische testen [3,4], en door de introductie van nieuwe bevolkingsonderzoeken.

Taakherschikking

De hoge werkdruk onder pathologen kan worden verminderd door het inzetten van PA's om bepaalde routinematige taken van pathologen over te nemen, waaronder het uitsnijden van relatief eenvoudige preparaten (zoals galblaas, appendix, en huidbiopten) en oncologische resectiepreparaten (zoals borst- en darmkanker), en het oogsten lymfeklieren. Voordat de opleiding voor PA's in Nederland werd ingevoerd in 2008 heerste er een discussie of het inzetten van PA's, die niet de 5-jarige opleiding in de pathologie hebben gehad, een bedreiging zou zijn voor de kwaliteit van de patiëntenzorg of dat het juist zou leiden tot een kwaliteitsverbetering. Wij onderzochten dit in één specifiek aspect, namelijk het oogsten van lymfeklieren uit darmkanker resectiepreparaten (**hoofdstuk 2**).

We toonden aan dat PA's bijdroegen aan een kwaliteitsverbetering, althans in dit specifieke aspect van hun takenpakket. Zij oogstten gemiddeld significant meer lymfeklieren per resectiepreparaat dan pathologen deden en oogstten in een significant hoger percentage resectiepreparaten voldoende lymfklieren (dat wil zeggen 10 of meer lymfeklieren volgens de Nederlandse richtlijn [5]) dan

pathologen deden. Dit resulteerde in nauwkeurigere stadiëring van de tumoren en betere selectie van vervolgbehandeling. Het aantal colonkanker patiënten dat alleen door te weinig geoogste lymfeklieren (minder dan 10) in aanmerking zou komen voor adjuvante (= na chirurgische verwijdering) chemotherapie was significant lager wanneer PA's in plaats van pathologen de lymfeklieren oogstten.

Verdere studies zijn nodig om de toegevoegde waarde van PA's in andere aspecten van hun takenpakket te bepalen. Men kan hierbij bijvoorbeeld denken aan het aantal gevallen waarbij het uitsnijden van extra weefsel nodig blijkt (resubmission rate). Goldstein en collega's [6] rapporteerden op basis van hun eigen ervaringen dat maar heel weinig weefselblokjes die PA's hadden uitgesneden opnieuw gesneden moesten worden en extra opwerking nodig hadden doordat zij in eerste instantie te dik waren. Galvis en collega's [7] rapporteerden dat wanneer het uitsnijden was gedaan door PA's in plaats van pathologen in opleiding er in minder gevallen extra weefselblokjes nodig waren voor microscopische beoordeling.

Momenteel zijn er in Nederland 83 PA's werkzaam en nog eens twaalf analisten zijn in opleiding tot PA. De Nederlandse Vereniging voor Pathologie (NVVP) erkent dat PA's zeer goed specifieke taken van pathologen kunnen overnemen, mits aan de volgende kwaliteitseisen wordt voldaan: 1) een gedegen opleiding, 2) duidelijke supervisie met regelmatige evaluatie van het functioneren, 3) duidelijke afspraken over taken en bevoegdheden, en 4) de aanwezigheid van actuele protocollen [8].

Doordat de werkdruk van pathologen verder toeneemt met de introductie van bevolkingsonderzoeken (BVO's) zouden PA's naast macroscopie taken ook microscopie taken kunnen overnemen van pathologen. Men kan hierbij denken aan pre-screening van colorectale poliepen die in het kader van het BVO darmkanker ontvangen zijn, zoals in een aantal laboratoria al het geval is, of huidtumoren, en het opstellen van een rapport voor deze tumoren dat na controle door de patholoog kan worden geautoriseerd [9]. De kwaliteit van microscopie pre-screening door PA's wordt constant gecheckt door alle casus opnieuw met een patholoog te bekijken, net zoals bij pathologen in opleiding.

Subspecialisatie van pathologen is een andere manier van taakherschikking om de kwaliteit van de zorg te verbeteren [10] en de werkdruk te verlagen door een toename in expertise en daarop volgend verminderde complexiteit.

In **hoofdstuk 6** toonden we een verbetering aan in de kwaliteit van cytologie diagnostiek door double reading van een selectie van cytologie casus door een team van gespecialiseerde cytopathologen. Daarnaast vonden we in **hoofdstuk 8** dat gespecialiseerde obductiepathologen significant sneller het obductieverslag afrondden dan 'algemene' pathologen en dat deze verslagen vaak ook uitgebreider waren. Deze resultaten benadrukken de noodzaak van subspecialisatie binnen de pathologie.

Verminderen van interobserver variatie

Het andere probleem waarop we ons in deze thesis hebben gericht is dat van interobserver variatie. Variatie in het diagnostische proces, en dus in de pathologie, is onwenselijk en standaardisatie is nodig wanneer variatie aanwezig is. De eerste stap naar standaardisatie is het inzichtelijk maken van de aanwezigheid van variatie.

In **hoofdstuk 3 en 4** brachten we de praktijkvariatie in het graderen van colorectale adenomen en carcinomen tussen Nederlandse pathologielaboratoria en individuele pathologen in kaart. Hiervoor maakten we gebruik van verslagen uit PALGA (de Nederlandse pathologie registratie). De variatie in gradering van colorectale adenomen (n = 32.391) en adenocarcinomen (n = 11.719) was aanzienlijk. Ook na correctie voor verschillen in case mix, diagnosticeerden een aantal laboratoria een significant afwijkend percentage adenomen met hooggradige dysplasie en / of slecht gedifferentieerde adenocarcinomen.

De resultaten van deze hoofdstukken illustreren dat er inderdaad variatie bestaat en dat standaardisatie van histologische gradering nodig is, bijvoorbeeld door educatie. Zo is er bijvoorbeeld een e-learning module (met afsluitend examen) ontwikkeld door pathologen in opdracht van het Rijksinstituut voor Volksgezondheid en Milieu (RIVM), met als doel de diagnostiek in het kader van het BVO darmkanker te standaardiseren. Met ingang van oktober 2014 mogen uitsluitend pathologen die de e-learning module hebben doorlopen en geslaagd zijn voor het examen de diagnostiek voor het BVO darmkanker doen. De variatie tussen laboratoria in het graderen van colorectale adenomen in 2013 (**hoofdstuk 3**) werd vergeleken met dat van niet-BVO adenomen van oktober 2014 t/m juni 2015. Onder de adenomen gediagnosticeerd op poliepectomieën was de variatie afgenomen, maar onder de adenomen gediagnosticeerd op biopten was de variatie juist toegenomen (SKMS onderzoeksrapport 'Vermindering van de praktijkvariatie: Analyse van een mogelijke strategie bij colorectale adenomen') [11]. Meer onderzoek is nodig om de exacte rol van de

e-learning module te bepalen.

Verder onderzoek zal moeten uitwijzen of er ook variatie is in andere aspecten, zoals in de histologische gradering van andere tumoren (bijvoorbeeld borstkanker) of andere kritische parameters (zoals histologisch subtype). We verwachten dat in deze aspecten vergelijkbare variatie tussen laboratoria en pathologen aanwezig is.

Andere manieren, naast educatie van individuele pathologen, om standaardisatie te verbeteren, zijn het inbouwen van extra controles in het diagnostische proces en het geven van feedback. Hieronder worden deze interventies in meer detail besproken.

Inbouwen van extra controles

In dit proefschrift hebben we twee manieren van extra controles onderzocht, namelijk double reading van een selectie casus (**hoofdstuk 5 en 6**) en het routinematig co-testen van patiënten met invasieve borstkanker met twee technieken om HER-2-status te bepalen (**hoofdstuk 7**).

Double reading

In **hoofdstuk 5** hebben we de toegevoegde waarde onderzocht van routine double reading van histologische preparaten voorafgaand aan bespreking op een multidisciplinair overleg. In ongeveer 1% van de gevallen resulteerde double reading in een discordante diagnose met mogelijk klinische relevantie ('major' discordantie). Major discordanties werden niet alleen gezien bij maligne, maar ook bij benigne gevallen, wat aangeeft dat double reading ook nuttig kan zijn voor (een selectie van) benigne casus.

Een hoger percentage (13%) major discordanties werd gezien in **hoofdstuk 6**, waar een selectie van cytologie preparaten double reading ondergingen door gespecialiseerde cytopathologen. In bijna alle gevallen (95,5%) waar histologische follow-up aanwezig was ondersteunde deze follow-up de diagnosen van de experts. Deze resultaten benadrukken dat cytopathologie een subspecialisatie binnen de pathologie is die de expertise van gespecialiseerde cytopathologen vereist.

Zowel uit de studies beschreven in dit proefschrift als in de literatuur [12-20] blijkt dat double reading een waardevol instrument is om het aantal

diagnostische fouten te meten en te verminderen. Echter, doordat steeds meer laboratoria fuseren wordt double reading op afstand tijds- en arbeidsintensief, omdat coupes per post verzonden moeten worden. Digitale pathologie door middel van 'whole slide images' (WSI) maakt tijdige consultatie en double reading op afstand op afstand mogelijk [21]. WSI zijn goed gevalideerd voor het beoordelen van histologische preparaten [21], maar worden in mindere mate gebruikt voor het beoordelen van cytologische preparaten, onder andere door de lange scantijd en focusproblemen, door de driedimensionale oriëntatie van de cellen. Echter, als gevolg van technologische verbeteringen, zoals betere focus- en Z-stack opties, kunnen nu ook cytologie WSI van betere kwaliteit worden gemaakt [22].

Co-testen

Ook het beoordelen van IHC kleuringen is onderhevig aan interobserver variatie, wat invloed kan hebben op diagnose, prognose en behandelkeuze. Dit is bijvoorbeeld het geval bij het beoordelen van HER-2 overexpressie [23] met consequenties voor selectie van patiënten die in aanmerking komen voor anti-HER-2 therapie, zoals trastuzumab en lapatinib. Daarnaast wordt HER-2 IHC beïnvloed door pre-analytische variatie en variatie in weefselverwerking.

Het is gebruikelijk om casus met een dubieus IHC resultaat (2+) verder te testen op de aanwezigheid van HER-2 genamplificatie, meestal door middel van fluorescentie (FISH) of chromogene *in situ* hybridisatie (CISH) of de op PCR gebaseerde kwantitatieve MLPA techniek [24,25]. Het is echter ook mogelijk dat bij IHC negatieve patiënten wel genamplificatie wordt aangetoond waardoor zij mogelijk toch reageren op anti-HER-2 therapie en dat bij IHC positieve patiënten geen genamplificatie wordt aangetoond waardoor zij mogelijk niet reageren.

In hoofdstuk 7 hebben we aangetoond dat het nuttig is om routinematig alle patiënten met invasieve borstkanker te co-testen met een genamplificatie test (MLPA en/of CISH), in plaats van alleen de patiënten met IHC 2+. De proporties van patiënten met invasieve borstkanker met IHC scores 0/1+ en 2+ verschilden significant tussen de laboratoria wijzend op moeilijkheden met pre-analytische standaardisatie en interpretatie van IHC. Met co-testen waren 11 gevallen discordant: 4 van de 490 IHC 0/1 + gevallen (0,8%) toonden genamplificatie aan en bij 7 van de 105 IHC 3+ gevallen (6,7%) werd er geen genamplificatie aangetoond. Na herbeoordeling en/of herhaalde kleuring in het kader van deze studie werd de IHC score in 8/11 discordante gevallen

aangepast, waarna er 3 echte discordante gevallen overbleven. Het routinematig co-testen van alle patiënten met invasieve borstkanker met IHC en MLPA kan de kwaliteit van HER-2 testen te verbeteren.

De belangrijkste vraag of patiënten met een dubieuze HER-2 status en patiënten met een discordantie tussen eiwitexpressie en genamplificatie wel of niet reageren op anti-HER-2 therapie, zoals trastuzumab, blijft echter onbeantwoord. Deze vraag is helaas ook lastig te beantwoorden, omdat dit maar een beperkt aantal patiënten betreft. Momenteel loopt er een klinische trial (NCT01275677) met "HER-2 low" patiënten met IHC 1+ of 2+ en/of geen genamplificatie met ISH. Om deze vraag te beantwoorden is een meta-analyse nodig waarin de resultaten van klinische trials met patiënten met dubieuze of discordante HER-2 status worden gecombineerd. Zolang hierover nog onduidelijkheid is, denken wij dat co-testen een goede manier is om alle patiënten te identificeren die mogelijk reageren op anti-HER-2 therapie. Daarnaast kan door middel van co-testen een grote groep patiënten worden verzameld met discordantie tussen eiwitexpressie en genamplificatie om te includeren in wetenschappelijk onderzoek.

Terugkoppeling van verrichtingen

Ook door terugkoppeling van verrichtingen aan pathologen en transparantie hierover kan de diagnostische accuraatheid verbeterd worden. Pathologen ontvangen in de dagelijkse praktijk feedback door middel van het consulteren van collega's en het bespreken van moeilijke gevallen met collega's. Daarnaast worden discordante diagnosen, in het geval van routine double reading, idealiter teruggekoppeld aan de eerste patholoog. Dit geeft pathologen een extra mogelijkheid om te leren van fouten/onjuiste diagnosen, die hoogstwaarschijnlijk niet waren ontdekt als er geen routine double reading gedaan was. De patholoog was blijkbaar zeker genoeg over de diagnose om het verslag te autoriseren zonder een collega te consulteren.

Ook kunnen verrichtingen van individuele laboratoria ten opzichte van de andere Nederlandse laboratoria worden teruggekoppeld door het analyseren PALGA data, zoals wij hebben gedaan voor het graderen van colorectale tumoren. Dit stelt laboratoria met afwijkende resultaten (in dit geval een afwijkend percentage adenomen met hooggradige dysplasie en/of slecht gedifferentieerde adenocarcinomen) in staat om actie te ondernemen.

Een toenemend aantal pathologieverslagen in Nederland wordt synoptisch opgesteld, door het invullen van zogenaamde protocolmodules, in plaats van narratief [26,27]. Door het gebruik van 'synoptic reporting' wordt eenvoudige data extractie en analyse uit grote aantallen pathologierapporten mogelijk, omdat alle parameters in PALGA zijn opgeslagen als afzonderlijke variabelen. Vervolgens kan regelmatige en tijdige feedback aan pathologielaboratoria, maar ook aan de kliniek (bijvoorbeeld het aantal niet-vrije chirurgische marges), worden gegeven. Daarnaast kunnen pathologielaboratoria zelf gemakkelijk de verrichtingen van individuele pathologen analyseren. Sinds kort is er ook een protocolmodule beschikbaar voor moleculaire uitslagen, waardoor het in de nabije toekomst ook mogelijk zal zijn om moleculaire uitslagen relatief eenvoudig en snel te vergelijken tussen laboratoria. Dit is zeer interessant voor bijvoorbeeld darmkanker, melanoom, en niet-kleincellig longkanker vanwege de vele verschillende mutaties en/of translocaties waarvoor gerichte therapie mogelijk is. Synoptic reporting zal, naast de verwachte indirecte verbetering van patiëntenzorg door het gemakkelijker terugkoppelen van resultaten, waarschijnlijk ook rechtstreeks bijdragen aan zorgverbetering door meer gestandaardiseerde, uniforme en complete verslagen (review Sluijter en collega's, ingediend).

De rol van obducties in kwaliteitsverbetering van de zorg

De obductie wordt beschouwd als de gouden standaard voor evaluatie van het medisch handelen. Adequate en tijdige terugkoppeling van obductiebevindingen aan behandelend artsen is dan ook van belang om de gezondheidszorg te verbeteren [28]. Het geeft artsen de mogelijkheid te leren van eerdere beslissingen en verrichtingen, wat hen helpt bij het nemen van toekomstige beslissingen in soortgelijke gevallen. Onmiddellijke feedback is effectiever dan late feedback, omdat artsen de casus in hun gedachten dan nog niet hebben 'afgesloten' [29]. Late feedback van obductiebevindingen wordt daarnaast door artsen ook genoemd als reden om minder obducties aan te vragen [29,30]. In **hoofdstuk 8** lieten we zien dat gespecialiseerde obductiepathologen het obductieverslag significant eerder afgerond hadden dan 'algemene' pathologen.

Sommigen lijken te denken dat obducties overbodig zijn geworden als gevolg van de ontwikkelingen in de geneeskunde (zoals moderne beeldvormingstechnieken). Echter, in **hoofdstuk 8** toonden we aan dat er in 23,5% van de obducties een major discordantie (volgens de Goldman classificatie) was tussen de klinische diagnosen en de bevindingen bij obductie. Major discordante diagnosen zijn gerelateerd aan de doodsoorzaak en als men de diagnose vóór het overlijden

zou hebben gesteld zou het beleid anders zijn geweest, waardoor de patiënt langer zou hebben geleefd of zou zijn genezen (klasse I) of waardoor de uitkomst waarschijnlijk niet anders zou zijn (klasse II). Het percentage major discordanties is vergelijkbaar met de literatuur [31].

Een andere manier waardoor obducties bijdragen in de kwaliteit van zorg is doordat bij obductie eventuele erfelijke afwijkingen kunnen worden ontdekt die mogelijk gevolgen hebben voor familieleden van de overledene. Daarnaast is de obductie een uniek instrument voor wetenschappelijk onderzoek. Men kan bijvoorbeeld kennis opdoen over het metastaseringspatroon van verschillende soorten kanker [32,33] of over het effect van bepaalde behandelingen [34] om daarmee behandelstrategieën verder te optimaliseren.

Het aantal obducties is, ondanks alle voordelen, de laatste decennia sterk afgenomen tot minder dan 10%. In 1993 werden er in Nederland 9.811 obducties gedaan. In 2013 waren dit er slechts 3.785 (ongeveer 3% van alle sterfgevallen in dat jaar) [35]. De NVVP beschouwt dit als een zorgwekkende ontwikkeling en verwijst daarbij naar het hoge percentage major discordanties uit onze studie [35].

De afname in het aantal obducties heeft verschillende oorzaken, waaronder de hoge kosten (obducties worden niet vergoed door zorgverzekeraars, waardoor de kosten volledig voor het ziekenhuis zijn), angst voor een eventuele tuchtzaak, terughoudendheid in het vragen van toestemming aan nabestaanden, het toenemende vertrouwen in laboratoriumonderzoek en beeldvorming, en late terugkoppeling van obductiebevindingen. We hebben aangetoond dat het probleem van late terugkoppeling kan worden verminderd door het inzetten van gespecialiseerde obductiepathologen, omdat zij de obductieverslagen significant eerder afgerond hadden.

Hopelijk maken de resultaten van **hoofdstuk 8** clinici en pathologen meer bewust van de aanhoudende rol van obducties in de kwaliteitsverbetering van de zorg, ondanks de ontwikkelingen in de zorg, en stimuleert het clinici om weer meer obducties aan te vragen. Daarnaast hopen we zorgverzekeraars te kunnen overtuigen om vergoeding van obducties opnieuw te overwegen.

Kosteneffectiviteit

De kwaliteitsinterventies voorgesteld in dit proefschrift gaan gepaard met een lichte toename in kosten van de pathologie. Dit is waarschijnlijk de belangrijkste reden die pathologielaboratoria ervan weerhoud deze kwaliteitsinterventies te implementeren. Toch verwachten wij dat door te investeren in goede kwaliteit van pathologie ergens anders kosten kunnen worden bespaard. Men kan hierbij denken aan kosten door onjuiste behandeling [36], onnodige ziekenhuisopname, of mogelijke claims bij onjuiste diagnosen. We hebben geen kosteneffectiviteit analyses uitgevoerd en deze ontbreken ook in de literatuur, op één studie na [37], waaruit werd geconcludeerd dat double reading van prostaatkanker biopten voorafgaand aan radicale prostatectomie kosteneffectief is. Pathologielaboratoria moeten echter wel voor double reading en andere kwaliteitsinterventies worden gecompenseerd, door herverdeling van budgetten. Meer kosteneffectiviteitsstudies zijn nodig om pathologen en bestuurders ervan te overtuigen dat zij zich niet moeten laten ontmoedigen door financiële redenen om kwaliteitsinterventies te implementeren in de praktijk.

Conclusies

De belangrijkste conclusie van dit proefschrift is dat de kwaliteit van de pathologie en patiëntenzorg aanzienlijk kan worden verbeterd door het toepassen van een aantal relatief eenvoudige kwaliteitsinterventies. De interventies leiden tot een lichte stijging in pathologiekosten, maar de globale kosten van de pathologie blijven nog steeds heel laag en uiteindelijk zullen elders in de gezondheidszorg kosten bespaard worden door betere diagnostiek en daaropvolgend behandelkeuzes.

Maar nog belangrijker is dat elke patiënt recht heeft op gezondheidszorg van optimale kwaliteit. De resultaten beschreven in dit proefschrift zullen hier zeker aan bijdragen en zullen hopelijk pathologen en/of bestuurders stimuleren om één of meerdere kwaliteitsinterventies te implementeren in hun eigen laboratorium.

References

1. Nakhleh RE. (2006) What is quality in surgical pathology? J Clin Pathol 59(7) 669-72.
2. Raab SS, Grzybicki DM. (2006) Anatomic pathology workload and error. Am J Clin Pathol 125(6): 809-12.
3. Warth A, Stenzinger A, Andrulis M, Schlake W, Kempny G, Schirmacher P, et al. (2015) Individualized medicine and demographic change as determining workload factors in pathology: quo vadis? Virchows Arch (epub ahead of print).
4. Cheung CC, Torlakovic EE, Chow H, Snover DC, Asa SL. (2015) Modeling complexity in pathologist workload measurement: the Automatable Activity-Based Approach to Complexity Unit Scoring (AABACUS). Mod Pathol 28(3): 324-39.
5. Dutch guideline for colorectal cancer. Available from: http://www.oncoline.nl/index.php?pagina=/richtlijn/item/pagina.php&id=37094&richtlijn_id=933.
6. Goldstein NS. (2001) Recent pathology related advances in colorectal adenocarcinomas. Eur J Surg Oncol 27(5): 446-50.
7. Galvis CO, Raab SS, D'Amico F, Grzybicki DM. (2001) Pathologists' assistants practice: a measurement of performance. Am J Clin Pathol 116(6): 816-22.
8. Dutch Society for Pathology (NVVP): Standpunt taakherschikking binnen de pathologie. 2011; Available from: http://www.pathology.nl/images/Voorlichting/Definitief_standpunt_Taakherschikking_in_de_pathologie_11jan2012.pdf.
9. Blaauwgeers JL. (2013) Task differentiation, in pathology as well. Ned Tijdschr Geneeskd 157(19): A6187.
10. Sarewitz SJ. (2014) Subspecialization in community pathology practice. Arch Pathol Lab Med 138(7): 871-2.
11. Madani A, Kuijpers CCHJ, von der Thüsen J, Seegers PA, Grünberg K, Overbeek LIH, Nagtegaal ID. (2015) Onderzoeksrapport SKMS Vermindering van de praktijkvariatie: Analyse van een mogelijke strategie bij colorectale adenomen. Available from: http://pathology.nl/images/Onderzoeksrapport_SKMS_colorectale_adenomen_30112015_-_definitieve_versie.pdf.
12. Whitehead ME, Fitzwater JE, Lindley SK, Kern SB, Ulirsch RC, Winecoff WF 3rd. (1984) Quality assurance of histopathologic diagnoses: a prospective audit of three thousand cases. Am J Clin Pathol 81(4): 487-91.
13. Safrin RE, Bark CJ. (1993) Surgical pathology sign-out. Routine review of every case by a second pathologist. Am J Surg Pathol 17(11): 1190-2.
14. Lind AC, Bewtra C, Healy JC, Sims KL. (1995) Prospective peer review in surgical pathology. Am J Clin Pathol 104(5): 560-6.
15. Renshaw AA1, Pinnar NE, Jiroutek MR, Young ML. (2002) Quantifying the value of in-house consultation in surgical pathology. Am J Clin Pathol 117(5): 751-4.
16. Weydert JA, De Young BR, Cohen MB. (2005) A preliminary diagnosis service provides prospective blinded dual-review of all general surgical pathology cases in an academic practice. Am J Surg Pathol 29(6): 801-5.
17. Croce-Kleinmann S, Marcellin L, Neuville A, Onéa A, Lindner V, Casnedi S, et al. (2008) Internal quality control of histological diagnosis in pathology. A nine-year experience. Ann Pathol 28(1): 9-16.
18. Layfield LJ, Jones C, Rowe L, Gopez EV. (2002) Institutional review of outside cytology materials: a retrospective analysis of two institutions' experiences. Diagn Cytopathol 26(1): 45-8.
19. Lueck N, Jensen C, Cohen MB, Weydert JA. (2009) Mandatory second opinion in cytopathology. Cancer 117(2): 82-91.

20. Bomeisl PE Jr, Alam S, Wakely PE Jr. (2009) Interinstitutional consultation in fine-needle aspiration cytopathology: a study of 742 cases. Cancer 117(4): 237-46.
21. Al-Janabi S, Huisman A, Van Diest PJ. (2012) Digital pathology: current status and future perspectives. Histopathology 61(1): 1-9.
22. Bongaerts O, van Diest PJ, Pieters M, Nap M. (2015) Working toward consensus among professionals in the identification of classical cervical cytomorphological characteristics in whole slide images. J Pathol Inform 6: 52.
23. Hoang MP, Sahin AA, Ordòñez NG, Sneige N. (2000) HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. Am J Clin Pathol 113(6): 852-9.
24. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. (2014) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Arch Pathol Lab Med 138(2): 241-56.
25. Moelans CB, de Weger RA, van Blokland MT, Ezendam C, Elshof S, Tilanus MG, et al. (2009) HER-2/neu amplification testing in breast cancer by multiplex ligation-dependent probe amplification in comparison with immunohistochemistry and in situ hybridization. Cell Oncol 31(1): 1-10.
26. Srigley J, Lankshear S, Brierley J, McGowan T, Divaris D, Yurcan M, et al. (2013) Closing the quality loop: facilitating improvement in oncology practice through timely access to clinical performance indicators. J Oncol Pract 9(5): e255-61.
27. Williams CL, Bjugn R, Hassell LA. (2015) Current status of discrete data capture in synoptic surgical pathology and cancer reporting. Pathol Lab Med Int 7: 11-22.
28. van den Boogert J, den Hoed PT, Veen HF, Harteloh PP. (2003) The necrology meeting as an instrument for quality of care improvement; experiences at the Ikazia Hospital, 1990-2000. Ned Tijdschr Geneeskd 147(13): 599-603.
29. van den Tweel JG. (2008) Autopsy pathology should become a recognised subspecialty. Virchows Arch 452(5): 585-7.
30. Yawson AE, Tette E, Tettey Y. (2014) Through the lens of the clinician: autopsy services and utilization in a large teaching hospital in Ghana. BMC Res Notes 7: 943.
31. Shojania KG, Burton EC, McDonald KM, Goldman L. (2003) Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. JAMA 289(21): 2849-56.
32. Hugen N, van de Velde CJ, de Wilt JH, Nagtegaal ID. (2014) Metastatic pattern in colorectal cancer is strongly influenced by histological subtype. Ann Oncol 25(3): 651-7.
33. Cummings MC, Simpson PT, Reid LE, Jayanthan J, Skerman J, Song S, et al. (2014) Metastatic progression of breast cancer: insights from 50 years of autopsies. J Pathol 232(1): 23-31.
34. Morrissey C, Roudier MP, Dowell A, True LD, Ketchanji M, Welty C, et al. (2013) Effects of androgen deprivation therapy and bisphosphonate treatment on bone in patients with metastatic castration-resistant prostate cancer: results from the University of Washington Rapid Autopsy Series. J Bone Miner Res 28(2): 333-40.
35. Dutch Society for Pathology (NVVP). (2014) Pathologen: verontrustend weinig obducties. Available from: http://www.medischcontact.nl/Actueel/Nieuws/Nieuwsbericht/143689/Pathologen-verontrustend-weinig-obducties.htm.

36. Ho J, Ahlers SM, Stratman C, Aridor O, Pantanowitz L, Fine JL, et al. (2014) Can digital pathology result in cost savings? A financial projection for digital pathology implementation at a large integrated health care organization. J Pathol Inform 5: 33.
37. Epstein JI, Walsh PC, Sanfilippo F. (1996) Clinical and cost impact of second-opinion pathology. Review of prostate biopsies prior to radical prostatectomy. Am J Surg Pathol 20(7): 851-7.

Appendix

Curriculum vitae
List of publications
Dankwoord

Curriculum vitae

Chantal Kuijpers was born on May 16th 1986 in Oosterhout, the Netherlands. After completing secondary school (athenaeum) at the Cambreur College in Dongen in 2004, she started studying physiotherapy at the Avans Hogeschool in Breda. During the first year, she realized that this was not the right study for her. In the summer of 2005 she moved to Amsterdam, for a study biomedical sciences at the Vrije Universiteit. After obtaining her bachelor degree, a master study in oncology followed, at VU Medical Centre, for which she graduated in 2010.

In 2011, she started her PhD project on "Quality assessment and improvements in pathology practice" at Symbiant Pathology Expert Centre, Alkmaar. Since April 2014, her PhD was expanded with another study on variation in grading of colorectal neoplasms at the department of pathology of the University Medical Centre Utrecht in combination with foundation PALGA, Houten. This research project perfectly fits within her thesis.

Currently, she is still working at the pathology department of the University Medical Centre Utrecht and foundation PALGA, on a research project on molecular testing in non-small cell lung cancer, with which she will continue after her dissertation.

Chantal is recently married to Arjen Epskamp.

List of publications

Publications

- **Kuijpers CCHJ**, Fronczek J, van de Goot FRW, Jiwa NM
  Hoofdstuk 30.4 "Obducties" in: Handboek patiëntveiligheid. Onder redactie van Rob Dillmann e.a. Tijdstroom Uitgeverij (2016)
  In press.

- **Kuijpers CCHJ**, Burger G, Al-Janabi S, Willems SM, van Diest PJ, Jiwa NM
  Improved quality of patient care through routine second review of histopathology specimens prior to multidisciplinary meetings
  Submitted for publication.

- Al-Janabi S, Horstman A, van Slooten HJ, **Kuijpers CCHJ**, Lai-A-Fat C, van Diest PJ, Jiwa NM
  Validity of whole slide images for scoring HER2 chromogenic in situ hybridization in breast cancer.
  Submitted for publication.

- **Kuijpers CCHJ**, Sluijter CE, von der Thüsen JH, Grünberg K, van Oijen MGH, van Diest PJ, Jiwa NM, Nagtegaal ID, Overbeek LIH, Willems SM
  Interlaboratory variability in the grading of colorectal adenocarcinomas in a nationwide cohort.
  Accepted for publication.

- Madani A, **Kuijpers CCHJ**, von der Thüsen J, Seegers PA, Grünberg K, Overbeek LIH, Nagtegaal ID
  SKMS onderzoeksrapport. Vermindering van de praktijkvariatie: Analyse van een mogelijke strategie bij colorectale adenomen (2015).
  Published at NVVP web site (www.pathology.nl)

- **Kuijpers CCHJ**, Sluijter CE, von der Thüsen JH, Grünberg K, van Oijen MGH, van Diest PJ, Jiwa NM, Nagtegaal ID, Overbeek LIH, Willems SM
  Interlaboratory variability in the grading of dysplasia in a nationwide cohort of colorectal adenomas.
  Histopathology, 2015 Dec 28. doi: 10.1111/his.12923 [Epub ahead of print].

- **Kuijpers CCHJ**, Visser M, Sie-Go DMDS, de Leeuw H, de Rooij MJ, van Diest PJ, Jiwa NM
  Improved cytodiagnostics and quality of patient care through double reading of selected cases by an expert cytopathologist.
  Virchows Archiv 2015; 466(6): 617-624.

- **Kuijpers CCHJ***, Fronczek J*, van de Goot FRW, Niessen HWM, van Diest PJ, Jiwa NM
  The value of autopsies in the era of high-tech medicine: discrepant findings persist.
  J Clin Pathol 2014; 67(6): 512-519.

- **Kuijpers CCHJ**, Moelans CB, van Slooten HJ, Horstman A, Hinrichs JWJ, Al-Janabi S, van Diest PJ, Jiwa NM
  Added value of HER-2 amplification testing by Multiplex Ligation-dependent Probe Amplification in invasive breast cancer.
  PLoS One 2013;8(12):e82018.

- **Kuijpers CCHJ**, van Slooten HJ, Schreurs WH, Moormann GRHM, Abtahi MA, Slappendel A, Cliteur V, van Diest PJ, Jiwa NM
  Better retrieval of lymph nodes in colorectal resection specimens by pathologists' assistants.
  Journal of Clinical Pathology 2013; 66(1): 18-23.

- Buijs JT, **Kuijpers CCHJ**, van der Pluijm G
  Targeted therapy options for treatment of bone metastases; beyond bisphosphonates. Review article.
  Current Pharmaceutical Design 2010; 16(27): 3015-27.

Oral presentations

- Pathasdag VAP. "Pathassers en kwaliteit pathologie: Lymfklier isolatie in colorectaal kanker". Utrecht, 21 November 2015.

- Kwaliteitssymposium pathologie dagen NVVP. "Praktijkvariatie in het graderen van colorectale adenomen". Nijmegen, 20 November 2015.

- Wetenschappelijk symposium PALGA. "Variatie in gradering van (pre-) maligne laesies tussen de verschillende pathologie-labs in Nederland". Utrecht, 7 October 2014.

Poster presentations

- Interlaboratory variability in grading of dysplasia in a nationwide cohort of colorectal adenomas.
  - 7th European Multidisciplinary Colorectal Cancer Congress. Amsterdam, November 2014.
  - Pathologendagen NVVP. Zeist, April 2015.
  - European Congress of Epidemiology – Healthy living. Maastricht, June 2015.
  - 27th European Congress of Pathology. Belgrade, September 2015.

- Interlaboratory variability in the histological grading of colorectal adenocarcinomas in a nationwide cohort.
  - 27th European Congress of Pathology. Belgrade, September 2015.

- Improved cytodiagnostics and quality of patient care by expert cytopathologists' second review.
  - Pathologendagen NVVP. Zeist, April 2014.

- The value of autopsies in the era of high-tech medicine: discrepant findings persist.
  - Pathologendagen NVVP. Zeist, April 2014. (Poster prize).

- Added value of HER-2 amplification testing by multiplex ligation-dependent probe amplification in breast cancer.
  - Pathologendagen NVVP. Zeist, April 2012.
  - 35th Annual San Antonio Breast Cancer Symposium, December 2012.

- Pathologists' assistants are important contributors to colorectal cancer pathology due to better lymph node retrieval.
  - Pathologendagen NVVP. Zeist, April 2012.
  - Nederlandse Chirurgendagen. Veldhoven, May 2012.

Dankwoord

Na al het harde werken is mijn proefschrift af en daar ben ik heel trots op. Maar dit heb ik uiteraard niet alleen gedaan. Ik wil iedereen bedanken die mij hierbij begeleid, geholpen en ondersteund heeft, op wat voor manier dan ook.

Allereerst wil ik mijn promotor, Prof. Dr. Paul van Diest, en copromotoren, Dr. Mehdi Jiwa en Dr. Stefan Willems, bedanken.

Beste Paul, dank voor het in mij gestelde vertrouwen om mij in dit promotietraject te begeleiden. Ik heb erg veel geleerd van je commentaren op en verbeteringen in mijn manuscripten, waardoor ik uiteindelijk steeds meer vertrouwen in mijn eigen schrijven heb gekregen. Dank daarvoor.

Beste Mehdi, ik heb, als biomedisch wetenschapper / niet-clinicus, erg veel geleerd over het reilen en zeilen binnen een pathologielaboratorium, wat mijn klinische blik heeft verruimd. Ik wil dan ook mijn dank naar je uitspreken voor de kans die je mij hebt gegeven om binnen Symbiant te starten met mijn promotietraject en voor alle werkoverleggen waarin, door jouw heldere visie, mijn onderzoeken steeds meer tot zijn recht kwamen en één geheel werden, met als uiteindelijke resultaat dit proefschrift.

Beste Stefan, dank dat je mij hebt willen begeleiden in de laatste fase van mijn promotie bij het UMC Utrecht en Stichting PALGA. Ik waardeer het enorm dat je, ondanks je drukke schema, altijd tijd vrijmaakt voor overleg. Ik vind het dan ook fijn dat wij onze samenwerking momenteel voortzetten in het project over moleculaire testen bij longkanker.

Pathologen van Symbiant: bedankt voor al jullie hulp bij het uitvoeren van mijn projecten, jullie inbreng tijdens het R&D en het lezen en verbeteren van mijn artikelen. Met name dank aan de volgende co-auteurs: Henk-Jan van Slooten, Vincent Cliteur, Gerard Burger, Mike Visser, Daisy Sie-Go, Judith Fronczek en Frank van de Goot.

Alle analisten van Symbiant (met name Anja, Henk, André, Mohammad, Marlon) die hebben meegewerkt aan de onderzoeken wil ik bedanken voor alle tijd en moeite die zij erin hebben gestoken. Martijn, bedankt voor het verzamelen van de vele coupes en blokjes.

Marcia, Anita, Sabine, Wendy en Willy, dank voor alle administratieve ondersteuning en het inplannen van de afspraken met Mehdi en Paul.

Alice, jou wil ik bedanken voor het verbeteren van het Engels in mijn artikelen (ik heb veel van je geleerd!). Voor alle hulp met de statistische analyses van de onderzoeken bij Symbiant wil ik Tjeerd van der Ploeg bedanken.

Lieve Shaimaa, Astrid, Mathilda en Judith: jullie hebben ervoor gezorgd dat ik een fantastische tijd heb gehad bij Symbiant. We hebben vele fijne gesprekken gevoerd, maar ook vooral veel gelachen. Ik ben heel blij dat ik jullie heb leren kennen en kijk nu alweer uit naar het volgende etentje samen!

Alle collega's van PALGA (Annette, Ariana, Bea, Bert, Caro, Esther, Hannelore, Ivette, Jolanda, Lucy, Paul, en Rick): bedankt voor de fijne werksfeer op het kantoor en jullie hulp wanneer dit nodig was. Jullie zijn een hele fijne groep collega's.

In het bijzonder dank aan Lucy en Caro. Lucy, bedankt voor de goede begeleiding bij mijn laatste twee onderzoeken. Jouw enthousiasme, adviezen en complimenten deden mij erg goed en hebben mij meer vertrouwen in mezelf gegeven. Caro, bedankt voor de gezellige tijd op de kamer bij PALGA (en in Belgrado!) en natuurlijk voor al je hulp bij alle statistische analyses. Door jullie epidemiologische blik ben ik wéér anders tegen de dingen gaan aankijken.

Het onderzoek dat ik bij het UMC Utrecht en Stichting PALGA deed (Hoofdstukken 4 en 5 van dit proefschrift) werd gesubsidieerd door het KWF, waarvoor mijn dank. Zonder deze subsidie was het niet mogelijk geweest dit onderzoek te doen.

Daarop aansluitend wil ik alle pathologielaboratoria die aan dit onderzoek hebben meegewerkt bedanken.

Natuurlijk wil ik ook mijn vriendinnen bedanken.

Lieve Esther, Linda en Kimberly: ik ken jullie al zo lang en we hebben al veel meegemaakt samen (waaronder zes prachtige vakanties!). Ondanks dat we wat verder uit elkaar wonen en elkaar niet zo heel vaak zien, weten we precies wat we aan elkaar hebben en is het altijd weer zoals vroeger. Dat is mij heel wat waard!

Lieve Kamaratski's (Ilona, Ilse, Jenny, Linda en Rosanne): wij hebben elkaar tijdens onze studie biomedische wetenschappen leren kennen en inmiddels, 10 jaar later, zijn we nog steeds een leuk clubje samen. Bedankt voor alle mooie en gezellige momenten die wij samen beleven. En: jullie hebben prachtig gezongen op onze bruiloft!

Lieve Simone, bedankt voor al je steun en begrip tijdens, maar ook al vóór, mijn promotietraject. Dit waardeer ik heel erg. Helaas woon je nu niet meer om de hoek, maar dat verandert zeker niets aan onze vriendschap!

Lieve schoonfamilie, Arie, Astrid, Niels en Lars: bedankt dat ik zo fijn ben opgenomen in jullie gezin en bedankt voor de interesse die jullie altijd tonen in mijn onderzoeken.

Lieve mama en Anton, ondanks dat jullie niet altijd helemaal begrijpen wat mijn werk precies inhoudt, waardeer ik jullie interesse ernaar enorm. Ik weet zeker dat jullie ontzettend trots op me zijn. Bedankt voor alle heerlijke weekenden in Dongen waarin we even volledig bij kunnen komen van alle drukte.

Lieve papa, helaas kun je dit allemaal niet meer meemaken. Wat zou je trots op me zijn geweest!

En als laatste wil ik mijn echtgenoot Arjen bedanken. Lieverd, je hebt heel veel geduld met mij moeten hebben en vaak gezegd dat het allemaal wel goed zou komen. En uiteindelijk is het dan ook goed gekomen, dit boekje is daarvan het grote bewijs. Jij bent de perfecte man voor mij. Op naar een geweldige toekomst samen!