Diploma Thesis

Interobserver variability in subtyping and estimating the extent of gastric intestinal metaplasia in patients with chronic atrophic gastritis

submitted by

Julia Lerch

to attain the academic degree

Doctor medicinae universae

(Dr. med. univ.)

at the

Medical University of Graz

conducted at the

Diagnostic & Research Institute for Pathology

under supervision of

Univ.-Doz. Univ. FA Dr.med. Cord Langner

Graz, 27.11.2021

Affidavit

I hereby confirm that the following diploma thesis has been written by myself without any support from third parties. For preparation no other sources than those indicated in the thesis have been used. The results of the study are currently under evaluation for publication in a peer-reviewed journal.

Graz, 27th of November 2021 *Julia Lerch eh.*

Acknowledgements

First and foremost, I want to thank my supervisor Univ. Doz. Dr. med Cord Langner for the opportunity to be a part of this study and to work together with an international group of expert pathologists. I am very grateful for his excellent guidance and encouragement in preparing this study for publication in a peer-reviewed journal and for sharing his passion for scientific work with me.

Furthermore, I want to thank all gastrointestinal expert pathologists who participated in this study. Without their reliable cooperation this diploma thesis would not have been possible. Finally, I thank Prof. Dr. Josef Haas for his statistical expertise in the sample size calculation for this study.

Last but not least, special acknowledgements go to my family, my partner and all my friends, who patiently supported and encouraged me during my study at the Medical University of Graz.

Table of contents

Abbreviations

GC	Gastric cancer
HP	Helicobacter pylori
WHO	World Health Organization
ICD	International Statistical Classification of Diseases and Related Health Problems
NSAID	Non-steroidal anti-inflammatory drug
GIM	Gastric intestinal metaplasia
OLGA	Operative Link on Atrophy Assessment
OLGIM	Operative Link on Intestinal Metaplasia Assessment
MALT	Mucosa associated lymphoid tissue
USA	United States of America
NO	Nitric oxide
Cag-A	Cytotoxin associated antigen
Vac-A	Vacuolating toxin
SNP	Single nucleotide polymorphism
IL	Interleukin
TNF	Tumour necrosis factor
TCGA	The Cancer Genome Atlas
MSI	Microsatellite-instable
MSS	Microsatellite-stable
ACGR	Asian Cancer Research Group
TNM	Tumor Node Metastasis
ESD	Endoscopic submucosal dissection
EMR	Endoscopic mucosal resection
CT	Computed tomography
PAS	Periodic acid Schiff
AB	Alcian-blue
HID	High iron diamine
H&E	Haematoxylin and eosin
IEN	Intraepithelial neoplasia
UICC	Union for International Cancer Control
pTNM	Pathological Tumor Node Metastasis

ESGE	European Society of Gastroenterology and Endoscopy
MAPS	Management of Patients with Precancerous Conditions in the Stomach
AGA	American Gastroenterological Association
BSG	British Society of Gastroenterology
RCT	Randomized controlled studies
ICC	Intraclass correlation coefficient
CI	Confidence interval
IQR	Interquartile range
SCNA	Somatic copy number alterations
TERT	Telomerase reverse transcriptase
HR	Hazard ratio

Index of figures

Index of tables

Zusammenfassung

Hintergrund: Der Subtyp und das Ausmaß der intestinalen Metaplasie (IM) im Magen kann für die Risikostratifikation bei Patient*innen mit präkanzerösen Magenläsionen herangezogen werden. Der inkomplette Subtyp und eine ausgedehnte IM sind mit einem erhöhten Risiko für Magenkrebs verbunden.

Zielsetzung: Diese Diplomarbeit untersuchte die Interobserver-Variabilität in der Typisierung der IM (komplette *versus* inkomplette IM) in der histologischen Diagnose von Patienten mit chronischer Gastritis. Als sekundäres Ziel wurde die Interobserver-Variabilität in der prozentualen Schätzung des Ausmaßes der IM analysiert. Faktoren mit potenziellem Einfluss auf die histologische Begutachtung und die klinischen Konsequenzen von Fehlklassifikationen wurden evaluiert.

Methode: Neun internationale Expert*innen für gastrointestinale Pathologie beurteilten 46 Fälle mit kompletter, inkompletter oder gemischter IM auf gescannten, mit Hämatoxylin und Eosin gefärbten Objektträgern. Die Ergebnisse wurden mit der von zwei weiteren Experten erstellten Konsensdiagnose verglichen. Die Interobserver-Variabilität wurde unter Verwendung der Kappa-Statistik bestimmt. Zusätzlich wurden alle elf Expert*innen Patholog*innen in die Berechnung der Interobserver-Variabilität in der prozentualen Schätzung des Ausmaßes der IM einbezogen. Die Übereinstimmung zwischen den Beobachtern wurde mit dem Intraclass-Korrelationskoeffizienten (ICC) getestet.

Ergebnisse: Die Übereinstimmung der Patholog*innen mit der Konsensdiagnose in der Unterscheidung einer überwiegend kompletten *versus* einer überwiegend inkompletten IM lag zwischen 78% bis 98 %. Die Kappa-Statistik zeigten einem moderaten bis nahezu perfekten Grad an Übereinstimmung (gewichtetes Kappa=0,464-0,984). Die Gesamtübereinstimmung aller elf Pathologen kann als substanziell klassifiziert werden (Fleiss' Kappa=0,716, 95% Konfidenzintervall: 0,677-0,755). Fehlklassifikationen mit Auswirkungen auf die klinische Entscheidungsfindung traten in 19 von 333 (5,7%) histologischen Bewertungen auf. Die Übereinstimmung in Fällen mit reiner IM war signifikant höher als in Fällen mit IM vom Mischtyp (p=0,010). Pathologen, die die Subtypisierung in der täglichen Routine anwenden, schnitten besser ab als diejenigen, die dies nicht tun (p=0,040). Die Übereinstimmung in der Schätzung des prozentualen Ausmaßes der IM war sehr gut (ICC 0,983, 95% Konfidenzintervall: 0,975-0,990).

Schlussfolgerung: Zusammenfassend lässt sich sagen, dass die Subtypisierung und die prozentuale Schätzung des Ausmaßes der IM auf Hämatoxylin- und Eosin-gefärbten Objektträgern mit hoher Übereinstimmung von gastrointestinalen Expert*innen-Patholog*innen durchgeführt werden kann. Die Einführung der Subtypisierung der IM als risikostratifizierendes Instrument in den aktuellen Praxisleitlinien der European Society of Gastrointestinal Endoscopy (ESGE) und der American Gastroenterological Association (AGA) war im Rahmen dieser Studie mit einer geringen Rate an Fehlklassifizierungen verbunden. Diese Diplomarbeit bildet die Grundlage für künftige Forschung auf diesem Gebiet, zum Beispiel durch Ausweitung der Studie auf allgemeine Patholog*innen, und für eine mögliche Einführung der prozentualen Bestimmung des IM Ausmaßes in den jeweiligen Leitlinien für das Management von Patient*innen mit präkanzerösen Magenläsionen.

Abstract

Background: Extensive gastric intestinal metaplasia (GIM) and the incomplete subtype are associated with an increased risk of gastric cancer. The extent and subtype of GIM may therefore be utilized in the risk stratification of patients with gastric precancerous lesions.

Aims: This diploma thesis aimed to examine the interobserver variability of subtyping GIM (incomplete versus complete) in the histological diagnosis of patients with chronic atrophic gastritis. As a secondary objective, interobserver variability in the estimation of the percental extent of GIM was investigated. Parameters with potential impact on the assessment and the clinical consequences of misclassification were evaluated.

Methods: Nine international gastrointestinal expert pathologists assessed 46 cases with complete, incomplete or mixed-type GIM on scanned haematoxylin and eosin-stained slides. Results were compared with the consensus diagnosis driven by two experts. Interobserver variability was evaluated by applying kappa statistics. Additionally, all eleven pathologists were included in a secondary analysis on the interobserver variability of estimating the overall percental extent of GIM. Interobserver agreement was calculated with the intraclass correlation coefficient (ICC).

Results: Regarding the predominant GIM pattern (predominant complete *versus* predominant incomplete GIM), the agreement between each observer and the consensus diagnosis ranged from 78% to 98%. Kappa statistics showed a moderate to almost perfect level of agreement (weighted kappa=0.464-0.984). The overall agreement of the participating pathologists can be classified as substantial (Fleiss' kappa=0.716, 95% confidence interval: 0.677-0.755). Misclassification with impact on clinical decision-making occurred in 19 out of 333 (5.7%) of case ratings. Agreement in cases with pure GIM was significantly higher than in cases with mixed-type GIM (p=0.010). Pathologists who apply subtyping in daily routine performed better than those who do not (p=0.040). The interobserver agreement in estimating the percental extent of GIM was very good (ICC 0.983, 95% confidence interval: 0.975-0.990).

Conclusion: In conclusion, subtyping and estimating the percental extent of GIM on haematoxylin and eosin-stained slides can be achieved with high interobserver

agreement among gastrointestinal expert pathologists. The implementation of GIM subtyping as a risk-stratifying tool in current practice guidelines by the European Society of Gastrointestinal Endoscopy (ESGE) and the American Gastroenterological Association (AGA) carried a low rate of misclassification within this study. This diploma thesis provides the basis for future research in the field, e.g., by expanding the evaluation to general pathologists in a nation-wide setting, and for the potential implementation of percental GIM assessment in the respective guidelines on gastric precancerous lesions.

1. Introduction

1.1 Background

Gastric cancer (GC) remains the worldwide third leading cause of cancer-related mortality and represents the fifth most common cancer (1). In 2020, GC was responsible for more than one million new cancer cases and for approximately 760,000 deaths (1). *Helicobacter pylori* (HP) infection and chronic atrophic gastritis stand at the beginning of a multi-step cascade to GC (2). Early detection of precancerous lesions facilitates less invasive treatment and may reduce the global burden of GC and its mortality (3-7). Therefore, clinicians have recently published guidelines on the management of gastric precancerous lesions (8).

1.2 Gastritis

The World Health Organization (WHO) defines gastritis in the International Statistical Classification of Diseases and Related Health Problems (ICD) as an injury of the gastric mucosa involving epithelial damage, mucosal inflammation, and epithelial cell regeneration except any epithelial defect (9). The diagnosis of gastritis is made after endoscopic and histological examination of the gastric mucosa (10,11).

To date several systems to classify gastritis exist (Table 1). In ICD-11, the WHO classifies gastritis based on its aetiology and pathophysiological genesis (9). The international recognized Sydney classification on gastritis (11) recommends reporting the diagnosis of gastritis by providing and synthesizing information on etiological, topographical and morphological aspects of the biopsy specimen. The aim of this classification was to establish reproductive and useful histological diagnoses when reporting gastritis in endoscopic biopsies (11).

Table 1: **Classification systems of gastritis.** Modified after Dixon et. al. (11) and the World Health Organization (9).

International Statistical Classification of Diseases and Related Health Problems ICD-11 (9)	**The updated Sydney System** (11)
- Autoimmune gastritis - *Helicobacter pylori* induced gastritis - Eosinophilic gastritis - Lymphocytic gastritis - Allergic gastritis - Gastritis due to duodenogastric reflux - Menetrier disease - Gastritis of unknown aetiology with specific endoscopic or pathological features - Gastritis due to external causes - Gastric phlegmon	- Non-atrophic - Atrophic Autoimmune Multifocal atrophic - Special forms Chemical Radiation Lymphocytic Non-infectious Granulomatous Eosinophilic Other infectious gastritides

1.2.1 Diagnostic aspects

For the diagnosis of gastritis, the Sydney classification (11) recommends taking at least five biopsy specimens: two from the greater and lesser curvature of antrum and corpus and one from the *incisura angularis.* If further lesions are found, they should be biopsied as well (11). Biopsies from different topographical sites should be preserved and labelled in separate vials (11). Information on medication, especially on the intake of non-steroidal anti-inflammatory drugs (NSAIDs), proton pump inhibitors and antibiotics, as well as a short description of focal endoscopic lesions should be provided by the endoscopist (11). Special staining for HP should be undertaken before declaring an inflamed biopsy specimen as HP-negative (11). In the diagnosis of HP gastritis, the following features should be graded with the help of visual analogue scales according to their severity into normal (0), mild (1), moderate (2) and marked (3): presence of HP, atrophy in antrum and corpus, presence of neutrophils (“activity of gastritis”) and mononuclear cells (“chronicity of gastritis”), and gastric intestinal metaplasia (GIM) (11).

“Grading” and “staging” of gastritis is important in order to enable proper risk stratification for patients at risk of GC (12,13). “Grading” refers to measuring the amount of mononuclear and granulocytic inflammatory infiltrate, whereas “staging” refers to assessing the extent of atrophy with or without metaplastic changes (12). The “Atrophy 2000 Club” proved a high level of interobserver agreement for categorizing chronic gastritis into three categories: indefinite for atrophy, atrophic or non-atrophic chronic gastritis (14). At a later date, histological staging systems such as the Operative Link on Atrophy Assessment (OLGA) (15) and the Operative Link on Intestinal Metaplasia Assessment (OLGIM) (16) have been introduced for the purpose of GC risk stratification.

1.2.2 Autoimmune gastritis

Autoimmune gastritis is a chronic progressive condition, in which cluster-of-differentiation-4 T-cells target the H+/K+-ATPase of the gastric parietal cells (17-21). This autoimmune process results in parietal cell loss, mucosal atrophy and metaplasia and may also be triggered by a longstanding HP infection (17-21). It is estimated that approximately 1-2% of the population is affected, with peaks of 4-5% among females at a higher age (17,20). Autoimmune gastritis can be categorized into different histological stages: In an early phase, the lamina propria is infiltrated by focal lymphocytes, eosinophils (17) and plasmacytes causing to glandular destruction (17,20,22,23). The inflammatory infiltrate is found predominantly in the basal glandular epithelium (17,20,23). Parietal cells may be replaced by pseudopyloric metaplasia (22). The inflammation persists in the florid phase with lymphocytes causing further atrophy (22). Parietal cells are replaced by intestinal and/or pseudopyloric metaplasia (17,24). The last phase represents diffuse chronic atrophic gastritis with multifocal GIM (17,22). Additionally, pancreatic acinar cell metaplasia might be a common finding (24) strongly suggesting an autoimmune pathogenesis (25). The antrum mucosa is usually not affected (22-24) or shows solely minor inflammatory changes (11,24). The loss of parietal cells leads to achlorhydria (11,20) which may result in neuroendocrine cell hyperplasia (11,20,23) and carcinoids (20,22). Further complications include the development of a hyperplastic polyp (20), of a pseudopolyp (20), of a pyloric gland adenoma (20), of dysplasia of the epithelium (17), of GC (17,20) and of pernicious anaemia, respectively (11,20). The diagnosis of autoimmune gastritis relies on the distinctive

histological features and on the detection of autoantibodies against intrinsic factor and parietal cells (20).

1.2.3 *Helicobacter pylori* gastritis

HP represents the major causative agent of gastritis, in particular of chronic atrophic gastritis (10-12). Persistent infection with HP can lead to gastric and duodenal ulcer disease, pernicious anaemia, gastric mucosa-associated lymphoid tissue (MALT) lymphoma and GC (10,26-28). In order to minimize the risk of GC, HP eradication is recommended ideally before premalignant lesions such as atrophy and GIM occur (10).

Countries with the highest HP prevalence are Latin America and the Caribbean, whereas Northern America shows the lowest prevalence (29). In Asia and Europe prevalence rates differ greatly among countries. The highest rates are observed in Kazakhstan and Serbia, the lowest rates in Indonesia and Belgium (29). Different socio-economic and environmental factors across these countries could explain this broad range of prevalence (29). The prevalence of HP is declining in industrialized countries (20,30), including the high prevalence country Japan (31), *inter alia* due to adequate eradication therapy (30,31).

HP gastritis either presents as non-atrophic gastritis or multifocal atrophic gastritis (11). In cases with non-atrophic gastritis, inflammation is located predominantly within the antrum or is equally distributed between antrum and corpus and lacks glandular atrophy (11). In the Western World the most common pattern of HP gastritis is the antral-predominant non-atrophic gastritis (12). If atrophy is detected, it is usually restricted to the antrum (12). The severity of inflammation in multifocal atrophic gastritis is similar in the antrum and the corpus with patches of atrophy and GIM forming particularly around the *incisura angularis* (11). These patches may expand further proximally and distally until they confluence (11). Multifocal atrophic gastritis and atrophic pangastritis are more likely to occur in countries with lower sanitary standards such as Southern and Eastern Asia, Latin America and to some extent Central, Eastern and Southern Europe (12) and are most likely due to persistent HP infection (21). However, there are still exceptions such as Japan, a country with one of the highest prevalence of atrophic gastritis and GC (12). Of note, multifocal atrophy and diffuse chronic gastritis can not only appear in individuals with HP gastritis but also in HP-negative cases or in previously infected

individuals (11). They can be a consequence of chronic bile reflux, NSAID intake or mucosal damage due to dietary factors, i.e., chemical gastritis or reactive gastropathy (11). In a study from Austria (24), post-HP gastritis occurred more frequently than active HP gastritis. Post-HP gastritis is defined as mild, non-active, chronic inflammation, with or without lymphoid aggregates and GIM (24).

1.2.4 Reactive gastropathy

The most common type of gastritis found in Austria is reactive gastropathy which can occasionally be found in combination with post-HP gastritis (24). In general, reactive gastropathy represents the most common type of gastritis in Western countries (32,33). This distinct entity was first described by Dixon et al. as “reflux gastritis” (34). The Sydney classification refers to it as “chemical gastritis” (11). The term reactive gastropathy is favourably used by authors who want to emphasize that this entity originates from regeneration in the gastric mucosa rather than from inflammation (11). Main causes of this entity are duodenal-gastric reflux or the intake of NSAIDs or ethanol (11). Morphological features include foveolar hyperplasia, oedema of the lamina propria, vasodilatation and congestion of superficial mucosal capillaries and ascending interfoveolar smooth muscle fibres (11,34). Other key features are the lack of neutrophils and few to no chronic inflammatory cells (11,34). Of note, not all features have to be present in one patient and concomitant HP infection can hide histological changes (32).

1.3 Gastric cancer

1.3.1 Epidemiology

The epidemiology of GC partly mirrors the epidemiology of HP infection (35). The incidence of GC is highest in Eastern Asia and Eastern Europe, lower in Northern America and Northern Europe and lowest in African regions. (1). In regions with high GC incidence rates (Asia, South and Central America, Eastern Europe) 80% of all gastric adenocarcinomas are located further distal in the stomach, especially around the antral-pyloric region (35). On the contrary, in regions with lower incidence rates (Northern Europa, United States of America (USA)) 50-60% of all gastric adenocarcinomas are located in the cardia or fundus (35). Whereas intestinal-type GC (36) is more common in high-risk populations, the diffuse type (36) is more

frequent in low-risk populations (37). Some countries have reported a decrease in the rate of intestinal-type GC with a respective increase in the incidence of diffuse gastric adenocarcinoma (30,35,37). The increase in diffuse-type GC in adults younger than 50 years was mainly observed in low-risk countries such as the USA or UK with a low prevalence of HP infection (38). Increasing obesity among young people was identified as a possible causative factor for the increase in the incidence of cardia and diffuse-type GC within this age group (30,38-40). A population-based study from the USA revealed that the overall incidence of GC and the incidence of intestinal-type GC decreased from 1975 to 2015, whereas diffuse-type GC significantly increased among 40-to 49-year-old individuals (30). Arnold et al. (38) assessed global GC incidence trends and predicted a further decrease in the overall incidence rates of GC, especially in high-incidence countries. Still, the absolute burden of GC is set to remain stable, if not increase (38).

A significant male predominance in GC was noted in a study by Zamani et al. (29), whereas no significant gender difference was noted in the prevalence of HP infection. On the other hand, a meta-analysis by Ibrahim et al. (41) suggested a higher prevalence rate of HP infection among men, potentially contributing to the male predominance of GC.

1.3.2 Aetiology and pathogenesis

GC is a heterogenous and multifactorial disease (35,42). In 90% of cases, GC occurs sporadic, whereas in 10% of cases it occurs within a familial or hereditary setting (35). Hereditary diffuse GC can be caused by a germline mutation of the E-Cadherin-(CDH1)-gen (37). The pathogenesis of sporadic diffuse GC is currently unclear.

HP infection which leads to chronic atrophic gastritis in the antrum and consequently spreads to the corpus, causing atrophic gastritis in both antrum and corpus is the major risk factor for developing intestinal-type, non-cardia GC (2,13). The Correa's cascade describes the development of intestinal-type GC via an inflammation-atrophy-metaplasia-dysplasia-sequence (43,44). Although HP infection is the starting point of this precancerous sequence, not everyone infected with HP also develops GC (4,43). The severity of inflammation and its outcome depend on environmental factors (Table 2), on the infecting HP strain and its virulence factors as well as on the host genetics' susceptibility (4,43).

Table 2: **Environmental agents with sufficient and limited evidence of gastric carcinogenity**. Modified after Carneiro et al. (35).

Agents with sufficient evidence of gastric carcinogenicity	• *Helicobacter pylori* infection • Rubber manufacturing industry • Tobacco smoking • X-radiation and γ-radiation
Agents with limited evidence of gastric carcinogenity	• Asbestos • EBV infection • Lead compounds, inorganic • Nitrate or nitrite (ingested) under conditions that result in endogenous nitrosation • Pickled vegetables (traditional Asian) • Salted fish, Chinese-style • Consumption of processed meat

1.3.2.1 Environmental factors

Environmental agents that might contribute to the development of GC are presented in Table 2. Excessive consumption of salty foods damages the mucosa, makes it more vulnerable to HP and induces inflammatory changes (2). Within the process of reparation, excessive cell replication can lead to an increased rate of endogenous mutations (2). The prolonged inflammatory process may result in atrophy (2). The associated loss of parietal cells leads to higher pH levels, promoting the growth of anaerobic bacteria (2). Anaerobic bacteria can reduce dietary nitrite to nitrate. Nitrate reacts with nitrogen-containing compounds in the gastric juice and forms genotoxic N-nitroso agents (2,44). In addition to environmental influences, endogenous factors such as activated macrophages and polymorphonuclear leucocytes are also involved in reactions that generate mutagenic nitric oxide (NO) (2). The mutagenic potential of NO might induce mutations of the tumour suppressor gene TP53 (2). Nitrosoindoles, which are contained in Chinese cabbage or fava beans, might promote carcinogenesis by inducing the ornithine decarboxylase (2,44).

The transitional process from atrophy to complete GIM is promoted by the loss of protective factors, that are, the intake and mucosal secretion of ascorbic acid,

an important antioxidant (2). Therefore, a low intake of fresh fruit and vegetables, which include ascorbic acid among other natural inhibitors of carcinogenesis, is regarded as risk factor of GC (2). In advanced stages of the precancerous cascade, carotenoids such as ß-carotenes function as free radical scavengers (2). They may reduce the effect of carcinogens and prevent the development of dysplasia and early GC (2). Excessive salt intake, however, may additionally increase the mutagenic potential of nitrosated food at this stage (2). Other irritants that contribute to the precancerous process include the ingestion of aspirin, tobacco smoking, alcohol consumption and chronic bile reflux (2,43).

1.3.2.2 HP strains and virulence factors

Infection with cytotoxin associated antigen-(cag-A)-positive vacuolating toxin (vac-A) s1m1 HP strains are associated with the development of precancerous lesions and GC (4,43). Infection with cag-negative vac-A s2m2 HP strains can lead to persistent non-atrophic gastritis and is not associated with an increased risk for GC (43). Non-atrophic HP gastritis can either stay non-atrophic or enter the precancerous cascade: non-atrophic gastritis → multifocal atrophic gastritis → complete GIM → incomplete GIM → dysplasia → GC (43). Non-atrophic gastritis can be cured by HP eradication therapy or may lead to duodenal ulcer disease (43).

1.3.2.3 Host genetics

In addition to environmental agents, the hosts' genetic susceptibility plays a role in the development of GC (45). Single nucleotide polymorphisms (SNPs) in immune-related genes such as interleukin (IL)-1b, IL-1RC, tumour necrosis factor (TNF)-alpha and IL-10 showed an increased risk for developing GC among HP infected individuals (45).

On a molecular level certain events take place that contribute to different stages in the pathogenesis of intestinal-type GC (45). These include somatically acquired gene mutations, somatic copy number alterations, structural variants, and changes in epigenetics and transcription (45). Due to the advance in discovering the genetic and molecular pathogenesis of GC, new GC classification systems based on genetics and molecular features have been established (35). The Cancer Genome Atlas (TCGA) Research Network (46) proposed a classification of GC including the following four subtypes: EBV-positive, microsatellite-instable (MSI),

genomically stable and chromosomally unstable GC. The Asian Cancer Research Group (ACRG) (47) classified GC into the following subtypes: MSI GC, microsatellite-stable (MSS) with epithelial-mesenchymal transition gene signature GC, MSS and TP53-active GC, and MSS and TP53-inactive GC. Although molecular profiling gains on popularity, the clinical management of GC mostly still relies on traditional histology and the Tumor Node Metastasis (TNM) staging system (4,35).

1.3.3 Clinical features

Whereas early stages of GC may remain unnoticed by the patient, common symptoms of advanced stages include dysphagia, vomiting, anaemia, gastrointestinal bleeding, weight loss, fatigue, asthenia, indigestion, early satisfaction of appetite or chest burn (48). Diagnosis of GC is made by endoscopic examination and forceps biopsy. Abnormalities of the mucosa including depressions, nodules or amorphous surfaces are suspicious for invasive neoplastic processes (49). Besides forceps biopsy and excision of superficial lesions by endoscopic submucosal dissection (ESD) and endoscopic mucosal resection (EMR) for treatment and histological assessment (4,50), endoscopic ultrasound, computed tomography (CT) with contrast agents, magnet resonance imaging or fluorodeoxyglucose-positron emission tomography-CT may be used for staging (35).

1.3.4 Macroscopy

1.3.4.1 Atrophy and metaplasia

Macroscopically, atrophic gastritis shows a thin lined smooth mucosa with the loss of rugal folds. Severe atrophy features a translucent appearing mucosa with prominent veins in the submucosa shining through (51). GIM does not show any specific findings in endoscopy, gross features resemble the underlying disorder, whereas GIM usually occurs in an atrophic stomach (51).

1.3.4.2 Dysplasia

Dysplasia presents as a polypoid, flat, or depressed lesion in endoscopic examination that can be accompanied by erosions, ulcers or changes in colour

tones (52). Polypoid, elevated, or flat dysplastic lesions may also be referred to as intestinal-type gastric adenoma (52).

1.3.4.3 Gastric cancer

Early GC can be macroscopically classified into three types: (1) protruding, (2) superficial (elevated, flat, or depressed) and (3) excavated GC. Protruding carcinomas are defined as polypoid lesions protruding more than 3 millimetres whereas excavated lesions show a deep depression (35). Advanced GC can be macroscopically classified as polypoid mass, ulcerative tumour, infiltrative-ulcerative tumour, or diffuse infiltrative tumour (35,53).

1.3.5 Histology

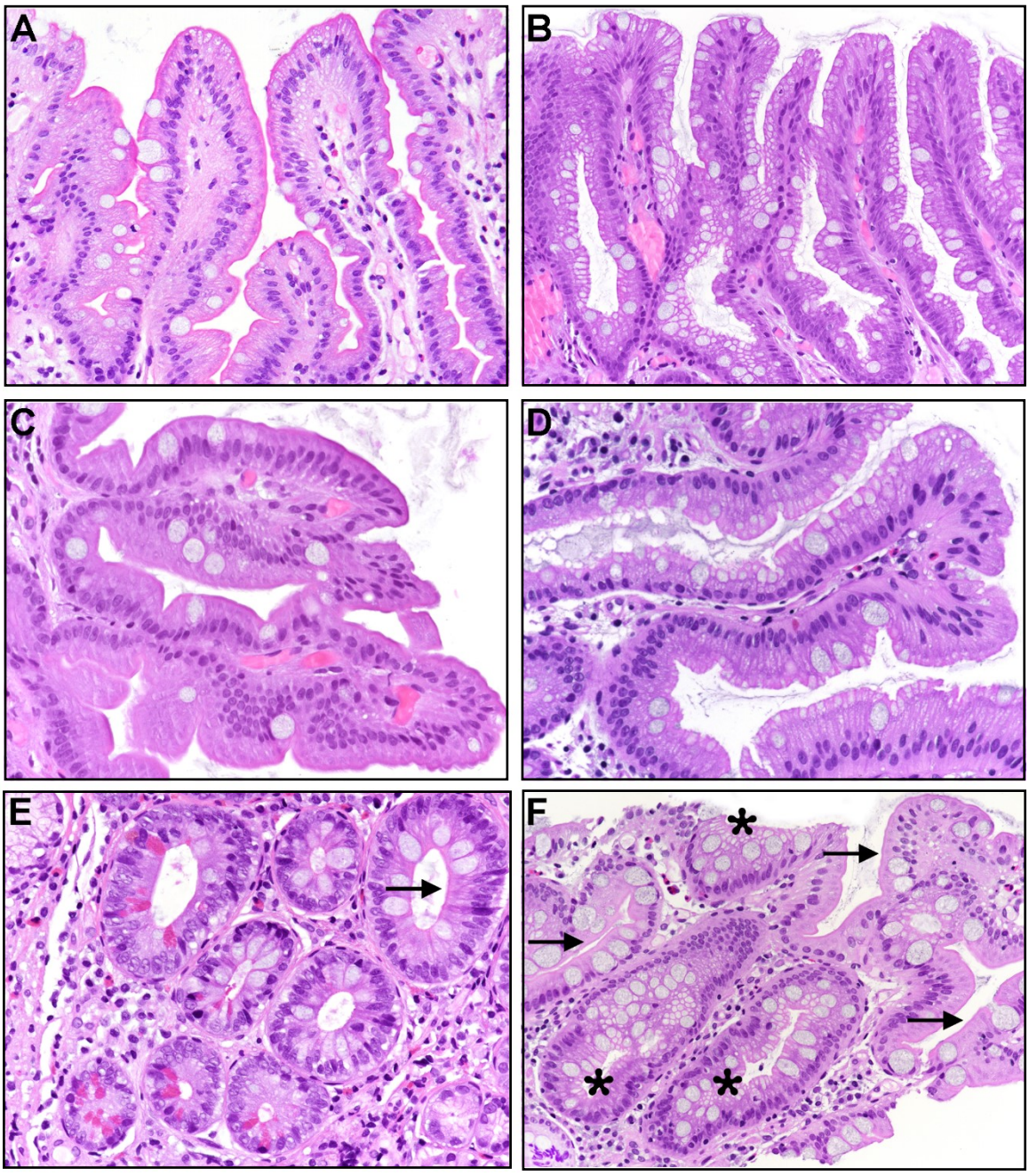
1.3.5.1 Atrophy and metaplasia

Gastric atrophy is characterized by the loss of the native glands in the gastric mucosa, that are, oxyntic or mucus secreting glands (54). Gastric atrophy can be either followed by fibrosis of the lamina propria or by metaplastic transformation, i.e., intestinal metaplasia, pseudopyloric metaplasia or pancreatic metaplasia (43,54,55). This metaplastic transformation evolves gradually and can be illustrated by the changing mucin secretion and enzyme production of the transforming cell (43,44). The term “complete GIM” emerged from the observation, that complete GIM produces the same, “complete”, range of digestive enzymes like cells from the small intestine do, i.e., sucrase, trehalose, leucine aminopeptidase and alkaline phosphatase. However, these digestive enzymes are vanishing with progression to “incomplete” GIM with sucrase disappearing at first and leucin aminopeptidase at last (44). Normal gastric epithelial cells secret neutral mucins, which are stained magenta with Periodic Acid Schiff (PAS) (43,56). Mucin core protein MUC5AC can be found at the surface and MUC6 in deeper glands of the gastric epithelium (43,56). Cells with an intestinal phenotype secrete neutral mucins in combination with sialomucins, .e.g., MUC2, or sialomucins alone (43,56). Sialomucins are stained blue with Alcian-Blue (AB) at pH 2.4 (56). In incomplete or colonic-type GIM, cells secret sulfomucins, which can be stained brown with high-iron diamine (HID)/AB (56-58). Based on mucin staining patterns, Jass et Filipe (58) differentiated GIM into three types, that is, type I as complete and types II and III as incomplete GIM.

However, the traditional approach to differentiate GIM requires sulfomucin or sialomucin staining, which is not performed in routine diagnostics (56,57).

Two types of GIM can be differentiated on haematoxylin and eosin (H&E)-stained slides (Figure 1): Complete GIM shows well-developed goblet cells admixed with eosinophilic absorptive enterocytes with brush borders at the surface and/or Paneth cells at the base of the crypts (56,57). Incomplete GIM lacks absorptive enterocytes and Paneth cells, but instead shows goblet cells with variable size admixed with mucin-secreting columnar cells without brush border (56,57). Mixed intestinal metaplasia can occur within a single biopsy (56).

Figure 1: Subtypes of gastric intestinal metaplasia (GIM). (**A**) Complete GIM is characterized by eosinophilic enterocytes with well-defined brush border admixed with well-formed goblet cells (H&E, original x100). (**B**) Incomplete GIM shows goblet cells of variable size and intervening mucin-secreting columnar cells (gastric foveolar cells) without brush border (H&E, original x100). (**C**) Metaplastic enterocytes and goblet cells in complete GIM at higher magnification (H&E, original x200). (**D**) Goblet cells and intervening mucin-secreting columnar cells in incomplete GIM at higher magnification (H&E, original x200). (**E**) Paneth cells are frequently observed at the base of the crypts in complete GIM, note enterocytes with brush border (arrow; H&E, original x150). (**F**) Mixed type with characteristics of complete (arrow) and incomplete (asterisk) subtypes within a single biopsy (H&E, original x100).



1.3.5.2 Dysplasia

Gastric dysplasia represents the advanced stage in the gastric precancerous cascade (43,44). It is defined as a neoplastic lesion within the gastric epithelium without evidence of stromal invasion (54). On a histological level, the WHO classifies gastric dysplasia based on its phenotype into five categories: (1) negative for dysplasia/intraepithelial neoplasia (IEN), (2) indefinite for dysplasia/IEN, (3) low-grade dysplasia/IEN (low-grade adenoma), (4) high-grade dysplasia/IEN (high-grade adenoma) and (5) intramucosal invasive carcinoma (52). Category 2 “Indefinite for dysplasia” is used for biopsy specimens that lack important histological features for establishing the diagnosis of a neoplastic or dysplastic lesion, mostly due to reactive or regenerative changes within severe inflammation (52). Category 5 “Intramucosal adenocarcinoma” refers to neoplastic phenotypes that show invasion of the lamina propria or muscularis mucosae (52). These phenotypes feature distinctive structural abnormalities, including glandular crowding, excessive branching and budding, necrotic debris within glands, infiltration of single cells or irregular fusion of glands (52).

The grading of dysplasia determines the GC risk (59). In a prospective study, 15% of the patients with low-grade dysplasia showed progression, whereas 69% of the patients with high-grade dysplasia developed invasive GC within 12 months of follow-up (59). Grading is based upon nuclear atypia, mitotic activity, disruption of architecture and cytoplasmic differentiation (52). Low-grade dysplasia shows minimal disruption of mucosal architecture, mild to moderate atypia and elongated (intestinal-type dysplasia) or round to oval (foveolar-type dysplasia), hyperchromatic nuclei (52). Atypical nuclei are located at the basement membrane and polarity of cells is maintained (52). High-grade dysplasia shows marked disruption of mucosal architecture with marked distortion of glands, formation of back-to-back glands and prominent atypia with enlarged, amphophilic nuclei (52). Atypical nuclei extend to the luminal surface and there is loss of nuclear polarity, a high nuclear to cytoplasm ratio, frequent mitosis, and overexpression of P53 (52).

1.3.5.3 Histological subtypes of gastric cancer

Several classification systems have been developed for the heterogeneous histological phenotypes of GC (35). Two widely applied systems are the Lauren (36) and the WHO classification (35). The Lauren classification (36) is based upon

pathological, epidemiological, and aetiologic features of GC (37), whereas the WHO classification is based solely on the histomorphological phenotype of GC (35) (Table 3). Intestinal-type GC is characterized by cohesive malignant epithelial cells, forming tubules or glands, and is often associated the concomitant presence of GIM (36). It mainly occurs in elderly and male, is located in the antrum and is associated with a longer course of disease and a better prognosis (60,61).

Diffuse GC is characterized by non-cohesive, scattered malignant cells with poor differentiation and a diffuse, highly infiltrative growth (62). It usually occurs in younger patients and is equally distributed between male and females (62). It is mainly located in the corpus and is associated with a shorter duration of disease and a worse prognosis (61,63).

Table 3: Lauren (36) **and WHO classification of gastric cancer**. Modified after Carneiro et al. (35)

Lauren (1965) (36)	**WHO (2019)** (35)
Intestinal	Papillary Tubular, well-differentiated Tubular, moderately differentiated
Indeterminate	Tubular (Solid), poorly differentiated
Diffuse	Poorly cohesive, signet-ring cell phenotype Poorly cohesive, other types
Intestinal/diffuse/indeterminate	Mucinous
Mixed	Mixed
Not defined	Other histological subtypes: adenosquamous carcinoma; squamous cell carcinoma; undifferentiated carcinoma; carcinoma with lymphoid stroma; hepatoid carcinoma; adenocarcinoma with enteroblastic differentiation; adenocarcinoma of fundic gland type micropapillary adenocarcinoma

1.3.6 Grading and staging

Tubular and papillary gastric tumours can be graded into G1 (well differentiated), G2 (moderately differentiated) and G3 (poorly differentiated) adenocarcinomas (35,64). Focusing on a two-tiered histopathological grading system, G1 and G2

carcinomas can be classified as low-grade GC while G3 carcinomas can be classified as high-grade GC (35). Low-grade GC resembles well-formed glands, whereas high-grade GC resembles poorly differentiated glands along with solid components or individual cells (35). For prognosis prediction the eight edition of the Union for International Cancer Control (UICC) TNM staging system is used (64) (Table 4-5). Tumours at the gastro-oesophageal junction with an epicentre more than 2 cm into the proximal stomach are staged as GC, whereas tumours ≤2 cm are staged as esophageal cancer (64).

Table 4: The pathological Tumor Node Metastasis (pTNM) staging system. Modified after the Union for International Cancer Control (64)**.** Histological examination of regional lymphadenectomy should include 16 or more lymph nodes. If the number of lymph nodes is <16 and lymph nodes are negative, the specimen should be classified as pN0.

pTX	Primary tumour cannot be assessed
pT0	No evidence of primary tumour
pTis	Carcinoma in situ: intraepithelial tumour without invasion of lamina propria, high-grade dysplasia
pT1a pT1b	Tumour invades lamina propria or muscularis mucosae Tumour invades submucosa
pT2	Tumour invades muscularis propria
pT3	Tumour invades subserosa
pT4a pT4b	Tumour perforates serosa Tumour invades adjacent structures (spleen, kidney, adrenal gland, small intestine, transverse colon, abdominal wall, retroperitoneum, liver, pancreas, diaphragm)

NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1 to 2 regional lymph nodes
N2	Metastasis in 3 to 6 regional lymph nodes
N3a N3b	Metastasis in 7 to 15 regional lymph nodes Metastasis in 16 or more regional lymph nodes

Table 5: Pathological stage. Modified after the Union for International Cancer Control (64).

Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T1	N1	M0
	T2	N0	M0
Stage IIA	T1	N2	M0
	T2	N1	M0
	T3	N0	M0
Stage IIB	T1	N3a	M0
	T2	N2	M0
	T3	N1	M0
	T4a	N0	M0
Stage IIIA	T2	N3a	M0
	T3	N2	M0
	T4a	N1, N2	M0
	T4b	N0	M0
Stage IIIB	T1, T2	N3b	M0
	T3, T4a	N3a	M0
	T4b	N1, N2	M0
Stage IIIC	T3, T4a	N3b	M0
	T4b	N3a, N3b	M0
Stage IV	Any T	Any N	M1

1.3.7 Prognosis

The pTNM staging system (64) is used as main tool for prognosis prediction in GC patients (35) (Table 4-5). If neoplasms are small, non-ulcerated, low-grade on histology and staged as pT1a, endoscopic treatment with ESD or EMR can be performed (42,65). In GC staged >T1N0 macroscopic and microscopic R0 tumour

resection should be performed (66). Neoadjuvant/perioperative chemotherapy may improve survival of patients with clinically staged ≥T2 or N-positive (66). For advanced GC with distant metastatic spread, palliative systemic treatment is recommended (62).

Patients with pT1 and pN0 show 5-year survival rates of >90% (35,67). Survival rates are decreasing consistently with increasing pT and pN category (35,67). Patients with GC stage IIIC show 5-year survival rates of 20% (35,67). In high-incidence countries, prognosis has been improved due to the implementation of screening programs that enable the detection of GC at earlier stages (68). However, in Western countries, where incidence rates are low, more than half of GC patients are diagnosed at advanced stages with 5-year survival rates ranging from 20% to 30% (42)

1.4 Management of gastric precancerous lesions

1.4.1 Risk stratification

Risk stratification in patients with gastric precancerous lesions using staging systems may enable targeted endoscopic surveillance, early detection of GC, reduction of GC mortality rates, and cost-effectiveness in GC screening, especially in low-intermediate risk countries (7). Staging systems that assess the severity and topography of atrophy (OLGA) or GIM (OLGIM) have been established (Table 6). OLGA (15) and OLGIM (16) are built upon the visual analogue scales and sampling protocols introduced by the Sydney classification of gastritis (11). Patients diagnosed with gastritis OLGA/OLGIM III and IV bear a significantly higher risk for developing GC than patients diagnosed with low-risk OLGA/OLGIM stages I and II (69,70).

Several studies have been conducted to assess interobserver agreement for OLGA/OLGIM staging systems (69-71). The majority of studies showed poor agreement for atrophy and fair agreement for intestinal metaplasia (71). Therefore, Lee et al. (7) advocate using GIM instead of atrophy as a uniform staging parameter. Although the OLGA system holds lower interobserver agreement among general pathologists, it has a higher sensitivity in detecting individuals at risk of GC compared to the OLGIM system (69,70). Individuals who show severe atrophy without GIM are “downstaged” when using the OLGIM instead of the OLGA staging

system. Nevertheless, GIM is regarded as a more reliable and commonly recognized histological parameter (72). Therefore, some recommend using a combination of both systems in routine diagnostics (69,70).

Table 6: **The OLGA** (15) **and OLGIM** (16) **staging systems of gastritis**. OLGA relies on atrophy as a marker for disease progression, whereas OLGIM relies on GIM. Modified after Capelle et al. (16).

OLGA					
Antrum	Corpus				
	Atrophy score	No atrophy	Mild atrophy	Moderate atrophy	Severe atrophy
	No atrophy	Stage 0	Stage I	Stage II	Stage II
	Mild atrophy	Stage I	Stage I	Stage II	Stage III
	Moderate atrophy	Stage II	Stage II	Stage III	Stage IV
	Severe atrophy	Stage III	Stage III	Stage IV	Stage IV

OLGIM					
Antrum	Corpus				
	Gastric intestinal metaplasia (GIM) score	No GIM	Mild GIM	Moderate GIM	Severe GIM
	No GIM	Stage 0	Stage I	Stage II	Stage II
	Mild GIM	Stage I	Stage I	Stage II	Stage III
	Moderate GIM	Stage II	Stage II	Stage III	Stage IV
	Severe GIM	Stage III	Stage III	Stage IV	Stage IV

1.4.2 Neoplastic risk in population-based studies

A nationwide study from the Netherlands demonstrated that increasing severity of premalignant lesions (i.e., severe dysplasia), increased age (i.e., 75-84 years) and male gender were risk factors for the development of GC (73). Another study showed that GIM occurs more often in males, in elderly (over 50 years) and in individuals with persistent HP infection (74).

A nationwide study from Sweden reported that only 1 in 256 individuals with normal mucosa will develop GC within 20 years, while the incidence increases with every step in the Correa's cascade, ultimately resulting in the development of GC in 1 in 85 individuals with gastritis, in 1 in 50 individuals with atrophic gastritis, in 1 in 39 individuals with GIM, and in 1 in 19 individuals with dysplasia, respectively (75). GC incidence rates were highest in patients who have already shown dysplasia at index endoscopy and lowest in patients who showed non-atrophic gastritis at index endoscopy (75).

A population-based study from the USA, showed a significantly increased risk for the development of GC among the Hispanic ethnicity (76). Therefore, surveillance is recommended for this ethnicity, for elderly and for patients with extensive GIM (76).

For individuals with extensive GIM, an approximately 2-fold increased risk of neoplastic progression was reported compared to individuals with limited GIM in a meta-analysis based on two studies (77).

Studies have not only proven increased GC risk for severe and extensive GIM, but also for the incomplete subtype (78-87). According to a meta-analysis based on seven studies, incomplete GIM bears a 1.7-fold higher risk of progression to dysplasia and a 3.3-fold higher risk of cancer compared to complete GIM (77). The presence of incomplete GIM is known to increase with increasing extent of GIM (5,57,85,88). This association can partly be reflected by the increasing stages of OLGA/OLGIM (57). However, incomplete GIM can not only be detected in individuals with advanced OLGIM stages III and IV, but also in individuals with low-risk OLGIM stages, such as OLGIM I (84).

1.4.3 Guidelines

Surveillance of patients with advanced stages of GIM enhances early detection of GC and therefore improves prognosis and patients' survival (4,7,89,90).

Surveillance programs will not significantly increase the burden of care since these patients represent only the minority, especially in low- to intermediate-risk countries (91). However, appropriate prevention measures for patients at high-risk of GC may reduce the clinical burden caused by disease progression (91).

The European Society of Gastroenterology and Endoscopy (ESGE) established and recently revised guidelines on the Management of Patients with Precancerous Conditions in the Stomach (MAPS II) (4). Currently, no surveillance is recommended for patients with mild to moderate atrophy restricted to the antrum or with GIM restricted to a single location (4). If additional risk factors such as incomplete GIM, persistent HP infection or a family history of GC are present, surveillance endoscopy at a three-year interval is indicated (4). Additionally, extensive GIM, defined as GIM involving both antrum and corpus, warrants surveillance (4). Patients with GIM restricted to the antrum are of particular interest, as these would not qualify for surveillance based upon GIM extent alone. In these individuals the GIM subtype is the determining factor for follow-up endoscopy.

For patients with advanced stages of atrophic gastritis, i.e., OLGA/OLGIM III and IV, endoscopic follow-up at a three-year interval is recommended (4). Intensified follow-up, e.g., every one to two years, is recommended, if family history of GC is present (4). At current date, the ESGE does not recommend targeted management based on age, gender, HP virulence factors or host-genetic variations (4). Moreover, patients should ideally undergo high-definition-chromoendoscopy with guided biopsy or receive at least four non-targeted biopsies from the greater and lesser curvature of antrum and corpus (4). If OLGA/OLGIM staging systems are used for risk stratification, another biopsy from the incisura may be considered (4). Biopsies from different topographical sites should be clearly labelled in separate vials (4).

In contrast to the ESGE, the American Gastroenterological Association (AGA) does not state any specific recommendations regarding routine surveillance for patients with GIM but prompts individual risk assessment (90). The AGA does recognize incomplete GIM, extensive GIM and a family history of GC as indicators for patients with high GC risk, potentially benefiting from surveillance endoscopy (77,90). Furthermore, immigrants from high-risk GC regions (East Asia, South America) and racial/ethnic minorities (e.g., Hispanics, American Indian) in the USA bear a higher risk for GC (92) and may benefit from short interval repeat endoscopy

to determine anatomic extent and histologic subtype of GIM for risk stratification (90).

The different approaches in the guidelines from the ESGE and the AGA might result from the fact, that the AGA regards the lack of evidence from randomized controlled studies (RCTs) as a limitation, whereas the ESGE accepts evidence from long-term observational studies. Different recommendations may also result from the difference in American and European health care systems and the potential economic burden (91).

The British Society of Gastroenterology (BSG) also identifies patients with extensive GIM, i.e., GIM affecting both antrum and corpus, at risk and recommends surveillance at a three-year interval. Whereas the ESGE prefers the OLGIM system for staging chronic atrophic gastritis, the BSG does not recommend using OLGA/OLGIM staging systems due to lack of prognostic evidence. The BSG does not recommend routine subtyping, but rather relies on the association between extensive GIM and the presence of the incomplete subtype (89). However, incomplete GIM might also occur in low-risk stages OLGIM I and II (84) and in cases with limited GIM.

Both AGA and BSG recommend against the use of any biomarkers, further emphasizing the importance of the histological assessment for adequate risk stratification.

1.5 Aims

Endoscopic surveillance based on GIM subtyping was implemented in the updated MAPS-Guidelines of the ESGE to assign patients to different clinical pathways. GIM subtyping is also considered as a valuable tool in the recently released AGA Clinical Practical Guidelines on the management of GIM.

Despite these recent recommendations from clinicians, it has not been proven that pathologists can differentiate between the two GIM subtypes with sufficient agreement, because a systematic interobserver variability study is currently lacking.

Moreover, the ESGE, AGA and BSG recognizes the extent of GIM as the key marker for risk stratification. Apart from studies addressing the OLGIM staging system, interobserver variability in estimating the percental extent of GIM has not been assessed yet.

This diploma thesis aims to evaluate the interobserver variation in GIM subtyping involving an international group of gastrointestinal expert pathologists. Furthermore, this study addresses the impact of misclassification on patient management and the identification of potential influencing factors. As secondary objective, this diploma thesis aims to evaluate the interobserver variability in estimating the overall percental extent of GIM and to identify parameters with impact on the assessment.

2. Materials und methods

2.1 Cases

The study included antral biopsies from 46 patients with chronic atrophic gastritis, diagnosed at the Diagnostic- and Research- (D&F) Institute of Pathology, Medical University of Graz, Austria, within the period of 11/2019 to 03/2020. All biopsies had been obtained based on Sydney criteria (11), that is, targeting the lesser and greater curvature, excluding the normal gastroduodenal transitional mucosa (4). It may be of note that corpus and/or fundus biopsies, which had been submitted in separate vials, lacked GIM in all cases and were therefore not part of the evaluation.

Since Austria is a country with a low prevalence of HP, resulting in a low incidence of GIM in general (with a low portion of mucosal surface involved) and a predominance of complete GIM (93), we selected the study sample in order to enrich for cases with a high portion of mucosa involved by GIM and for cases with the incomplete subtype.

Table 7 shows case characteristics including the age, the mean extent of GIM and the number of antral biopsy pieces. Gender distribution, HP status, type of background inflammation and GIM subtypes are displayed in Figure 2-5.

Before starting the study, sample size calculation was performed by Dr. Josef Haas, Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz with power of 80% and significance level at 5%. All samples were routinely stained with H&E and scanned thereafter (Pannoramic 1000 Whole-Slide Scanner, 3D Histech Ltd., Budapest, Hungary).

Table 7: **Case characteristics (n=46).**

Age (years)	Mean	65.8
	Median	69
	Range	27-87
Portion of mucosal surface involved by intestinal metaplasia	Mean	31%
	Median	20%
	Range	10-90%
Number of biopsy pieces	Mean	2.7
	Median	2
	Range	1-6

Figure 2: Gender distribution (n=46).

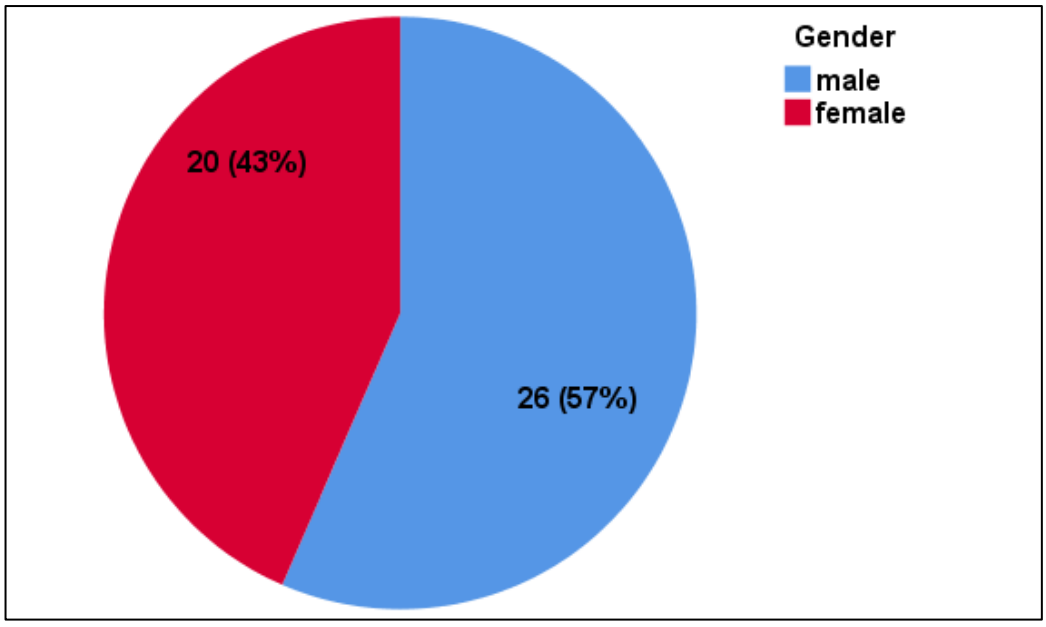


Figure 3: Helicobacter pylori status (n=46).

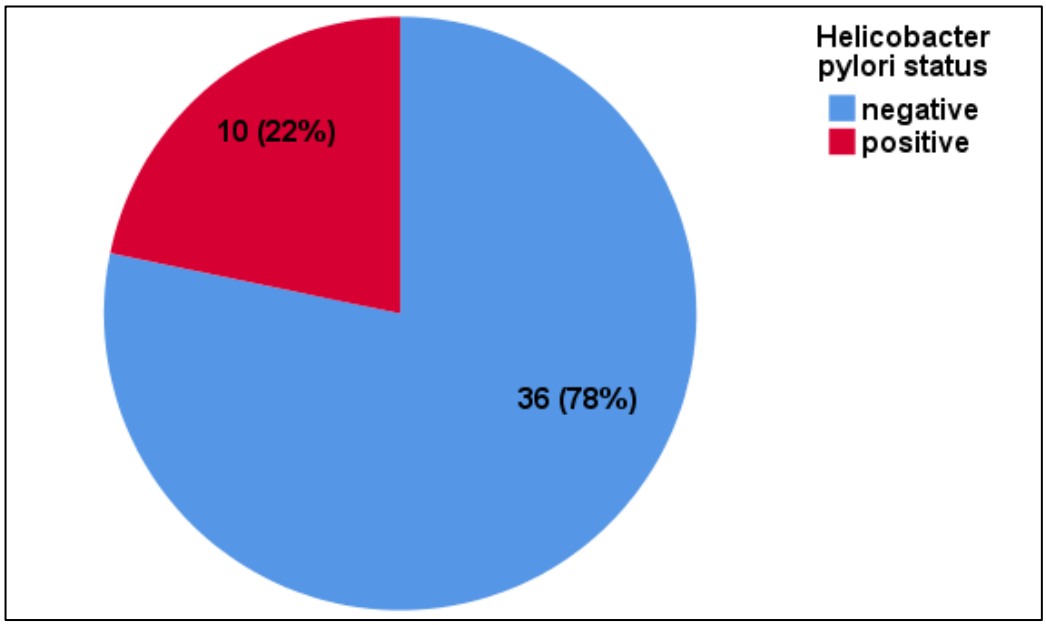


Figure 4: Background inflammation.

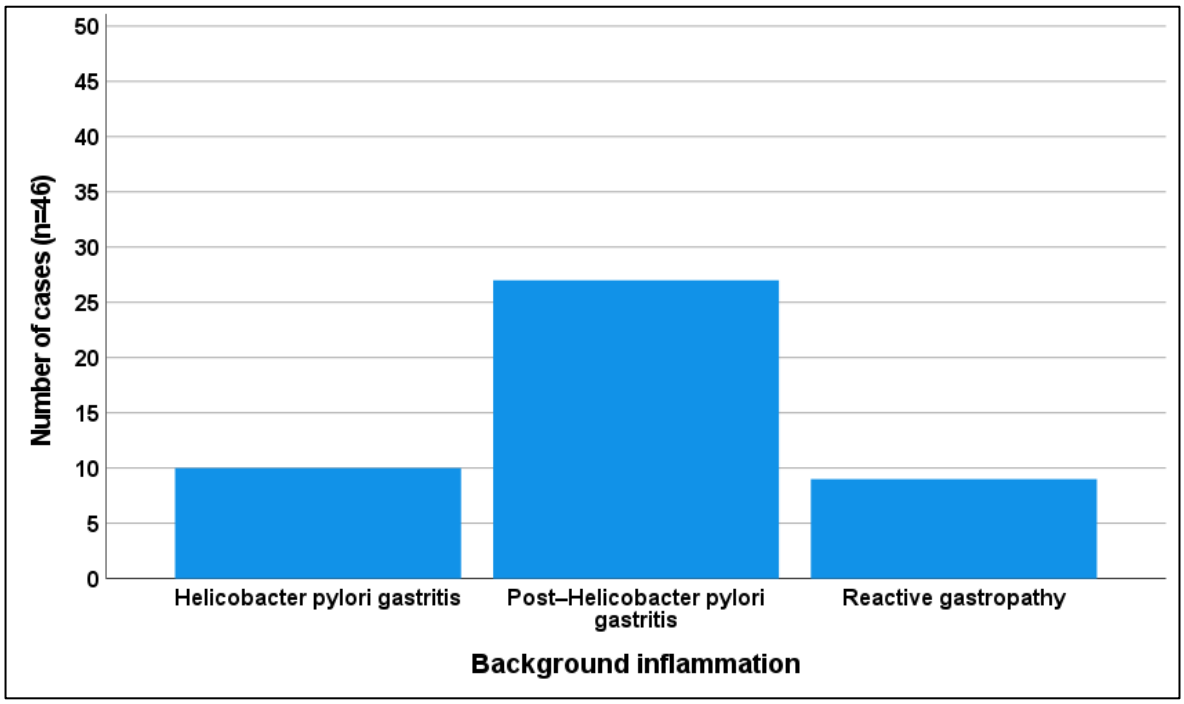
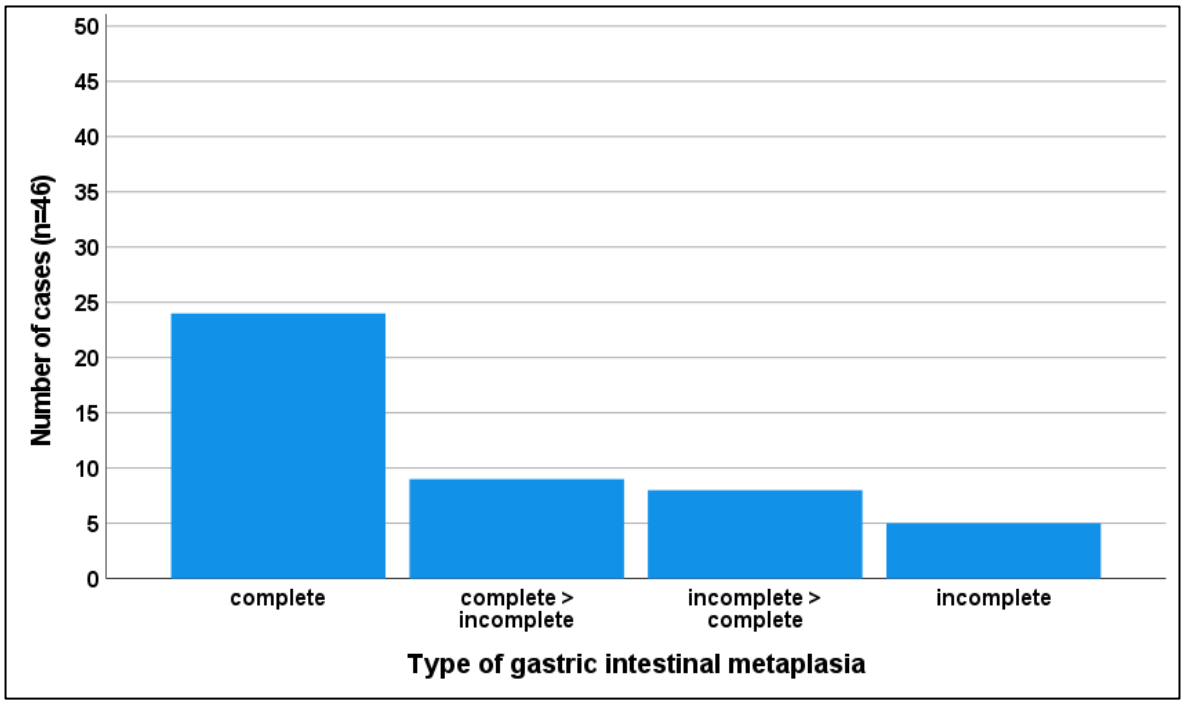


Figure 5: Type of gastric intestinal metaplasia.



2.2 Pathologists

Two gastrointestinal expert pathologists (Cord Langner and Gregory Y. Lauwers), who are used to apply GIM subtyping in routine diagnosis, independently viewed the scans, performed GIM subtyping, and classified the cases based on the morphological criteria into the following categories:

1. Complete GIM alone
2. Mixed, complete exceeding incomplete GIM
3. Mixed, incomplete exceeding complete GIM
4. Incomplete GIM alone

The two pathologists reached initial agreement in 38 cases out of 46 (83%). In cases of disagreement, a consensus (“gold standard”) was obtained by joint microscopy and case discussion. No consensus (“gold standard”) was established for the percental estimation of GIM extent.

Nine international gastrointestinal expert pathologists were invited to participate as observers in the study. Inclusion criteria were completed specialty training for pathology and proven publication activity in gastrointestinal pathology. For

evaluation of the interobserver variation in estimating the extent of GIM, all eleven pathologists were included as observers:

Gregory Y. Lauwers	H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida, USA.
Cord Langner	Diagnostic and Research Institute of Pathology, Diagnostic and Research Center for Molecular BioMedicine, Medical University of Graz, Graz, Austria.
Rish K. Pai	Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, Arizona, USA.
Ian Brown	Envoi Specialist Pathologists, Brisbane, Queensland, Australia. Faculty of Medicine, University of Queensland, Brisbane, Queensland, Australia.
Anthony J. Gill	Cancer Diagnosis and Pathology Group, Kolling Institute of Medical Research, Royal North Shore Hospital, St Leonards, New South Wales, Australia. New South Wales Health Pathology, Department of Anatomical Pathology, Royal North Shore Hospital, St Leonards, New South Wales, Australia. Sydney Medical School, University of Sydney, New South Wales, Australia.
Dhanpat Jain	Department of Pathology, Yale University School of Medicine, New Haven, Connecticut, USA.
Bence Kővári	Department of Pathology, University of Szeged, Szeged, Hungary. H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida, USA.
Ryoji Kushima	Department of Pathology, Shiga University of Medical Science, Shiga, Japan.
Kieran Sheahan	Department of Pathology & Centre for Colorectal Disease, St. Vincent's University Hospital, Dublin, Ireland. University College Dublin, Dublin, Ireland. University College Dublin, Dublin, Ireland.

Tomas Slavik	Ampath Pathology Laboratories, Pretoria, South Africa. Department of Anatomical Pathology, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa.
Amitabh Srivastava	Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts, USA.

Access to scanned slides was provided by an electronically transferred web link. The assessment was done independently in a blinded fashion on dynamic images (3D Histech Ltd. Case Viewer, Budapest, Hungary) with viewing between x2 and x40 magnification. Every observer classified each case applying the four categories mentioned above using a standardized evaluation sheet (see appendix). For evaluation of the extent of GIM, the pathologists were asked to estimate the overall percentage of mucosa involved by GIM in 5% increments, that is, across all biopsies included within a given sample

2.3 Ethics board approval

Institutional Review Board approval for this study was obtained from the Ethics Committee of the Medical University of Graz, Austria (EK 33-444 ex20/21).

2.4 Statistical analysis

Categorical variables are presented as absolute and relative frequencies, numerical variables as mean, median and ranges. Differences in categorical variables were examined using the chi-square test or Fisher's exact test, as appropriate. Differences in continuous variables between groups were examined using the t-test for two independent samples as parametric test procedure and the Mann-Whitney-U test as non-parametric test procedure. Parametric distribution was tested with the Shapiro-Wilk test. Interobserver variability in GIM subtyping was assessed by applying kappa statistics, which are used to quantify the degree of agreement beyond chance (94). The level of agreement between each observer and the consensus diagnosis was calculated using weighted kappa for all categories (1-4) as well as for combined categories focusing on the predominant GIM pattern (1 and 2 *versus* 3 and 4). Fleiss' kappa was used to calculate the agreement among the pathologists as follows: the scores from the nine observers and the original scores from the two pathologists who provided the consensus diagnosis were included in

this calculation (95). Kappa values were interpreted according to the scheme of Landis and Koch (96), modified by Altmann (97): kappa values <0.00 suggest no agreement, 0.00-0.20 poor agreement, 0.21-0.40 slight agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 almost perfect agreement, respectively.

The interobserver agreement in estimating the extent of GIM was assessed by applying the intraclass correlation coefficient (ICC), which is used to measure the degree of agreement for continuous variables for different observers when assessing the same cases (98). The calculation is based upon a two-way mixed model and absolute agreement. For interpretation, the scheme introduced by Altman (1991) (97) was used: an ICC value ≤0.20 suggests poor agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 good agreement and >0.80 very good agreement, respectively. Correlation between the standard deviation and the extent of GIM, defined by the mean score of GIM of the eleven observers, and the number of biopsy pieces per slide was evaluated by applying the Spearman correlation. Finally, a regression analysis was performed to establish a model that allows the prediction of the standard deviation from the extent of GIM. To account for non-consistent scattering in our dataset, we performed adjustment with heteroscedasticity-consistent standard error estimators (99).

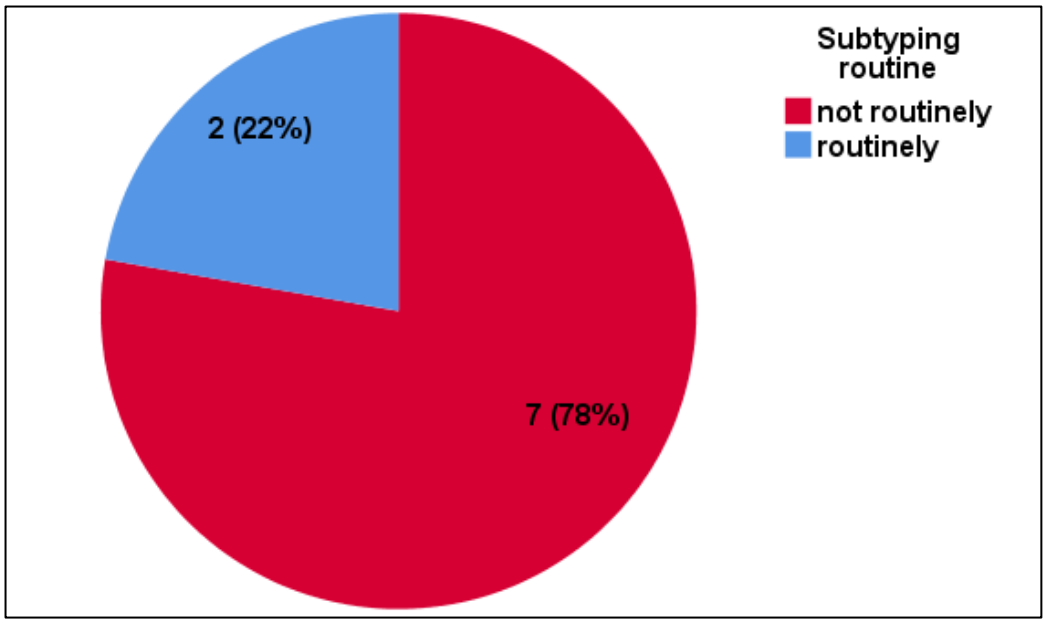
All statistical operations were performed using IBM SPSS Statistics Version 26, provided by the Medical University of Graz. P-values were two-sided, and values <0.05 were considered statistically significant.

3. Results

3.1 Pathologists' subtyping routine

Figure 6 shows the subtyping routine of nine pathologists, who participated in this study as observers. Two of the observers reported to apply GIM subtyping in their daily routine practice, the other seven did not.

Figure 6: Pathologists' subtyping routine.



3.2 Interobserver variability in GIM subtyping

Table 8 summarizes absolute and relative frequencies of the observers' assigned GIM subtypes.

Table 8: Absolute and relative frequencies of the observers' (OS) assigned gastric intestinal metaplasia subtypes in 46 cases.

Subtypes of gastric intestinal metaplasia (GIM); n (%)	OS #1	OS #2	OS #3	OS #4	OS #5	OS #6	OS #7	OS #8	OS #9
Category 1: pure complete GIM; 24 (52.2%)	24 (52.2%)	25 (54.3%)	17 (37.0%)	22 (47.8%)	20 (43.5%)	26 (56.5%)	24 (52.2%)	21 (45.7%)	18 (39.1%)
Category 2: Complete > incomplete GIM; 9 (19.6%)	8 (17.4%)	7 (15.2%)	13 (28.3%)	11 (23.9%)	15 (32.6%)	7 (15.2%)	8 (17.4%)	10 (21.7%)	14 (30.4%)
Category 3: Incomplete > complete GIM; 8 (17.4%)	4 (8.7%)	9 (19.6%)	3 (6.5%)	12 (26.1%)	7 (15.2%)	6 (13.0%)	11 (23.9%)	8 (17.4%)	8 (17.4%)
Category 4: Pure incomplete GIM. 5 (10.9%)	10 (21.7%)	5 (10.9%)	13 (28.3%)	1 (2.2%)	4 (8.7%)	7 (15.2%)	3 (6.5%)	7 (15.2%)	6 (13.0%)

Table 9 illustrates the interobserver variability between the nine observers and the consensus diagnosis. For four GIM categories, the interobserver variability ranged from slight to almost perfect agreement (kappa=0.368-0.961). Specifically, two observers reached almost perfect agreement with the consensus, five observers substantial agreement, two observers moderate agreement and one observer slight agreement, respectively. Diagnostic agreement between each observer and the consensus diagnosis ranged from 48% to 96%, with six observers reaching more than 70% agreement.

For two GIM categories, that is, addressing the predominant pattern (complete, 1 and 2, *versus* incomplete, 3 and 4), the interobserver variability ranged from moderate to almost perfect agreement (kappa=0.464-0.984). Specifically, five observers reached almost perfect agreement, two observers substantial agreement and another two observers moderate agreement, respectively. Diagnostic agreement between each observer and the consensus diagnosis ranged from 78% to 98%.

For four categories, the two observers who apply GIM subtyping in daily routine practice had a higher agreement and higher kappa values (p=0.040) than the other observers, while no significant difference was noted for two GIM categories (p=0.241).

Table 10 shows the interobserver variability among all eleven pathologists (the nine observers and the two pathologists who had agreed on the consensus diagnosis). For four categories, the overall agreement was moderate (Fleiss' kappa=0.447, 95% CI: 0.423-0.471). Applying two categories, the overall agreement was substantial (Fleiss' kappa=0.716, 95% CI: 0.677-0.755).

Table 9: Weighted kappa values (95% CI) and agreement (n; %) between the nine observers and the consensus diagnosis. The two observers who perform subtyping of gastric intestinal metaplasia (GIM) in daily routine practice are highlighted in grey.

Observer	Observer *versus* consensus (four GIM categories)	Observer *versus* consensus (two GIM categories, predominant pattern; 1 and 2 *versus* 3 and 4)	Agreement (four GIM categories)	Agreement (two GIM categories, predominant pattern; 1 and 2 *versus* 3 and 4)
#1	0.781 (0.664-0.899)	0.948 (0.846-1.049)	35 (76%)	45 (98%)
#2	0.961 (0.907-1.015)	0.948 (0.846-1.049)	44 (96%)	45 (98%)
#3	0.520 (0.393-0.678)	0.649 (0.415-0.884)	22 (48%)	39 (85%)
#4	0.368 (0.160-0.577)	0.464 (0.180-0.747)	22 (48%)	36 (78%)
#5	0.857 (0.756-0.959)	0.888 (0.736-1.039)	39 (85%)	44 (96%)
#6	0.769 (0.648-0.891)	0.786 (0.586-0.985)	34 (74%)	42 (91%)
#7	0.777 (0.657-0.897)	0.738 (0.523-0.953)	36 (78%)	41 (89%)
#8	0.793 (0.686-0.900)	0.898 (0.759-1.036)	35 (76%)	44 (96%)
#9	0.654 (0.502-0.806)	0.843 (0.671-1.014)	29 (63%)	43 (93%)

Table 10: Fleiss' kappa values (95% CI) including all pathologists (the nine observers and the two pathologists who produced the consensus diagnosis), who participated in subtyping gastric intestinal metaplasia (GIM).

	Overall	Per category
Four GIM categories	0.447 (0.423-0.471)	1 = 0.581 (0.542-0.620) 2 = 0.261 (0.222–0.300) 3 = 0.323 (0.284-0.362) 4 = 0.569 (0.530-0.608)
Two GIM categories (predominant pattern; 1 and 2 *versus* 3 and 4)	0.716 (0.677-0.755)	1 and 2 = 0.716 (0.677-0.755) 3 and 4 = 0.716 (0.677-0.755)

3.3 Factors with impact on agreement

Factors that may be associated with the level of agreement are illustrated in Table 11. Herein, we compared cases with the highest agreement (six cases with 100% agreement) with cases with the lowest agreement (one case with 22%, two cases with 33%, and five cases with 44% agreement). We identified three parameters with potential impact. Agreement was significantly lower (p=0.010) in mixed cases (categories 2 and 3) compared to cases with only one type of GIM (categories 1 and 4). Likewise, the number of biopsy specimens within the sample and the portion of mucosal surface involved by GIM (in particular when more than one biopsy piece was affected) were both higher in the cases with low agreement, but this difference was not statistically significant (p=0.886 and p=0.120, respectively), likely due to inherent small sample size.

Table 11: Factors with impact on the agreement between the observers and the consensus diagnosis in subtyping gastric intestinal metaplasia (GIM), illustrated by cases with highest and lowest agreement (n; %).

		Cases (n=6) with highest agreement	Cases (n=8) with lowest agreement	p-value
GIM categories	1 and 4 (pure GIM types)	6 (100%)	2 (20%)	0.010
	2 and 3 (mixed GIM types)	0 (0%)	6 (80%)	
Mean number of biopsy pieces		2.33	2.75	0.886
Mean portion of mucosal surface involved by intestinal metaplasia		15%	34%	0.120

3.4 Clinical consequences of misclassification

The updated MAPS Guidelines (4) recommend endoscopic surveillance for all patients with incomplete GIM, while patients with complete GIM do not need follow-up when metaplastic change is restricted to the antrum. Since all cases of our study cohort, who were biopsied according to the Sydney criteria (11), show metaplastic changes restricted to the antrum (limited GIM), no surveillance endoscopy would have been necessary based on the extent of GIM. However, in 13 out of 46 cases (28%) of our preselected study cohort incomplete intestinal metaplasia is present and hereby the sole determinant for necessary surveillance endoscopy. The recommendation regarding incomplete GIM is based upon the publications by Gonzalez et al. (82,83,85), who defined incomplete GIM as either pure incomplete GIM or incomplete GIM as predominant pattern in mixed cases (82,83,85).

Thus, patients with pure complete GIM (category 1) may be “overreported” when misclassified as predominantly incomplete or pure incomplete GIM (category 3 or 4), thereby inducing unnecessary follow-up investigations. On the contrary, patients with predominantly incomplete or pure incomplete GIM (category 3 or 4) may be lost to follow-up when misclassified and thereby “underreported” as pure

complete GIM (category 1). The definitive clinical significance of category 2 (incomplete GIM as a minor pattern) still needs to be elucidated.

Table 12 illustrates the clinical consequences of misclassification in our study. Overall, misclassification that might induce an erroneous follow-up decision occurred in 19 out of 333 (5.7%) ratings, with “underreporting” (n=14) occurring more often than “overreporting” (n=5).

Table 12: Clinical consequences of misclassification of gastric intestinal metaplasia. According to the Guidelines of the European Society of Gastrointestinal Endoscopy (4) patients with complete GIM (Category 1) do not need follow-up when metaplastic change is restricted to the antrum. In contrast, surveillance is recommended for all patients with predominant (Category 3) or pure (Category 4) incomplete GIM. The definitive clinical significance of category 2 (incomplete GIM as a minor pattern) is currently unclear. Therefore, this category is not included within the table. Cases in which overreporting (n=5) or underreporting (n=14) has impact on clinical management are highlighted in grey.

Consensus diagnosis (gold standard)	Observers' assigned categories	Observer #1	Observer #2	Observer #3	Observer #4	Observer #5	Observer #6	Observer #7	Observer #8	Observer #9
Category 1 (n=24; endoscopic surveillance not recommended)	1: Complete GIM	21	21	14	15	20	22	23	22	15
	2: Mixed GIM (complete > incomplete)	3	3	10	6	4	2	0	4	8
	3: Mixed GIM (incomplete > complete)	0	0	0	3	0	0	1	0	1
	4: Incomplete GIM	0	0	0	0	0	0	0	0	0
Categories 3 and 4 (n=13; endoscopic surveillance recommended)	1: Complete GIM	0	0	1	3	0	0	0	0	0
	2: Mixed GIM (complete > incomplete)	0	0	1	2	2	2	2	0	1
	3: Mixed GIM (incomplete > complete)	4	8	1	7	7	4	8	7	6
	4: Incomplete GIM	9	5	10	1	4	7	3	6	6
Misclassification with potential impact on endoscopic surveillance		0/37 (0%)	0/37 (0%)	2/37 (5%)	8/37 (22%)	2/37 (5%)	2/37 (5%)	3/37 (8%)	0/37 (0%)	2/37 (5%)

3.5 Interobserver agreement in estimating the extent of GIM

The mucosa of the 46 cases was involved by GIM in different quantities, with an overall mean value of 29%, ranging from 6.8% to 92.7%. Mean values for individual observers ranged from 23.3% to 33%. The interobserver agreement of the eleven observers was very good, with an ICC value of 0.983 (95% CI: 0.975-0.990). Table 13 shows the interobserver correlation matrix of the eleven pathologists estimating the percental extent of GIM.

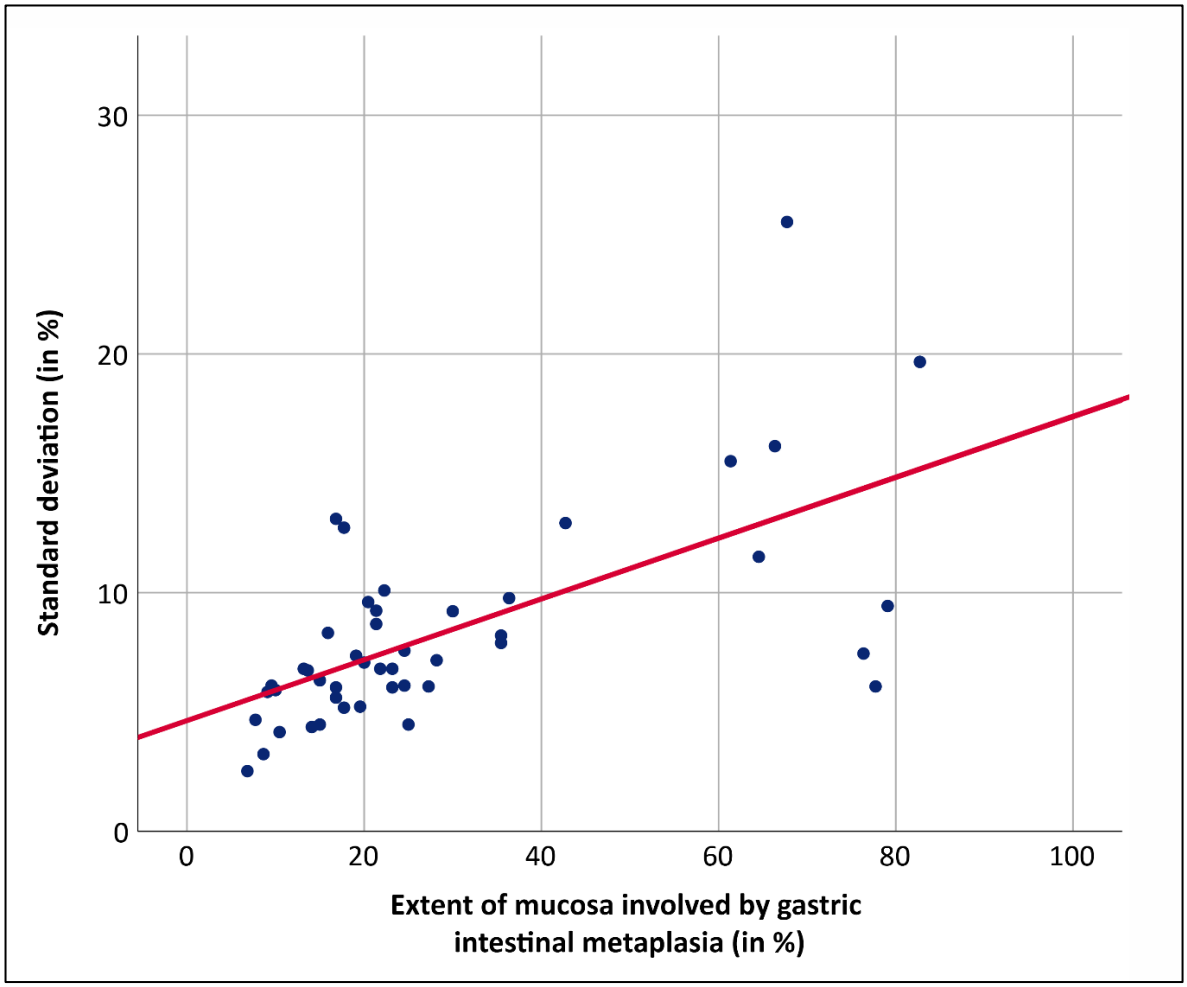
Table 13: Interobserver correlation matrix.

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11
#1	1.000	0.966	0.920	0.958	0.882	0.923	0.965	0.874	0.830	0.947	0.742
#2	0.966	1.000	0.927	0.953	0.871	0.944	0.949	0.901	0.817	0.952	0.717
#3	0.920	0.927	1.000	0.916	0.846	0.882	0.915	0.880	0.790	0.934	0.679
#4	0.958	0.953	0.916	1.000	0.869	0.902	0.934	0.894	0.778	0.924	0.749
#5	0.882	0.871	0.846	0.869	1.000	0.851	0.860	0.893	0.669	0.851	0.692
#6	0.923	0.944	0.882	0.902	0.851	1.000	0.930	0.866	0.802	0.911	0.691
#7	0.965	0.949	0.915	0.934	0.860	0.930	1.000	0.876	0.804	0.961	0.723
#8	0.874	0.901	0.880	0.894	0.893	0.866	0.876	1.000	0.725	0.901	0.677
#9	0.830	0.817	0.790	0.778	0.669	0.802	0.804	0.725	1.000	0.814	0.722
#10	0.947	0.952	0.934	0.924	0.851	0.911	0.961	0.901	0.814	1.000	0.674
#11	0.742	0.717	0.679	0.749	0.692	0.691	0.723	0.677	0.722	0.674	1.000

3.6 Factors with impact on the agreement

The number of biopsy pieces per slide ranged from 1 to 6 (mean 2.7, median 2). No correlation between the standard deviation and the number of biopsy pieces was observed ($p=0.059$). The six cases with the lowest standard deviation had a mean biopsy number of 3.0 (median 2.5, range 2-6) whereas the six cases with the highest standard deviation had a mean biopsy number of 2.3 (median 2.5, range 1-4). Cases with a higher amount of metaplastic epithelium had a higher standard deviation. The significant positive association between both parameters was verified by applying the Spearman correlation, $r_s=0.644$ ($p<0.01$). A simple linear regression was calculated to predict the standard deviation based upon the extent of GIM. According to the regression equation with $R2$ of 0.403, the standard deviation increased by 0.127 ($t=2.862$, $p<0.01$) for each percent of GIM extent. ($F(1,44)=29.749$, $p<0.01$; Figure 7).

Figure 7: Correlation between the extent of gastric intestinal metaplasia (GIM, in %) and the standard deviation (SD, in %): The SD increases with the amount of metaplastic epithelium identified on the mucosal surface (SD = 4.568 + 0.127 x extent GIM).



3.7 Histopathological assessment of cases

The 46 cases and their histopathological characteristics are presented in Figure 8-53. The bar charts summarize the nine observers' assigned GIM types. The consensus GIM type ("gold standard") is coloured in green and illustrates the absolute agreement per case (Figures 8a to 53a). Within the boxplots, each box shows the interquartile range (IQR) with lines representing the median of the eleven observers' scored percental extent of GIM (Figures 8b to 53b). Whiskers are extended to within 1.5 IQR of the first and third quartiles. Data points falling outside this range are displayed independently. The scanned H&E-stained slides show typical features of incomplete and complete GIM. The upper photograph was taken at lower magnification for overview of each slide (Figures 8c to 53c), the lower one at higher magnification for more precise assessment (Figures 8d to 53d).

Figure 8: Case 01. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

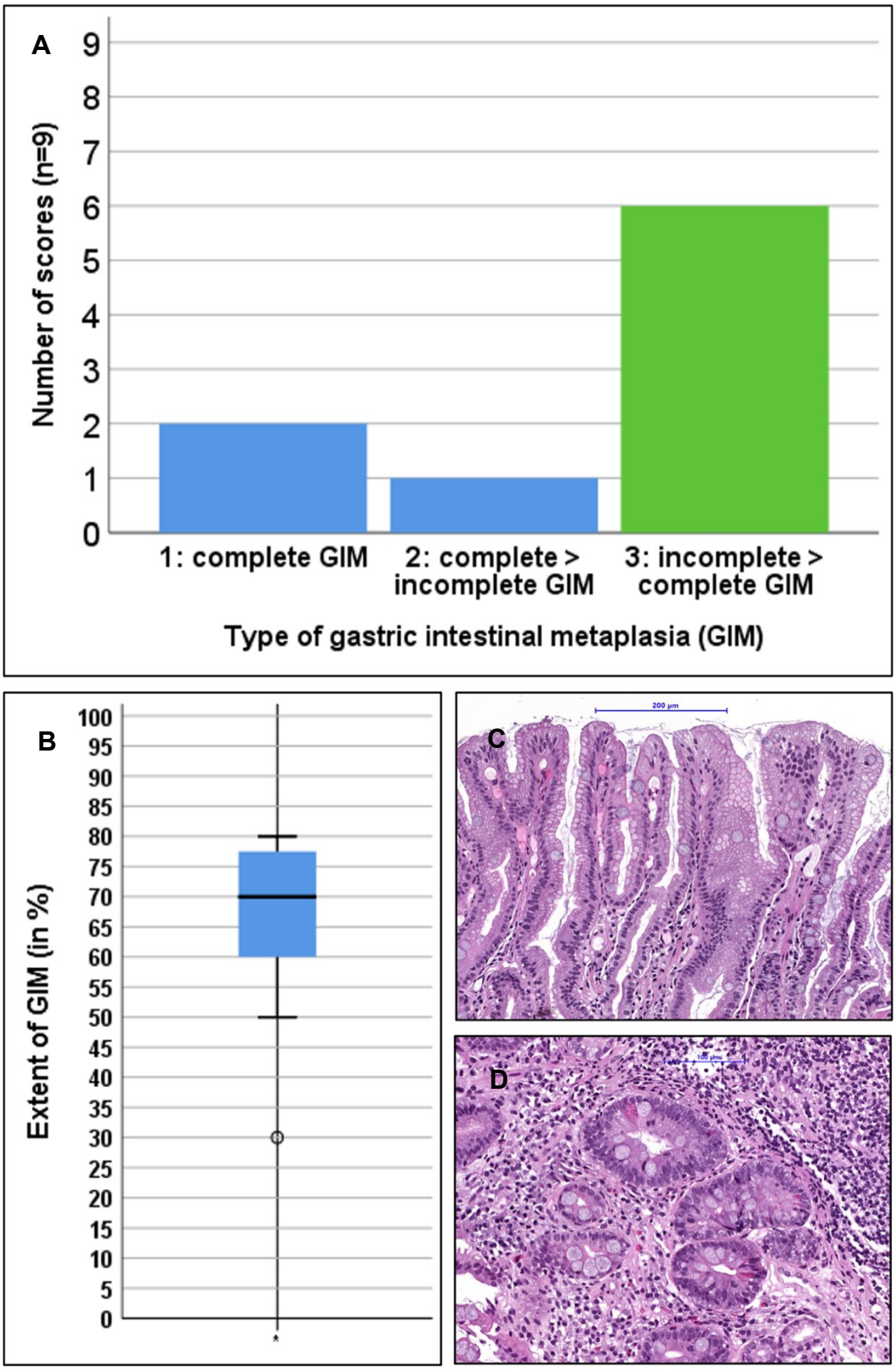


Figure 9: Case 03. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

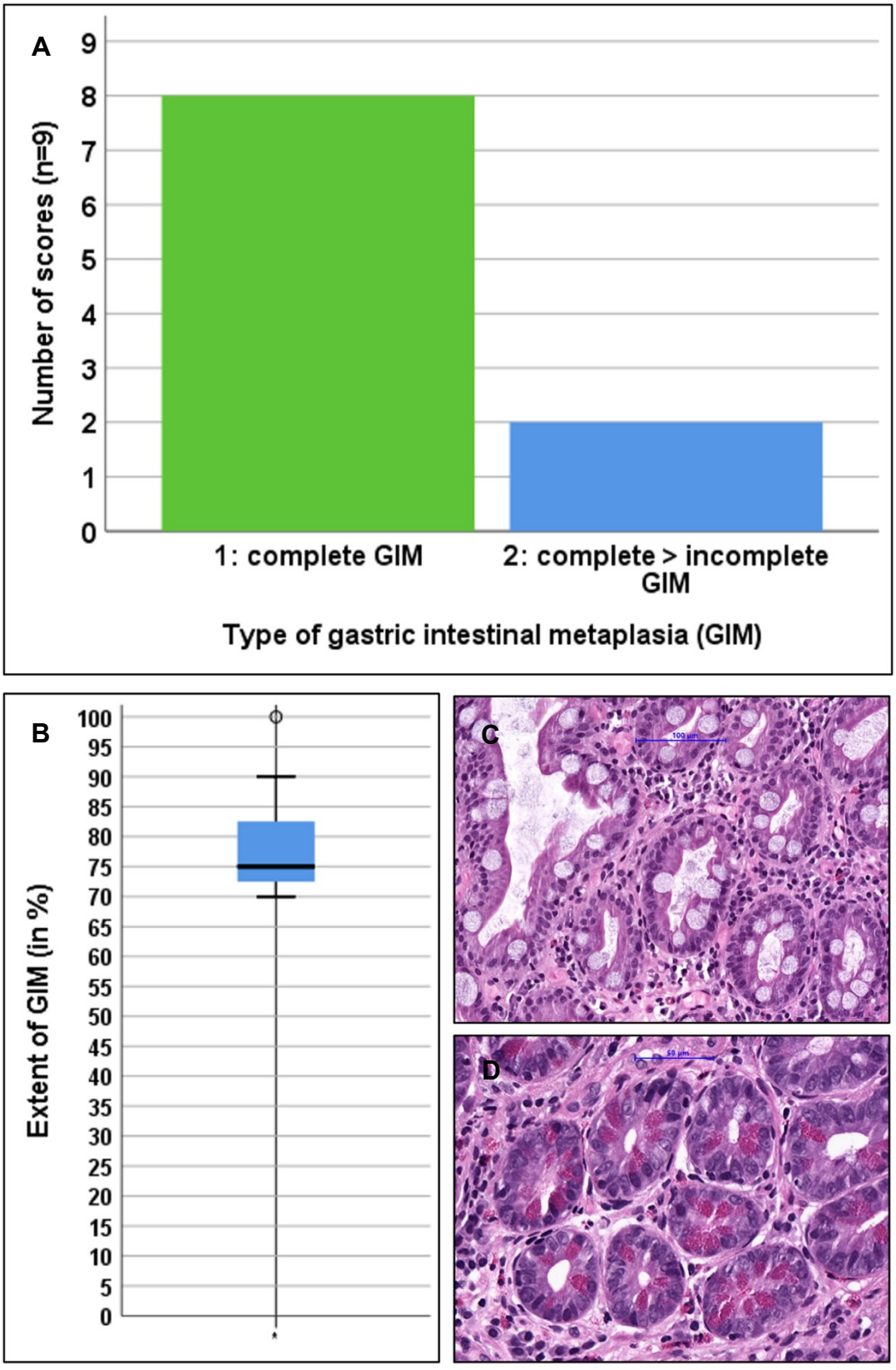


Figure 10: Case 03. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

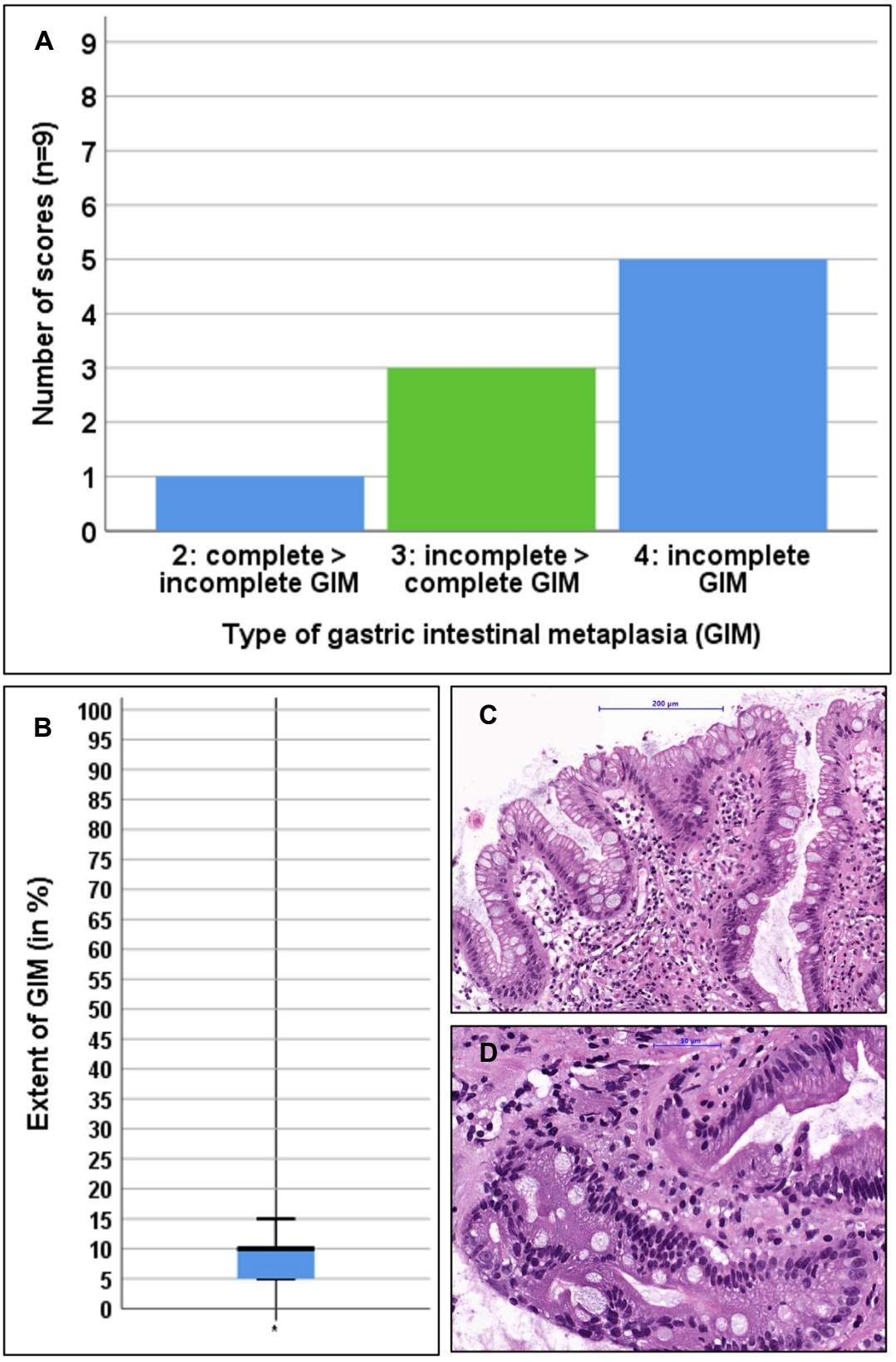
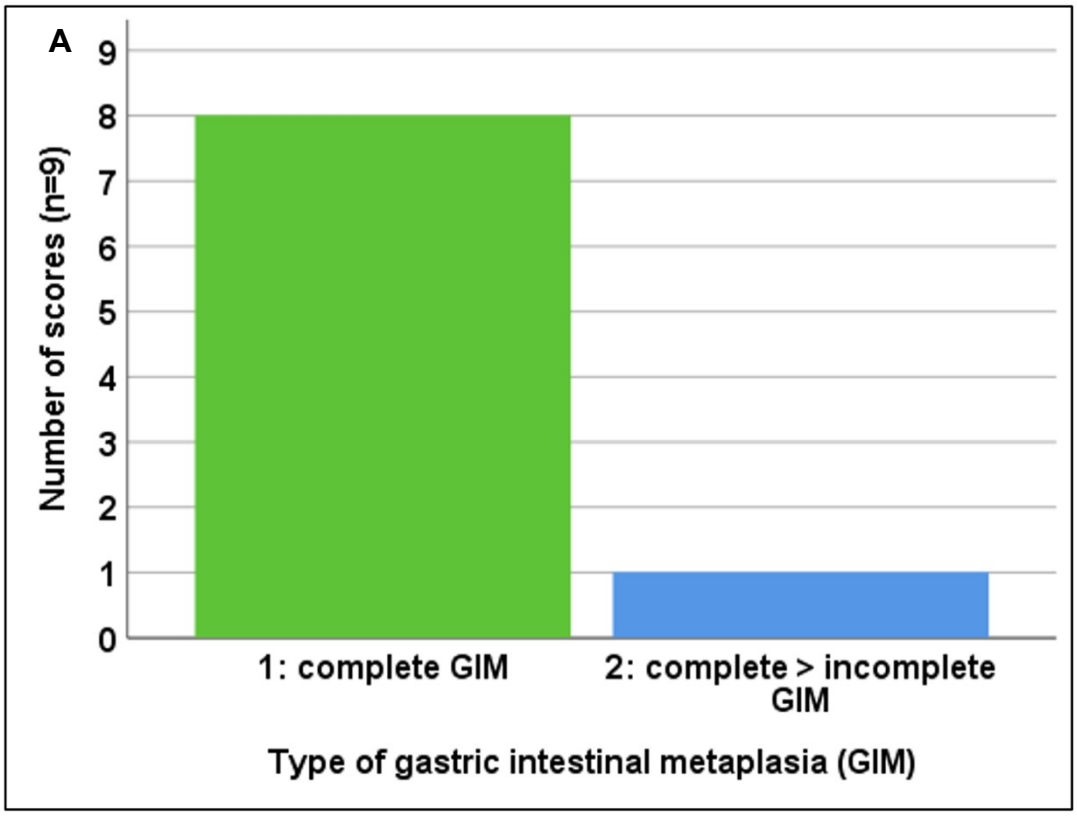
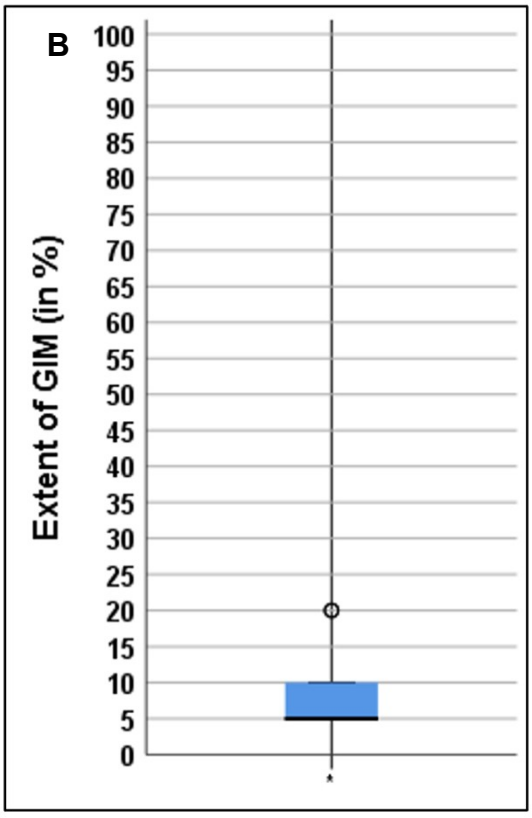
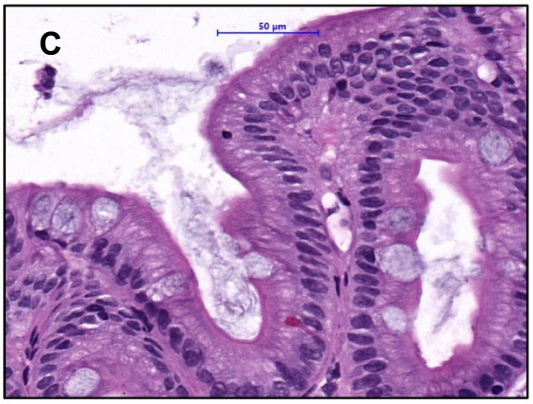


Figure 11: Case 04. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







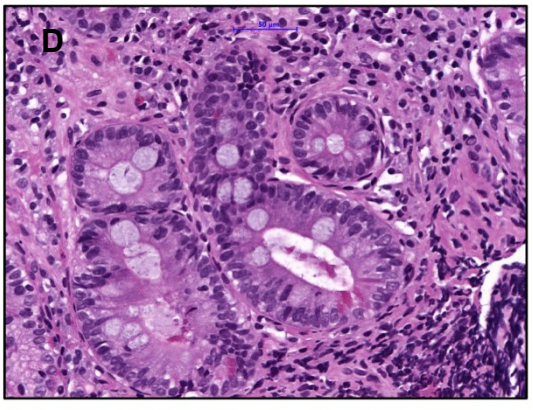
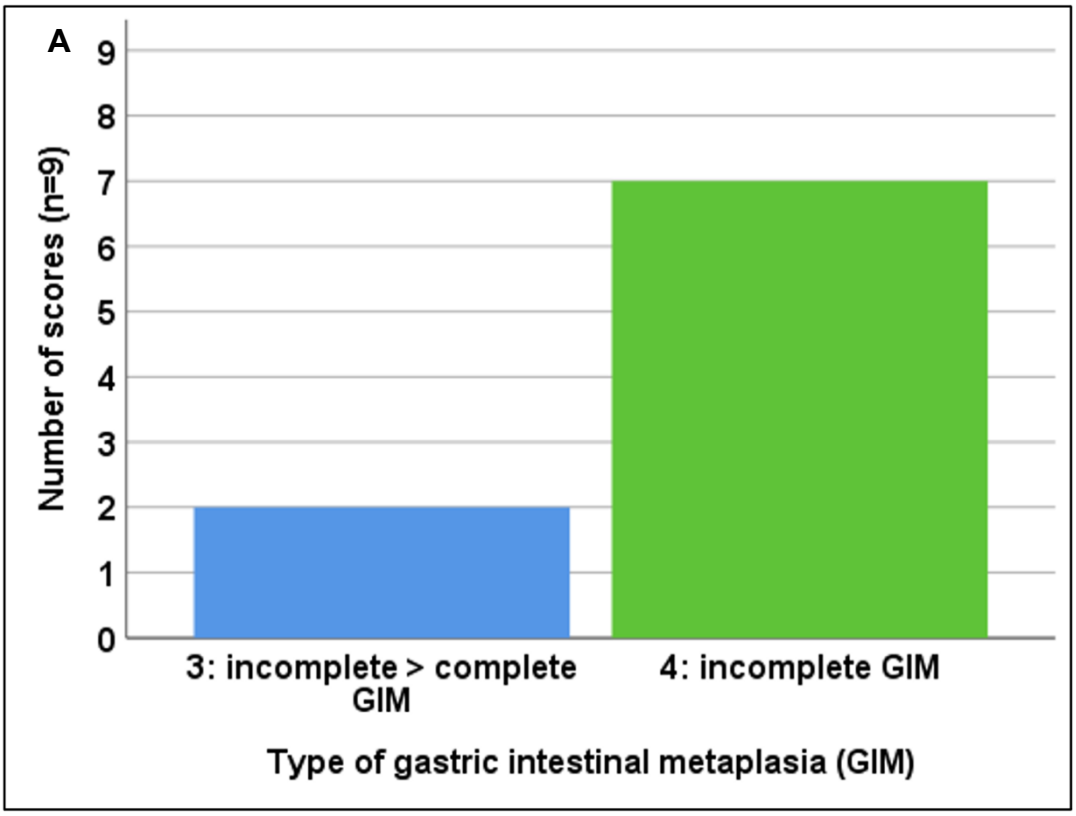
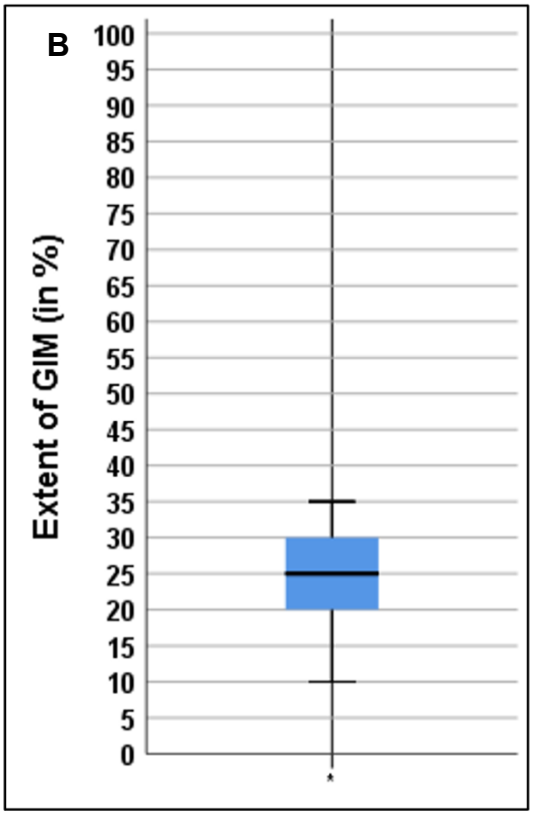
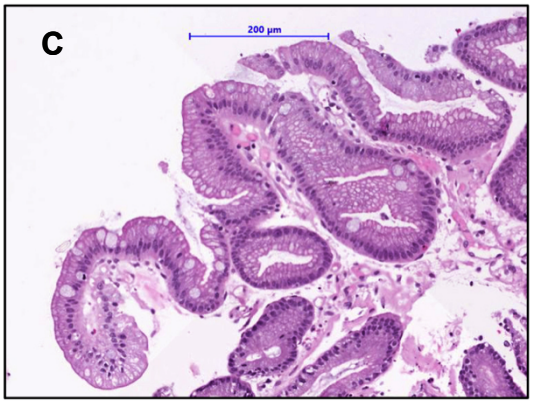


Figure 12: Case 05. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







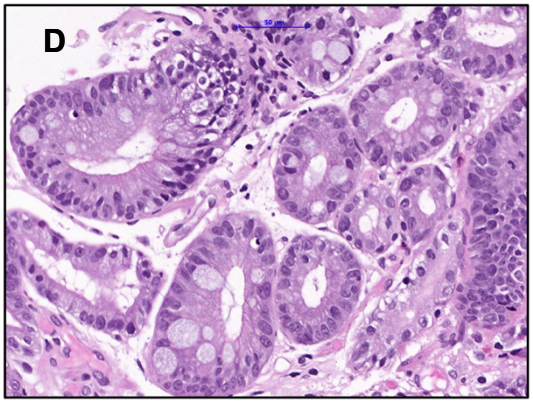


Figure 13: Case 06. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

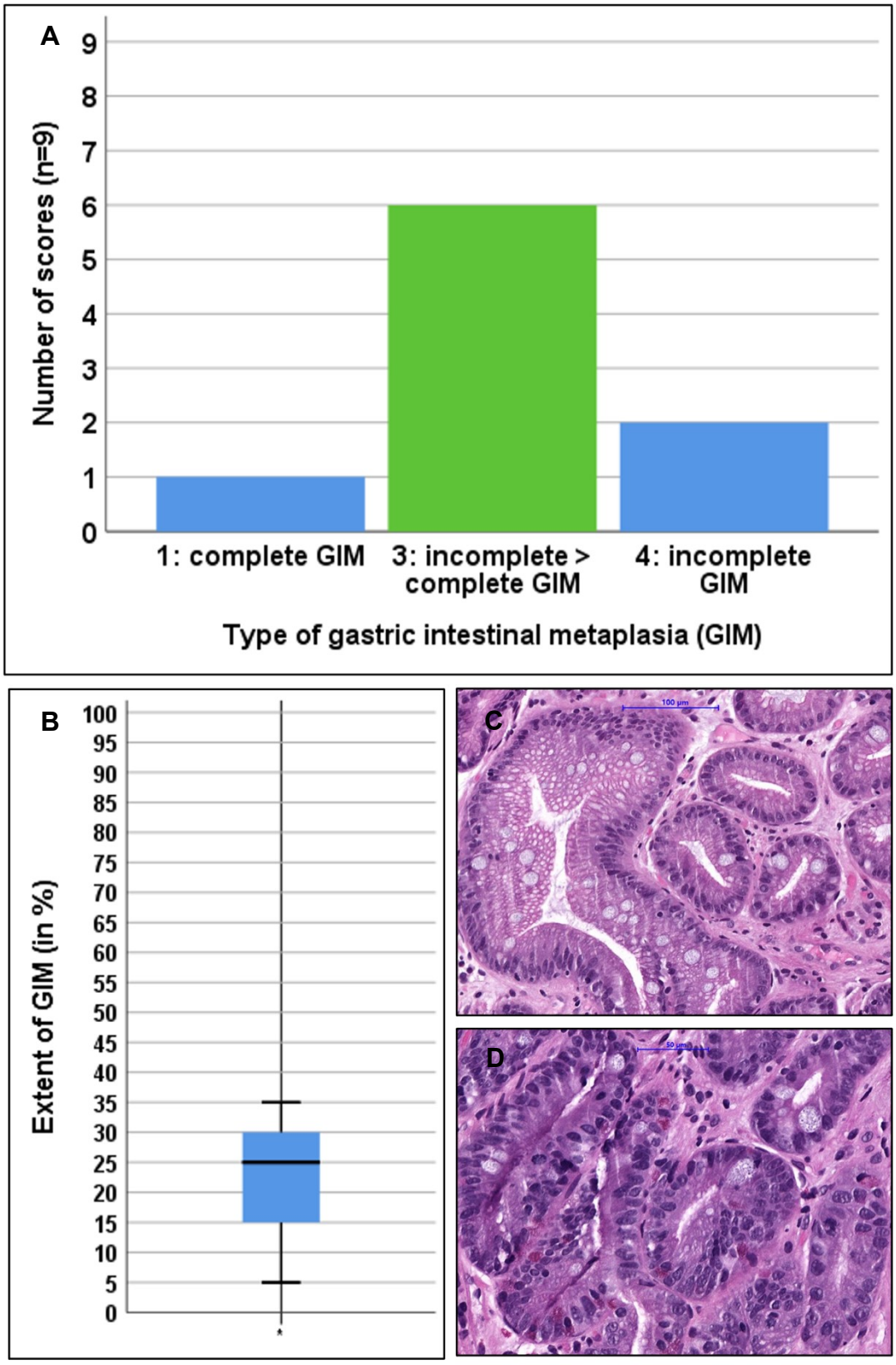


Figure 14: Case 07. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

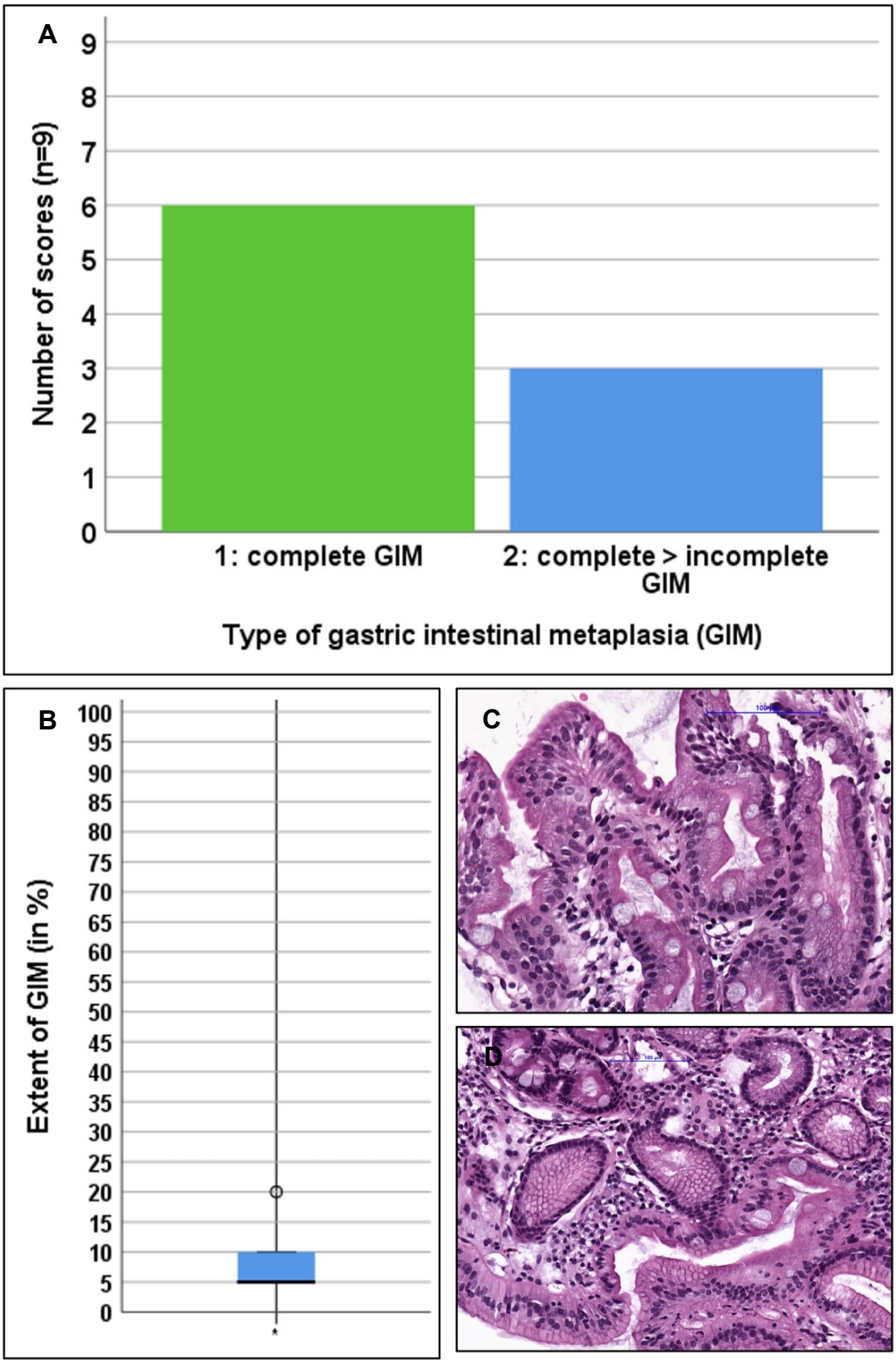
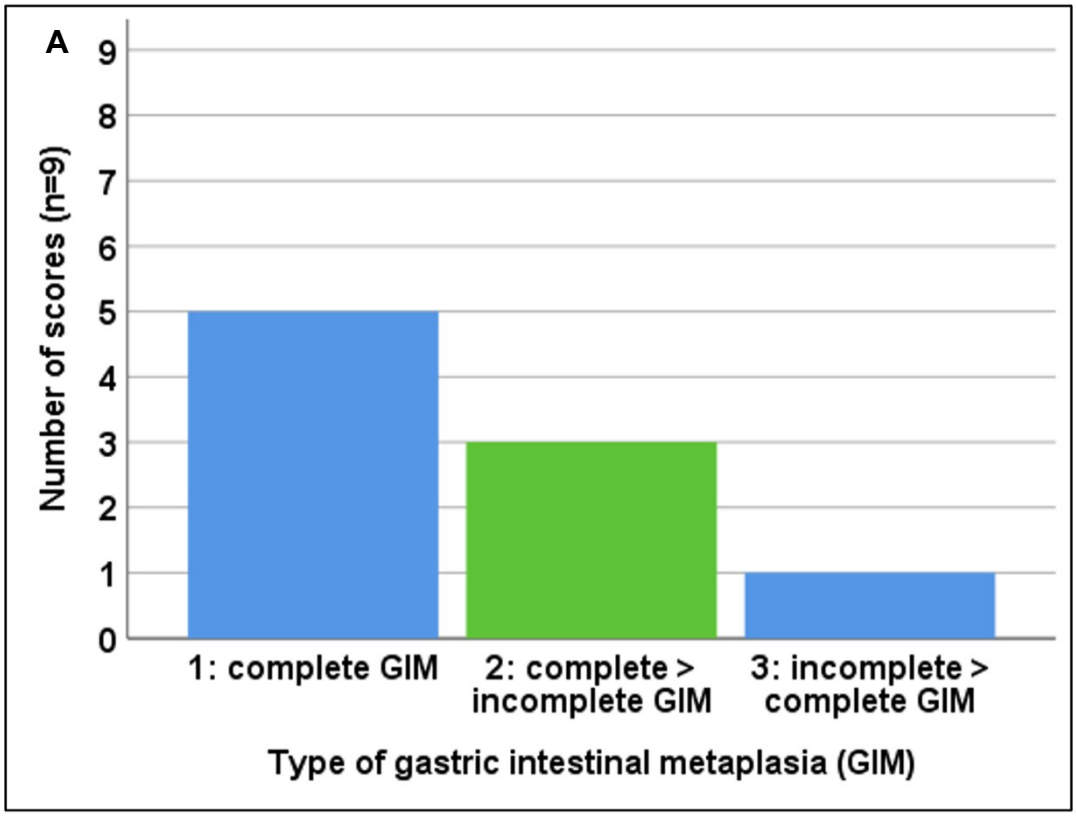
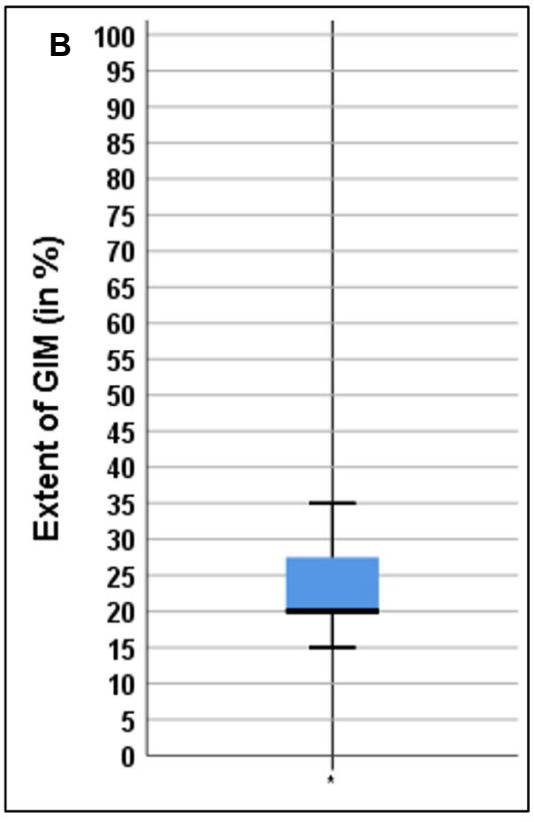
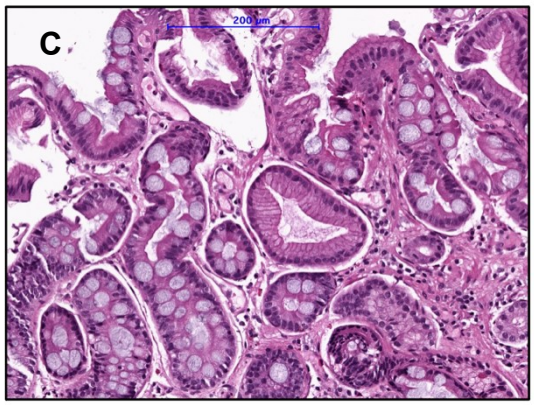


Figure 15: Case 08. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







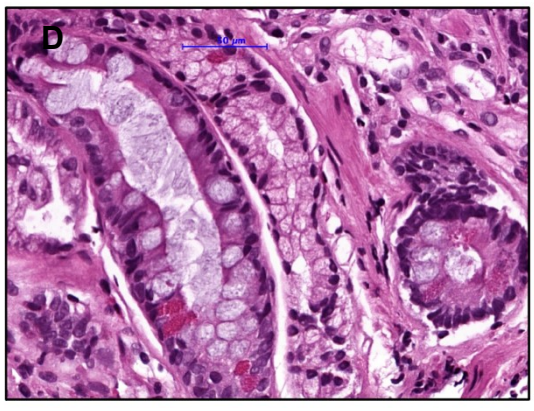


Figure 16: Case 09. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

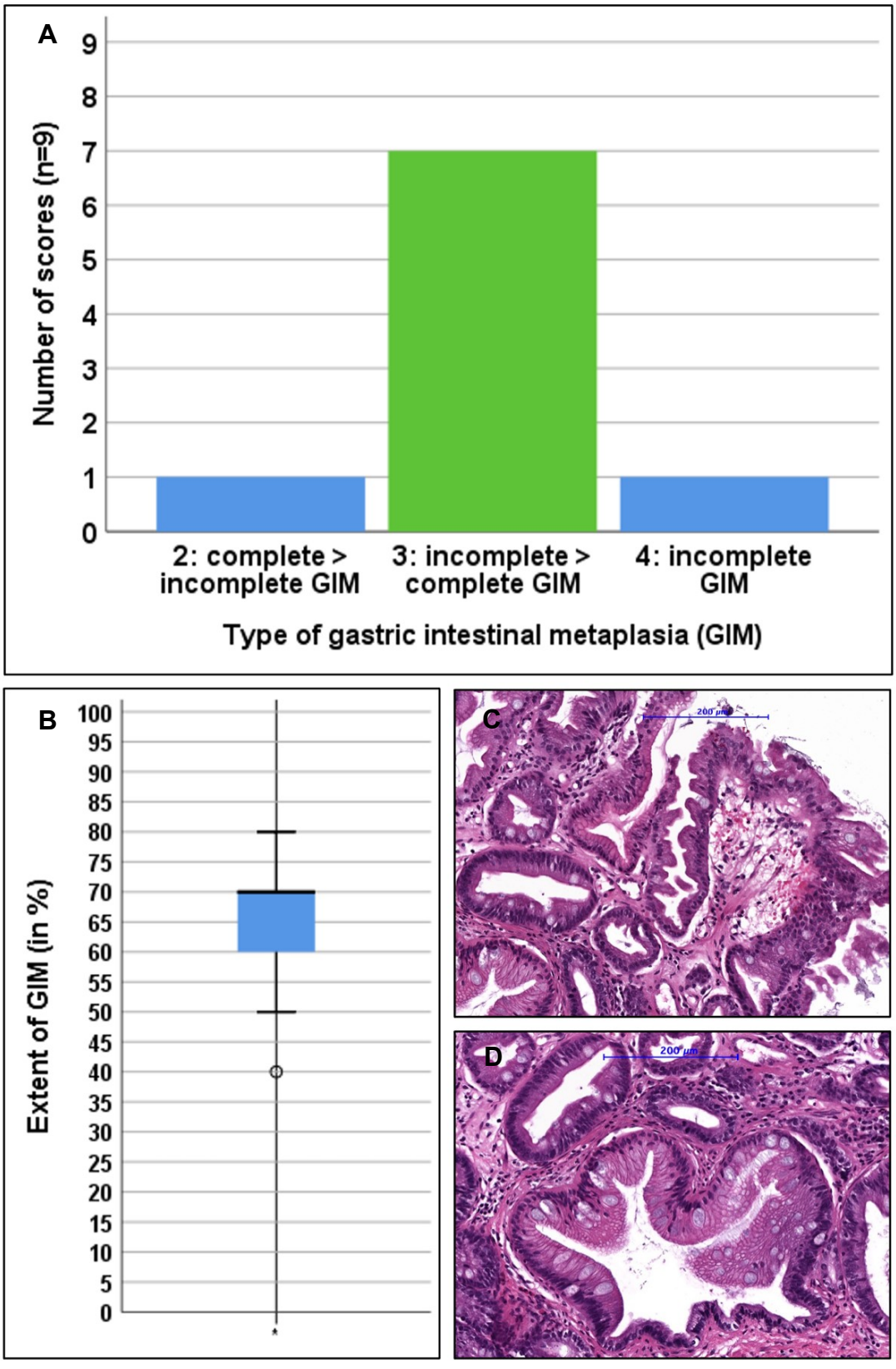


Figure 17: Case 10. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

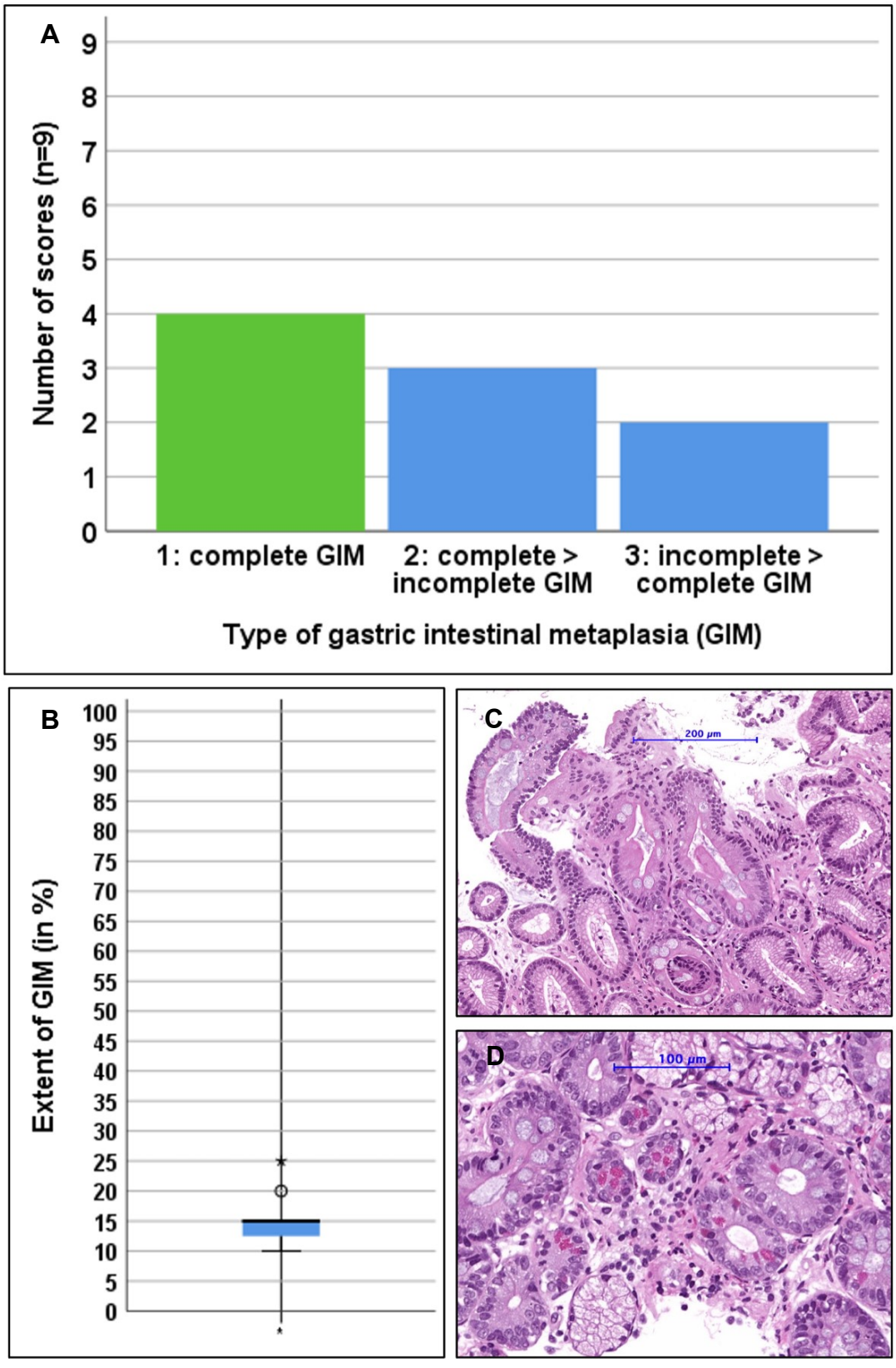
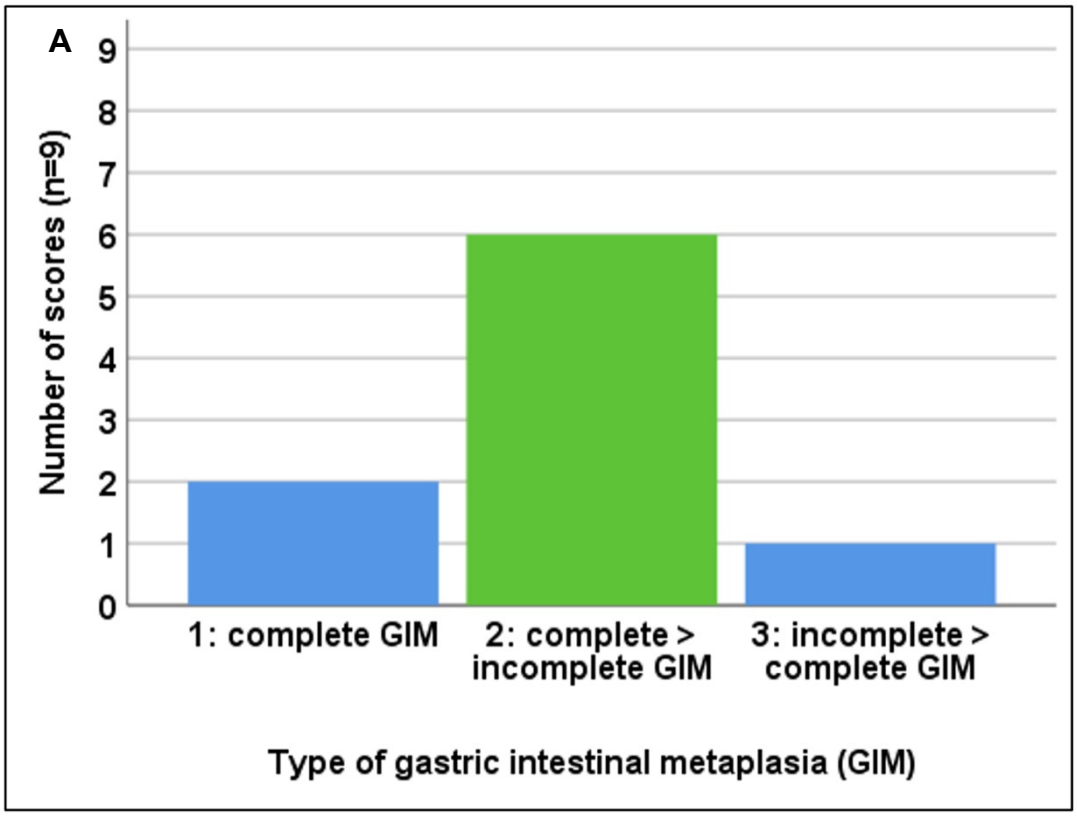
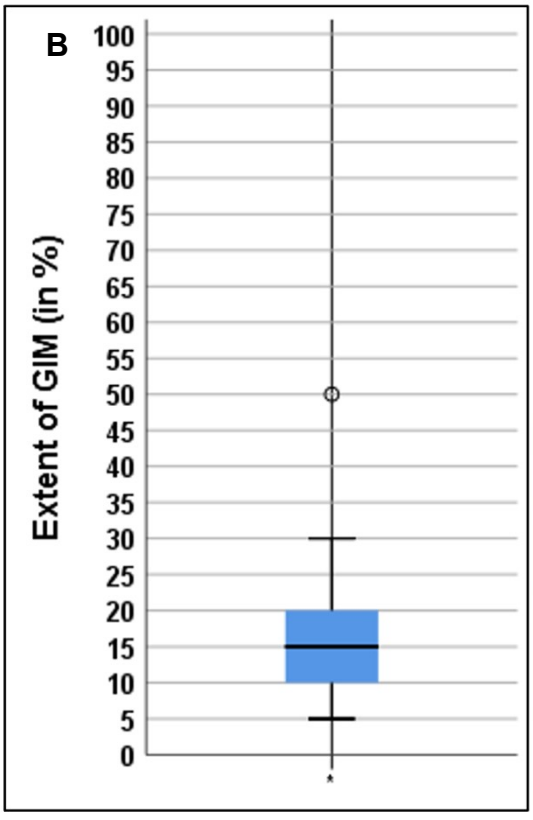
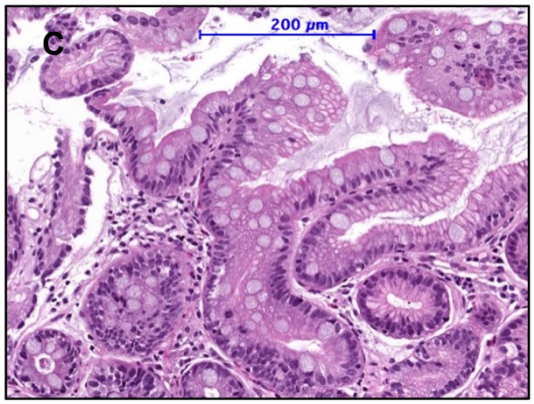


Figure 18: Case 11. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







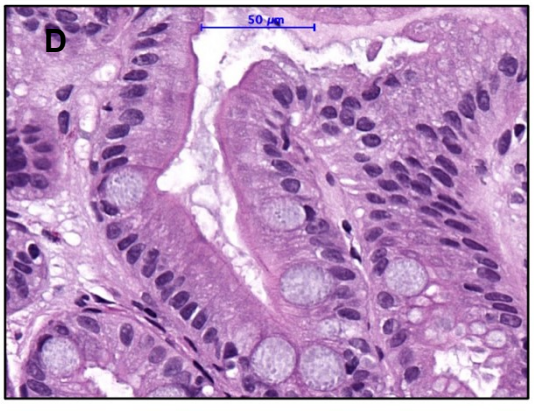
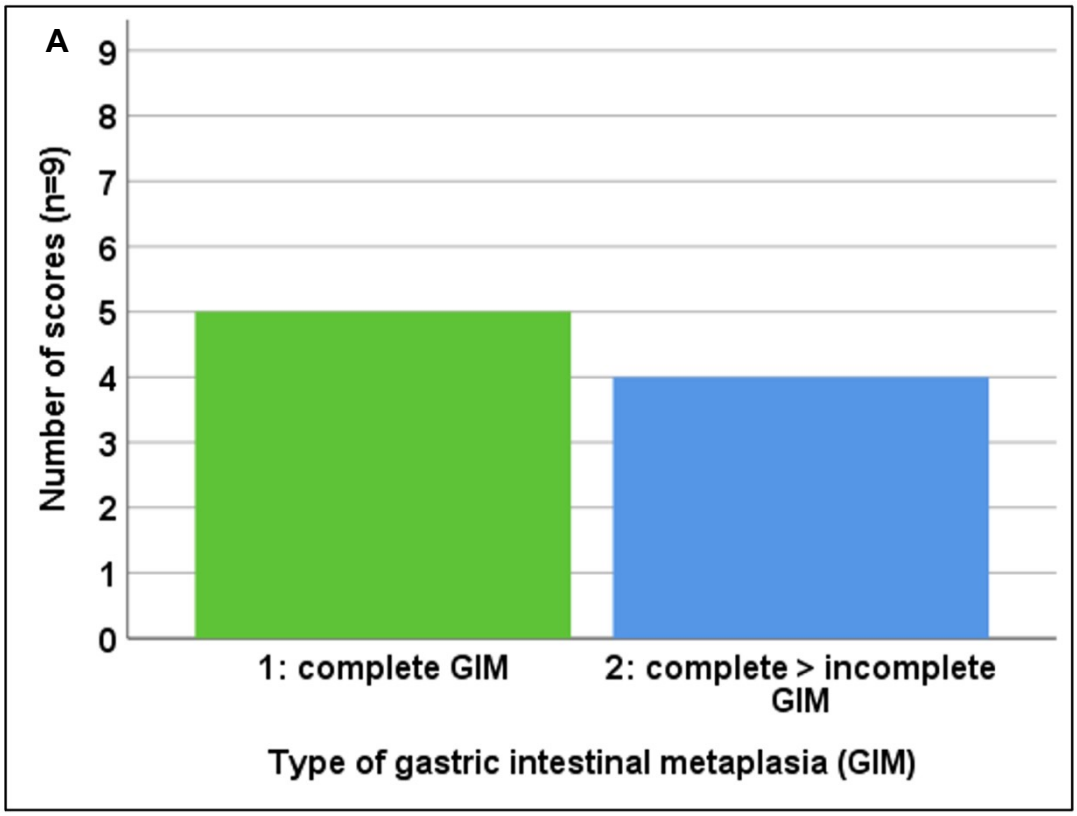
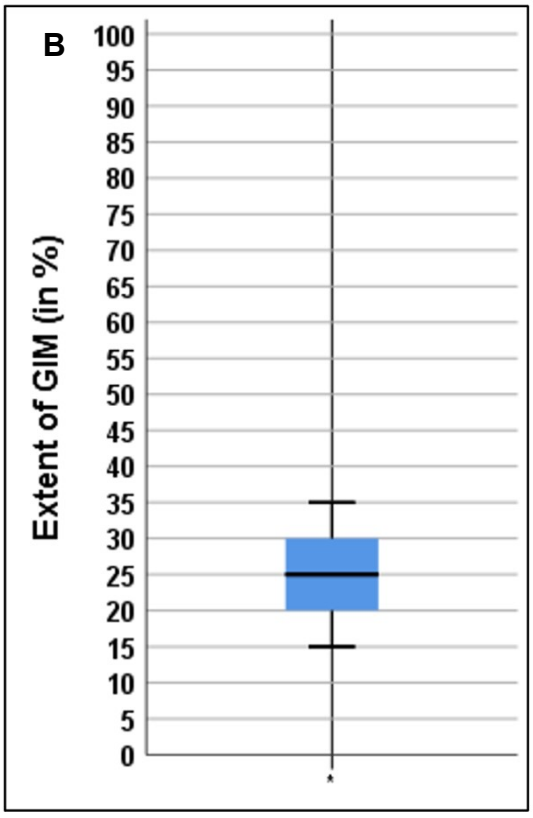
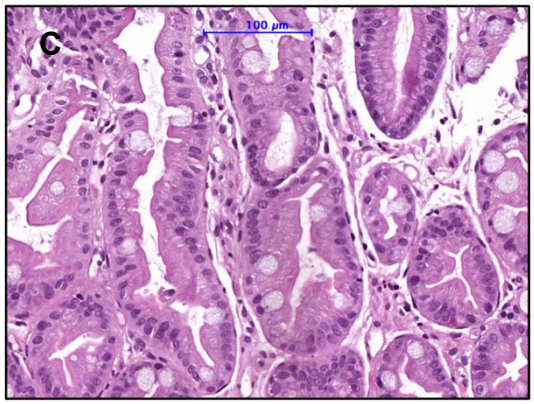


Figure 19: Case 12. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







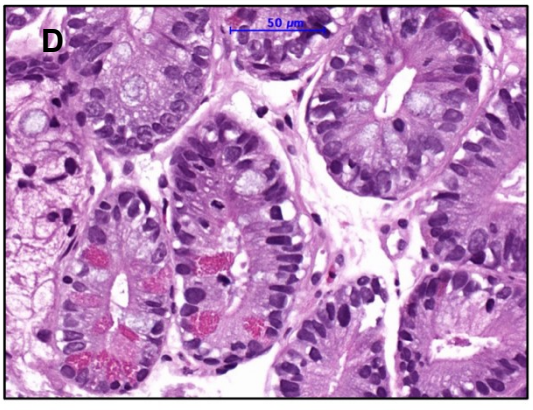
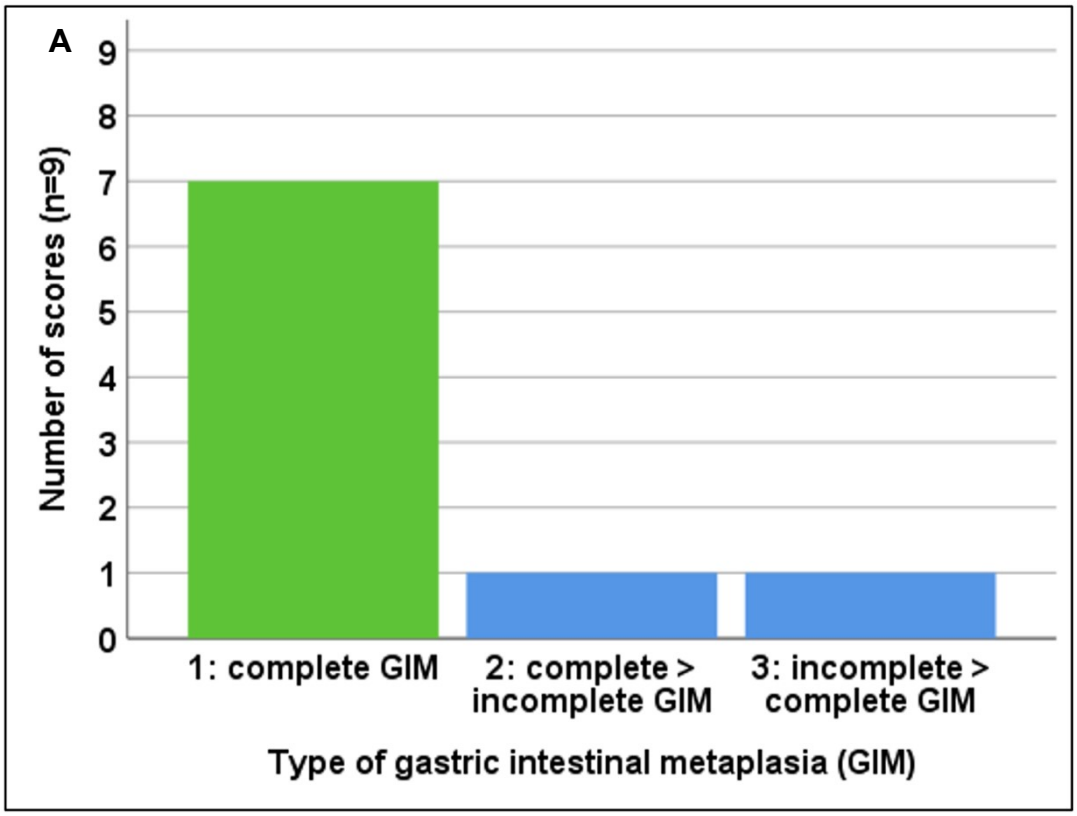
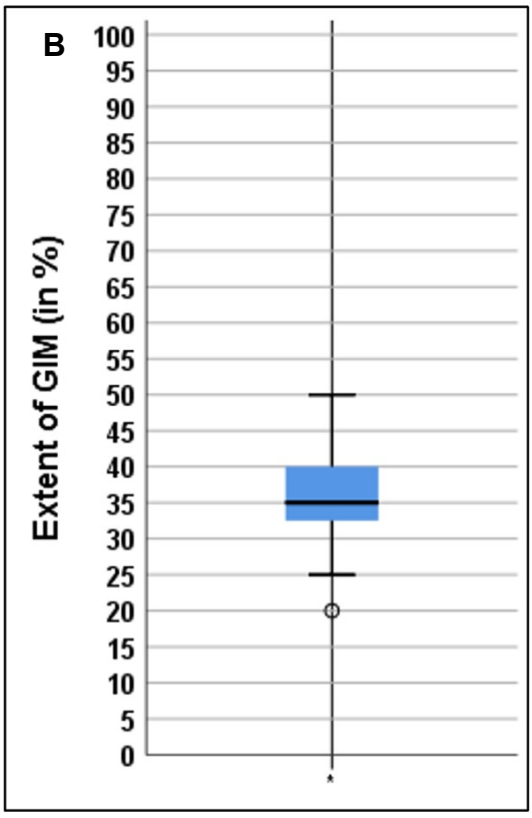
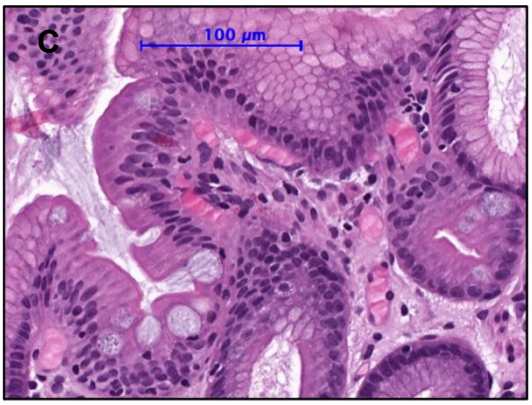


Figure 20: Case 13. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







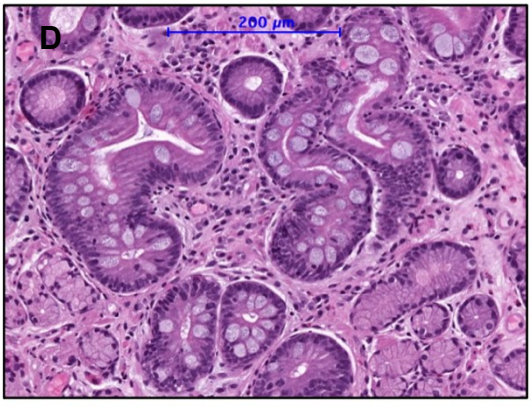


Figure 21: Case 14. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

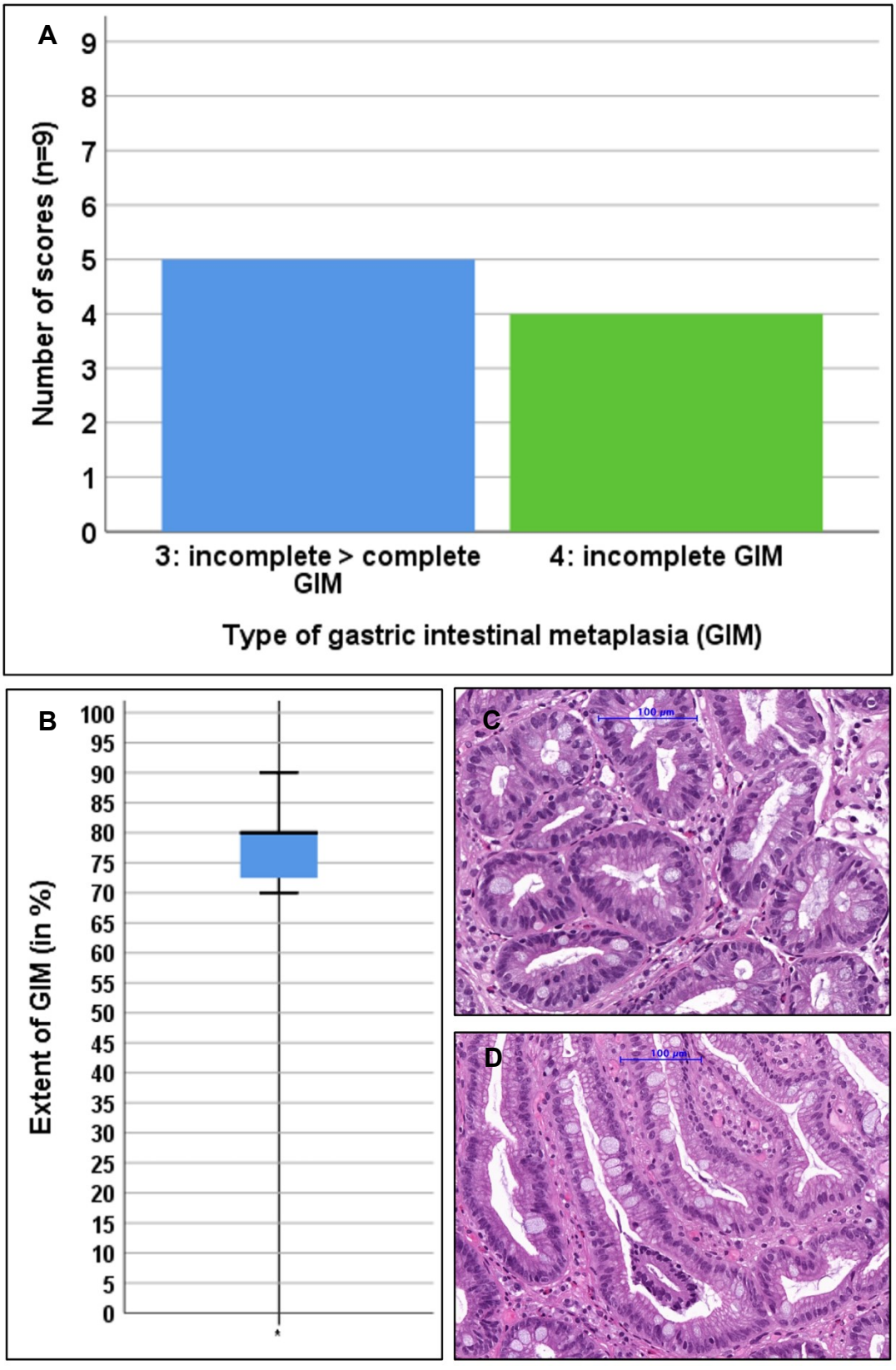


Figure 22: Case 15. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

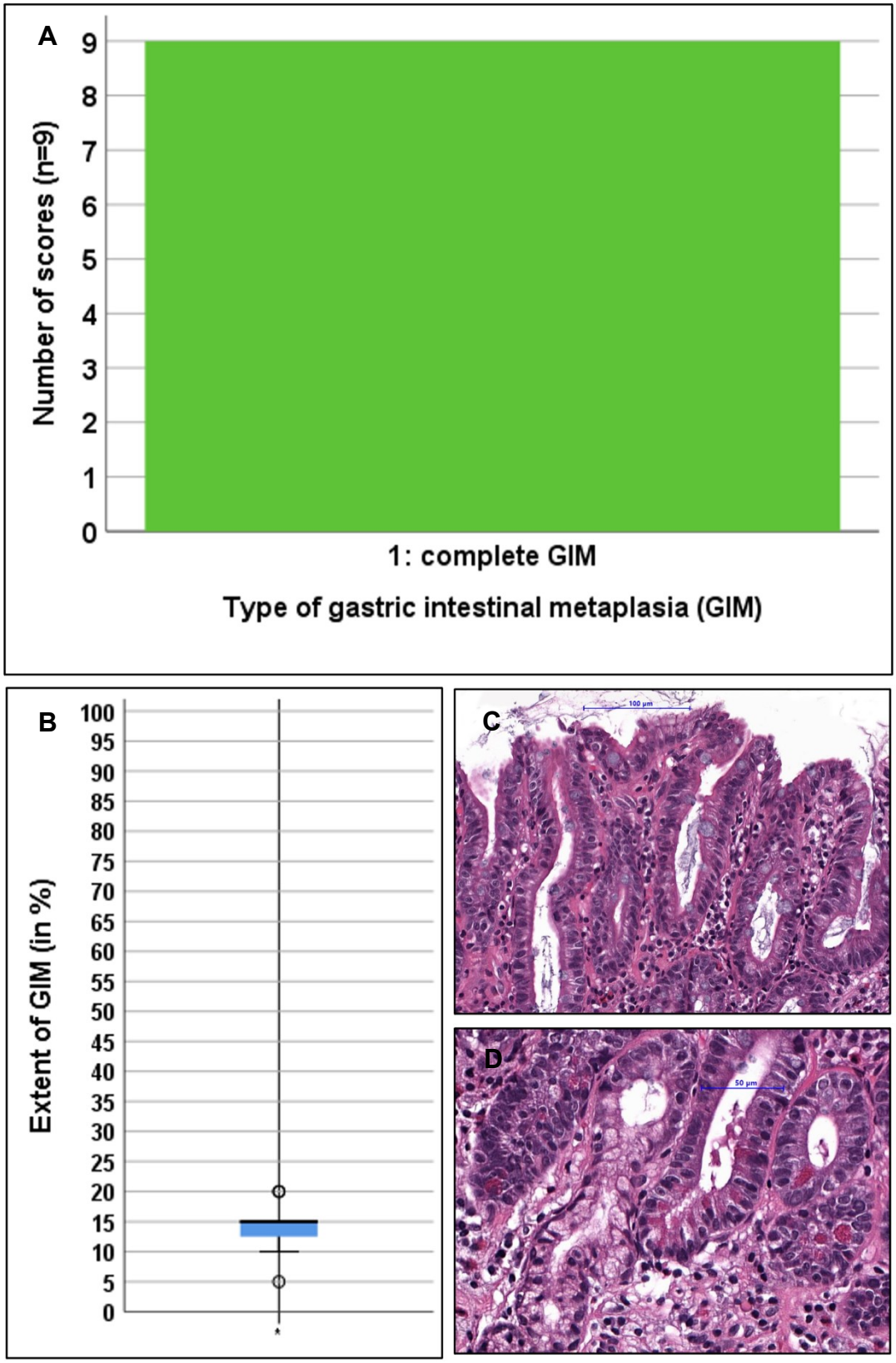
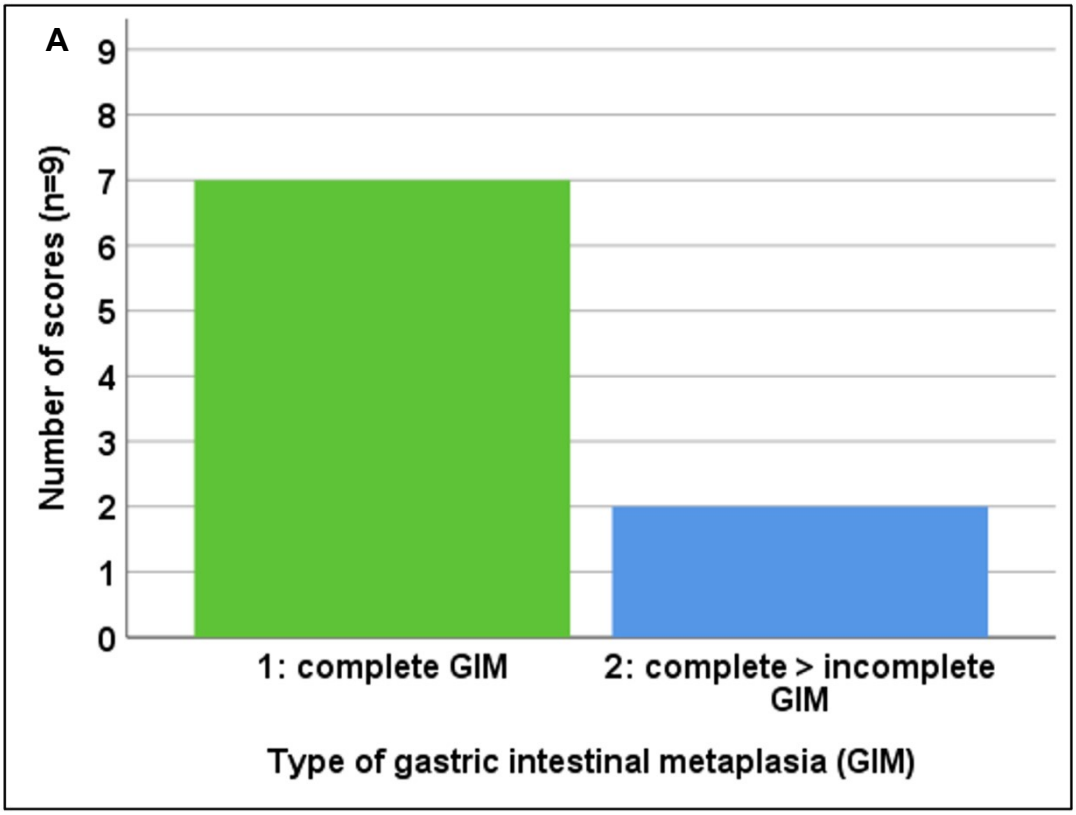
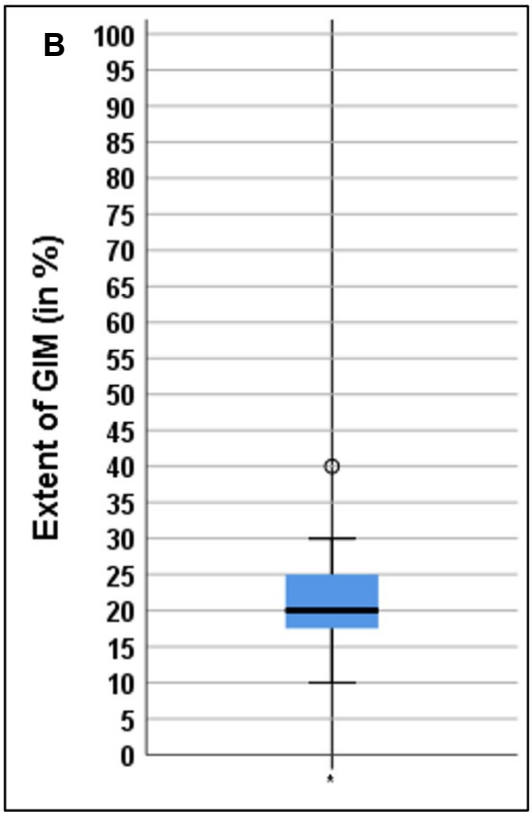
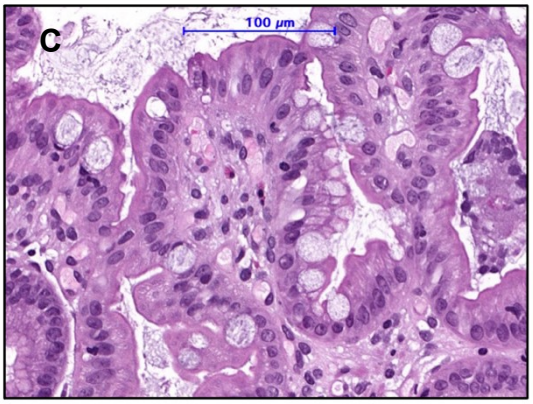


Figure 23: Case 16. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







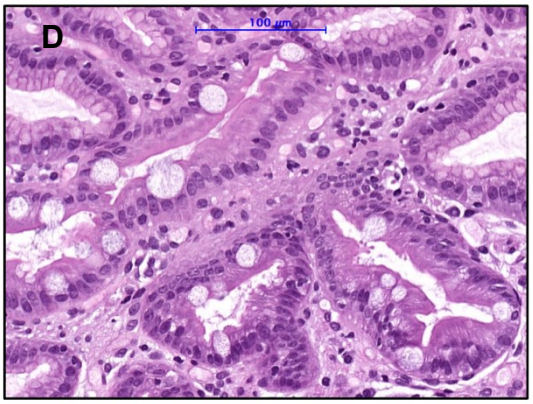
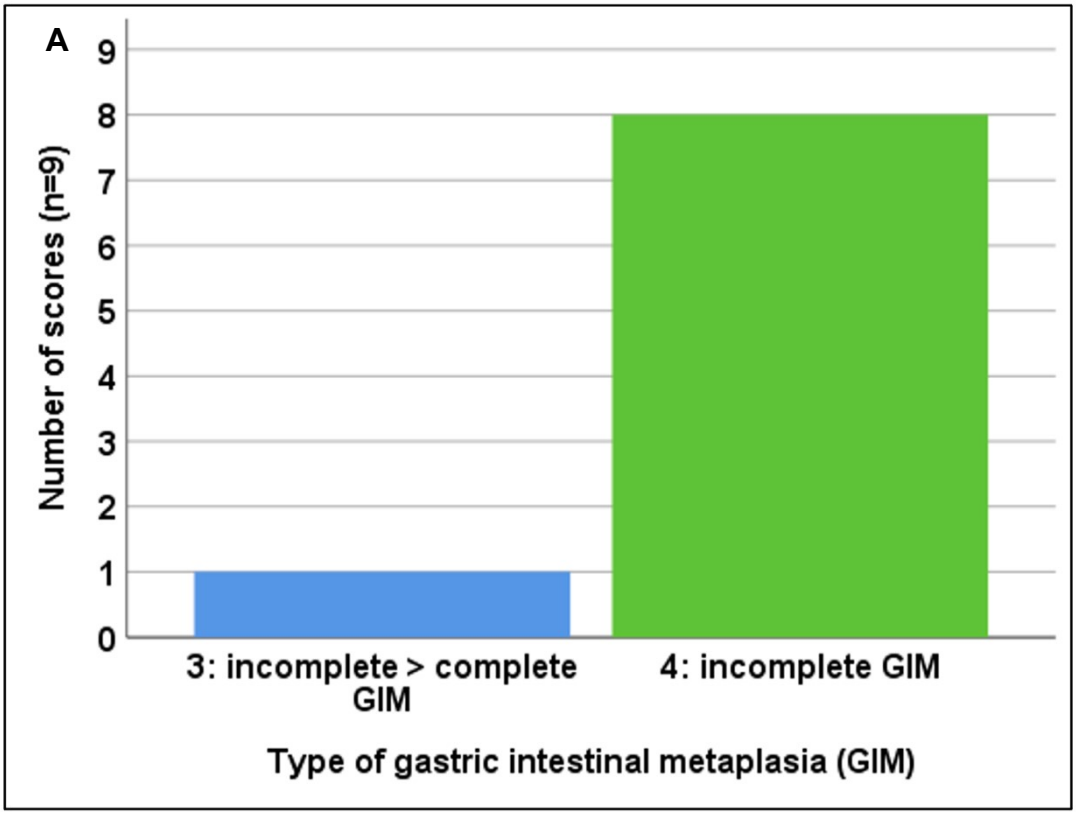
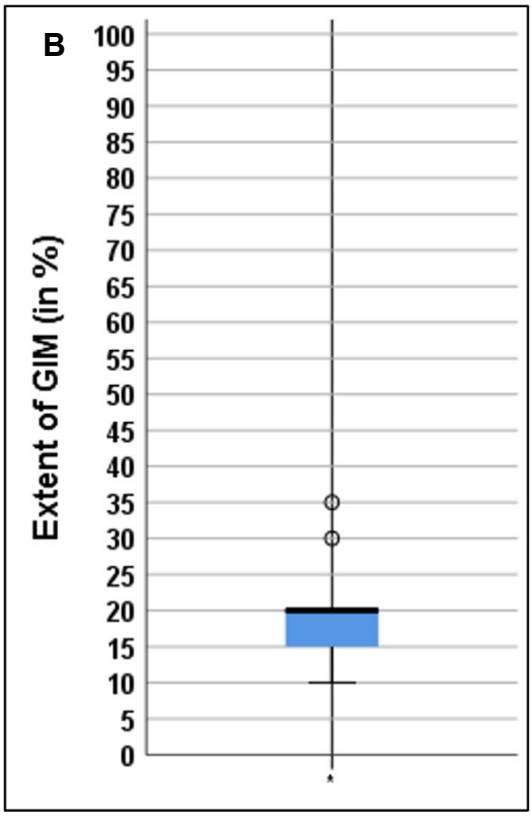
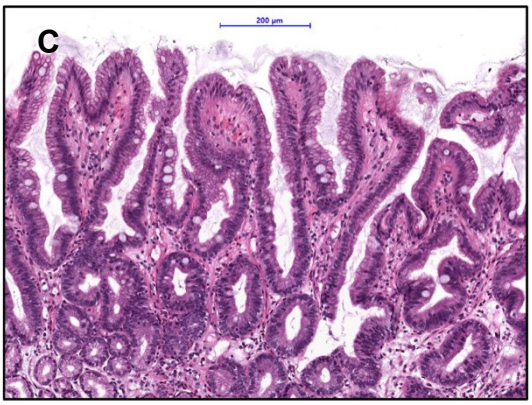


Figure 24: Case 17. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







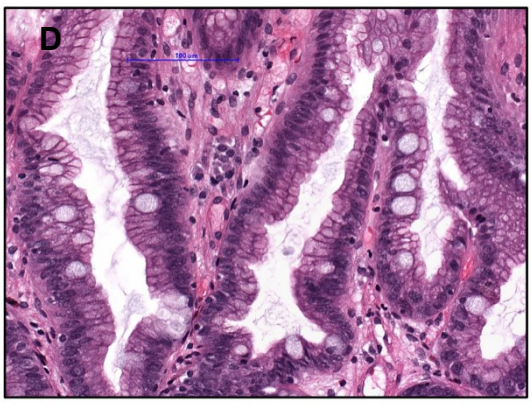
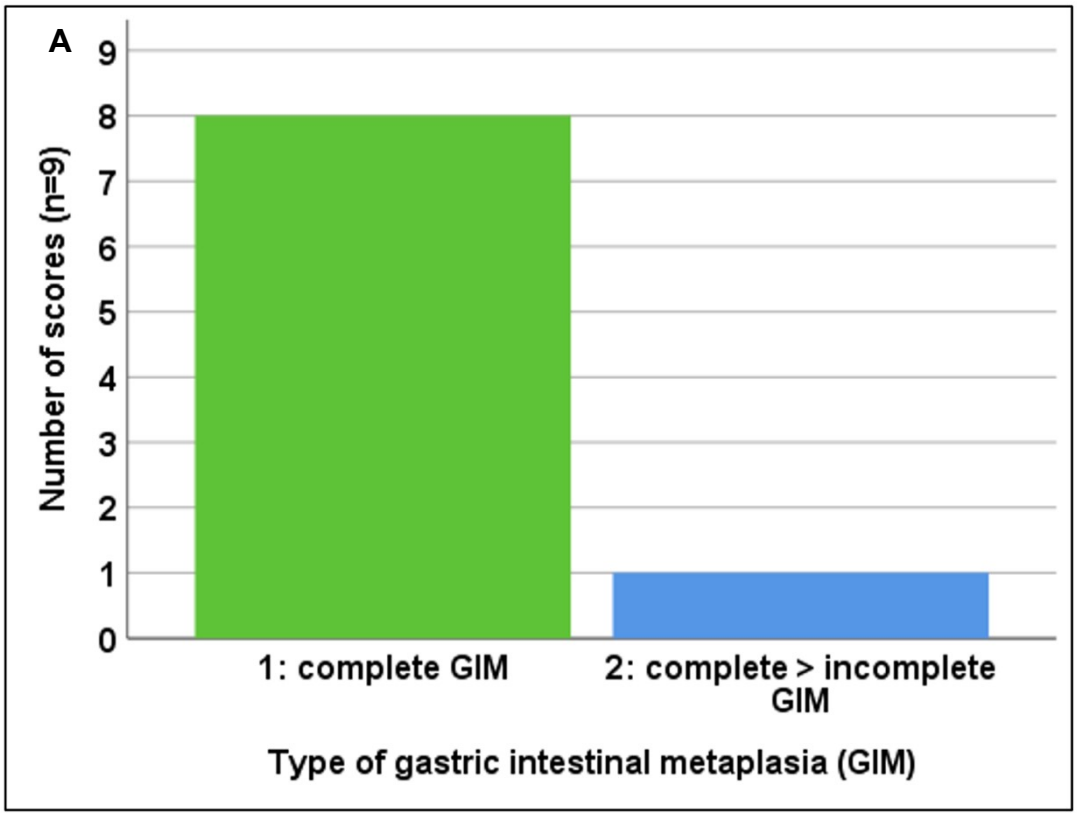
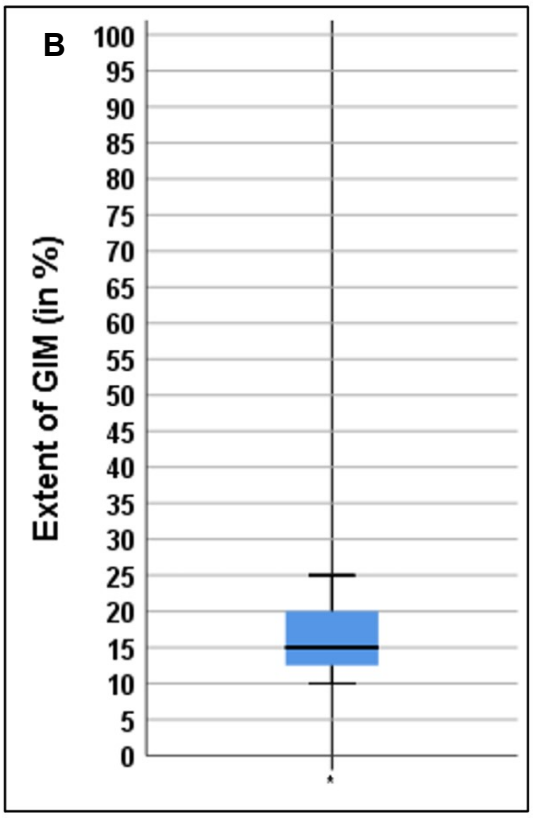
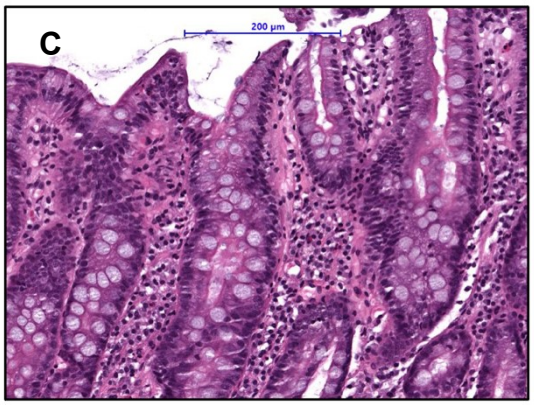


Figure 25: Case 18. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







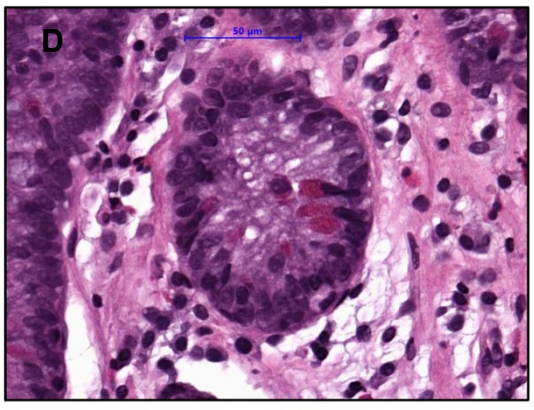
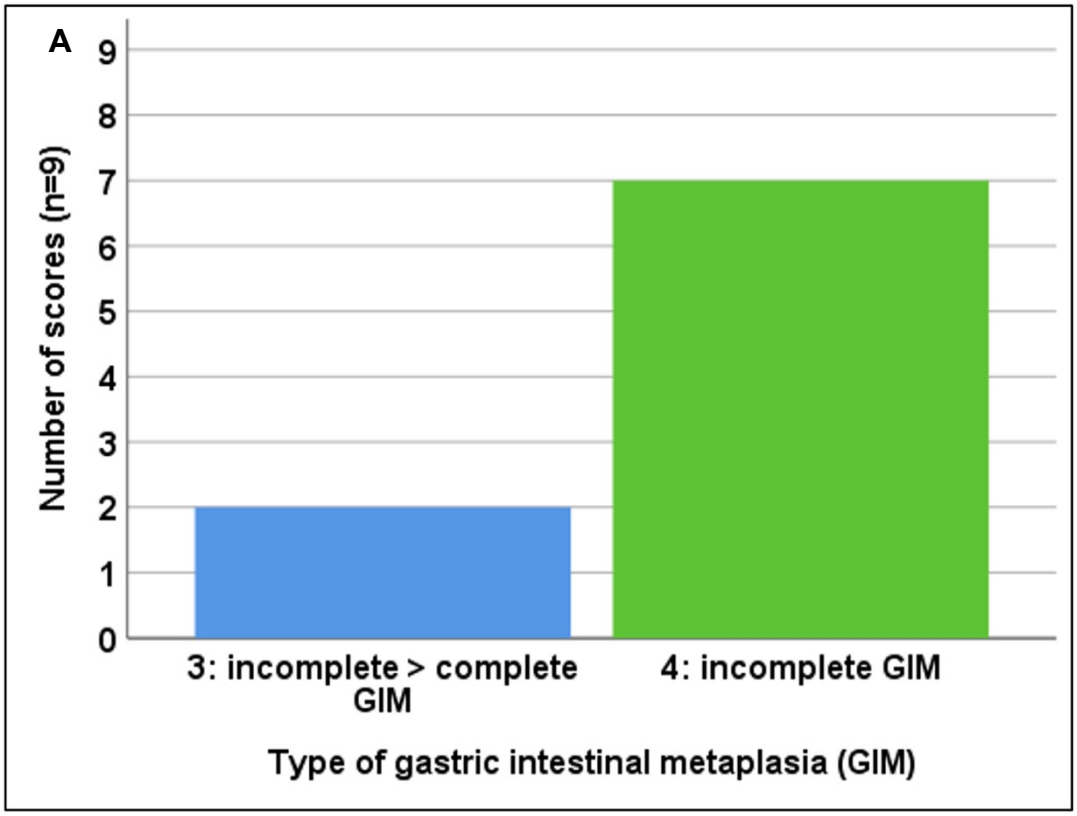
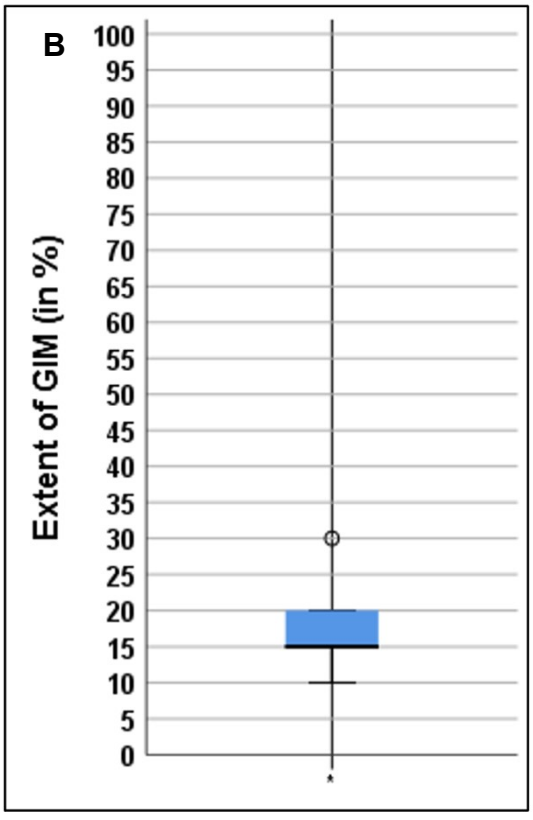
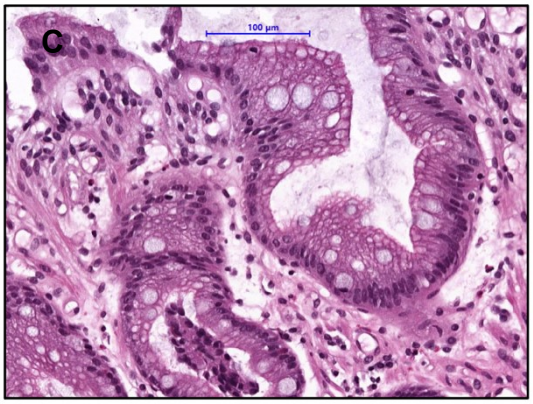


Figure 26: Case 19. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







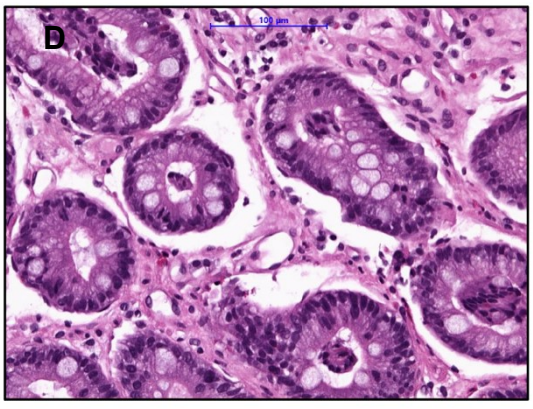
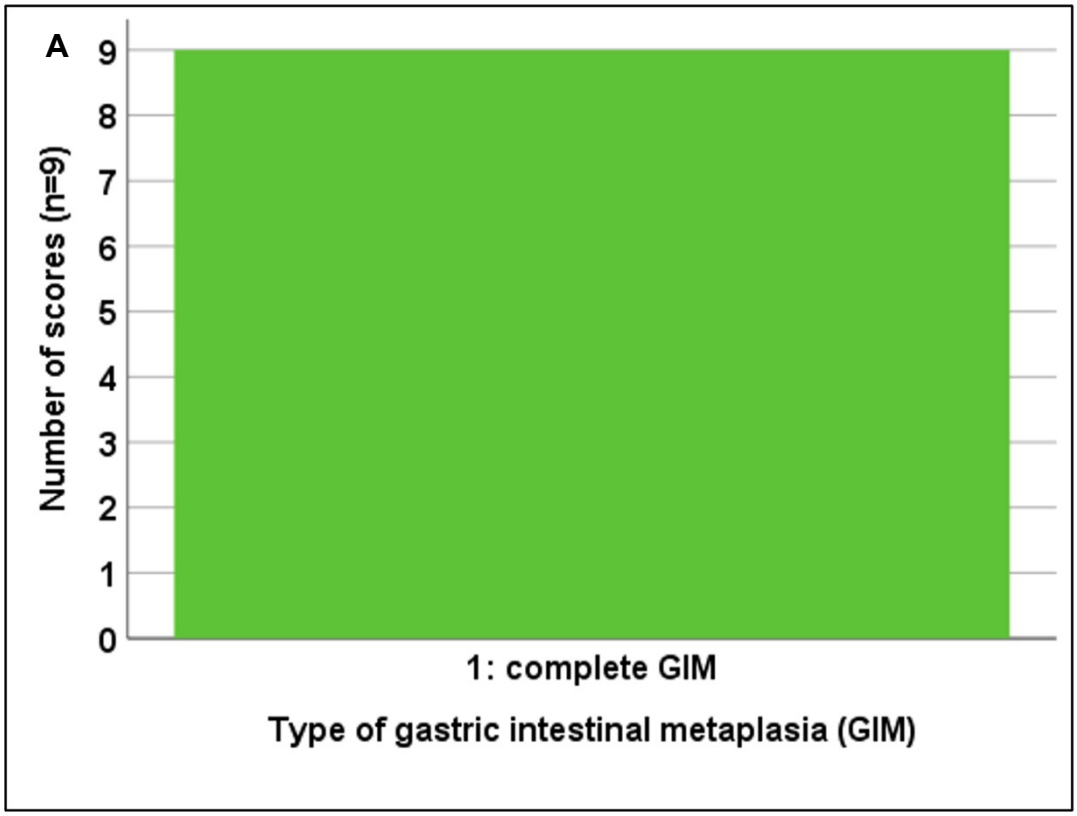
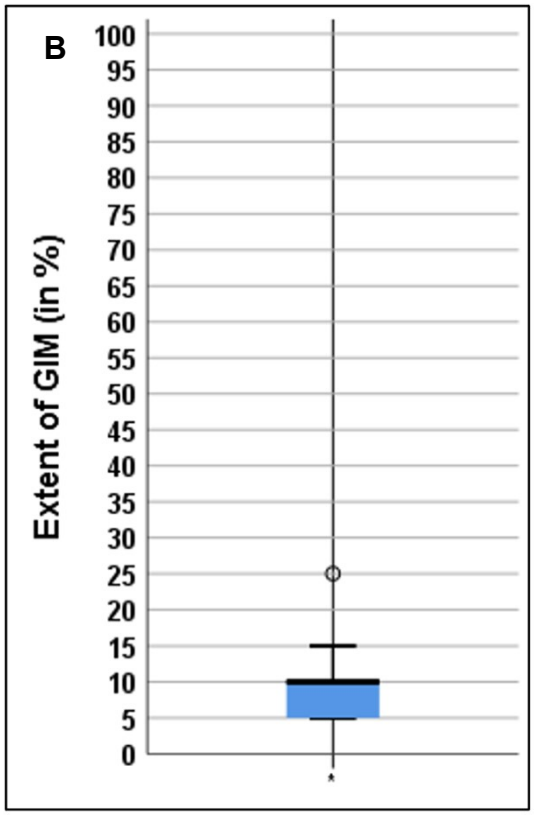
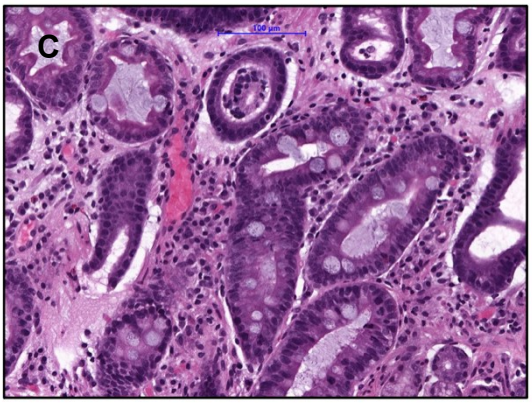


Figure 27: Case 20. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







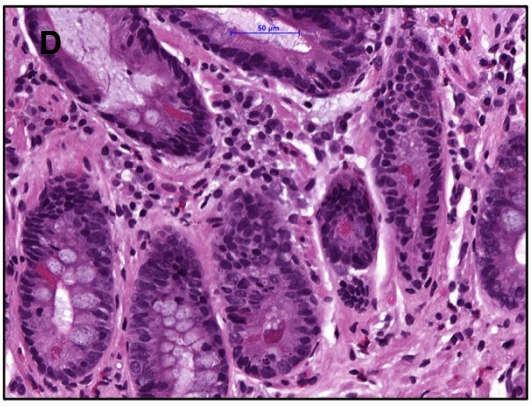


Figure 28: Case 21. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

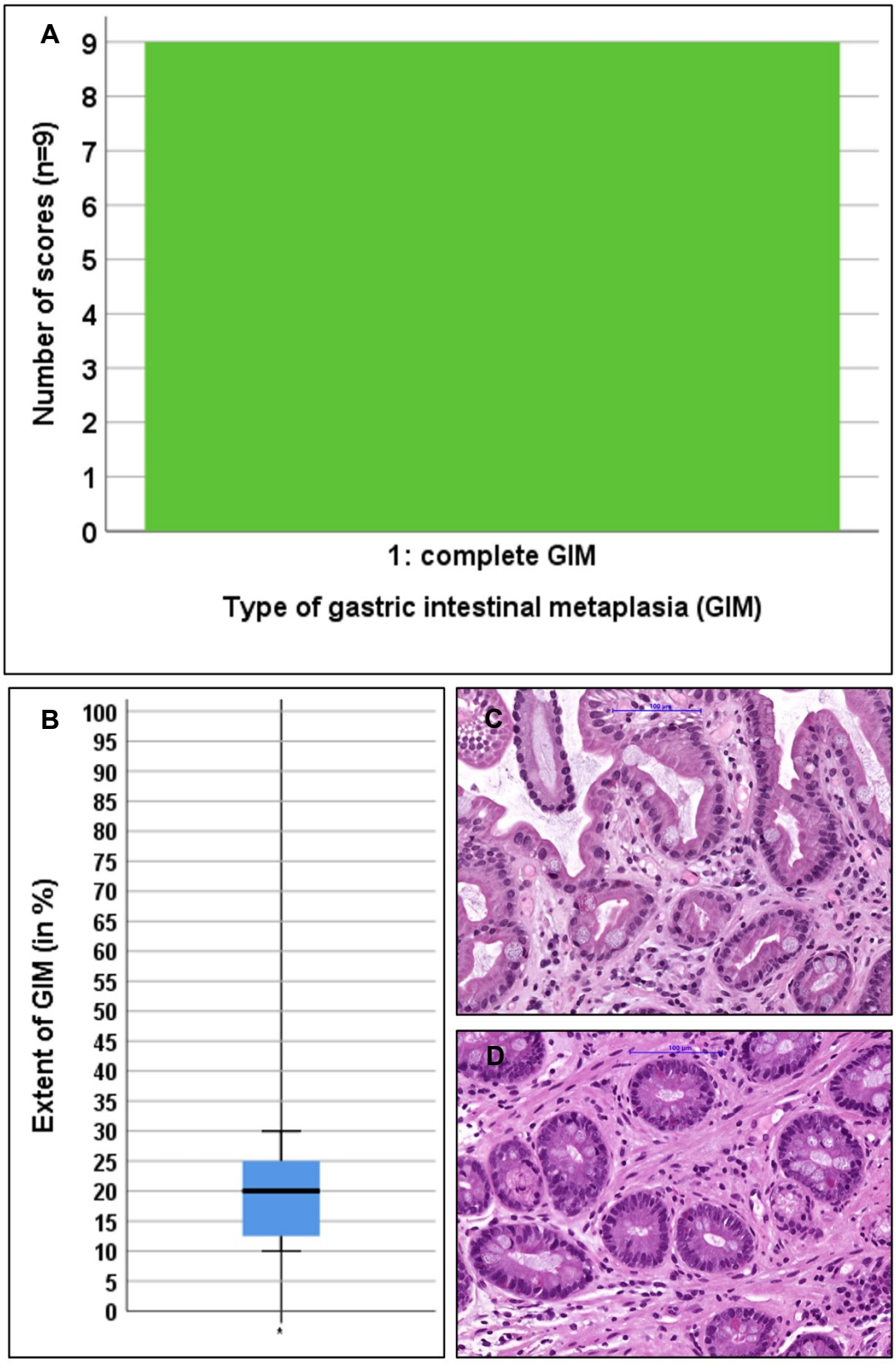
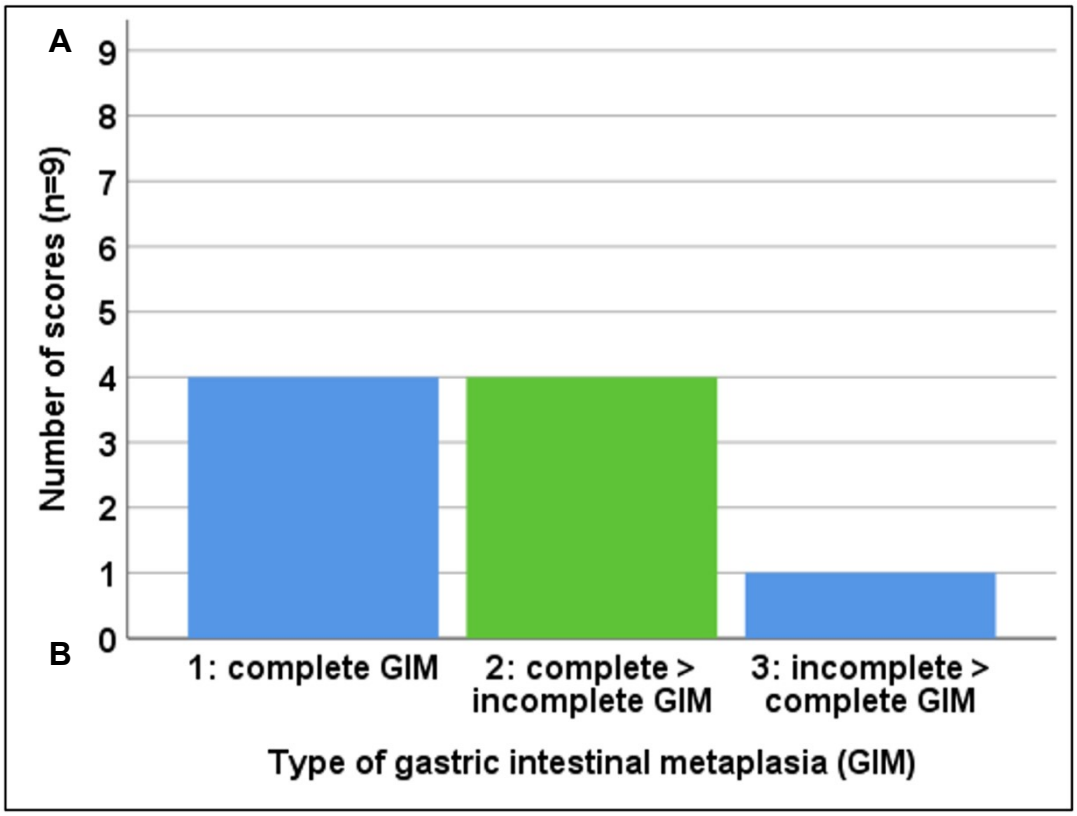
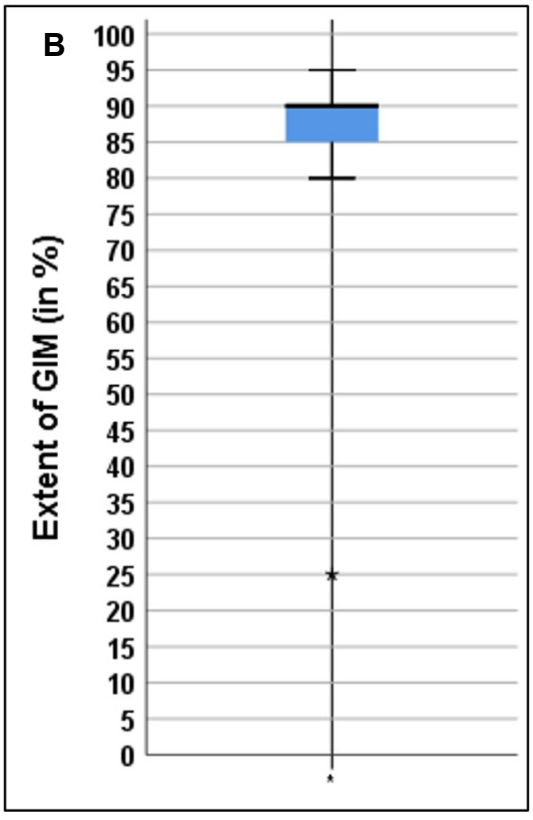
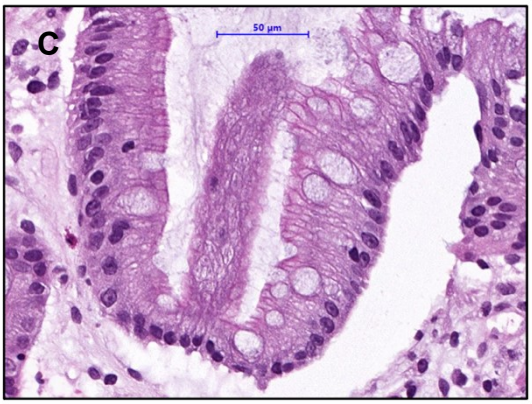


Figure 29: Case 22. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







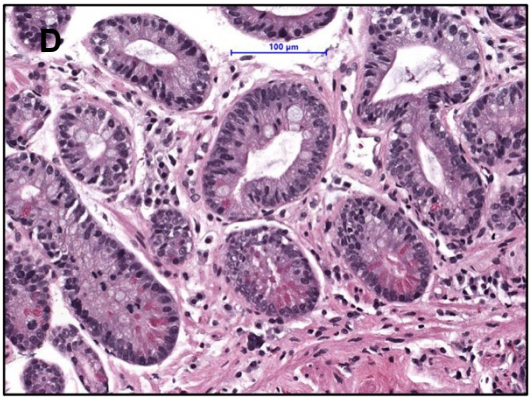
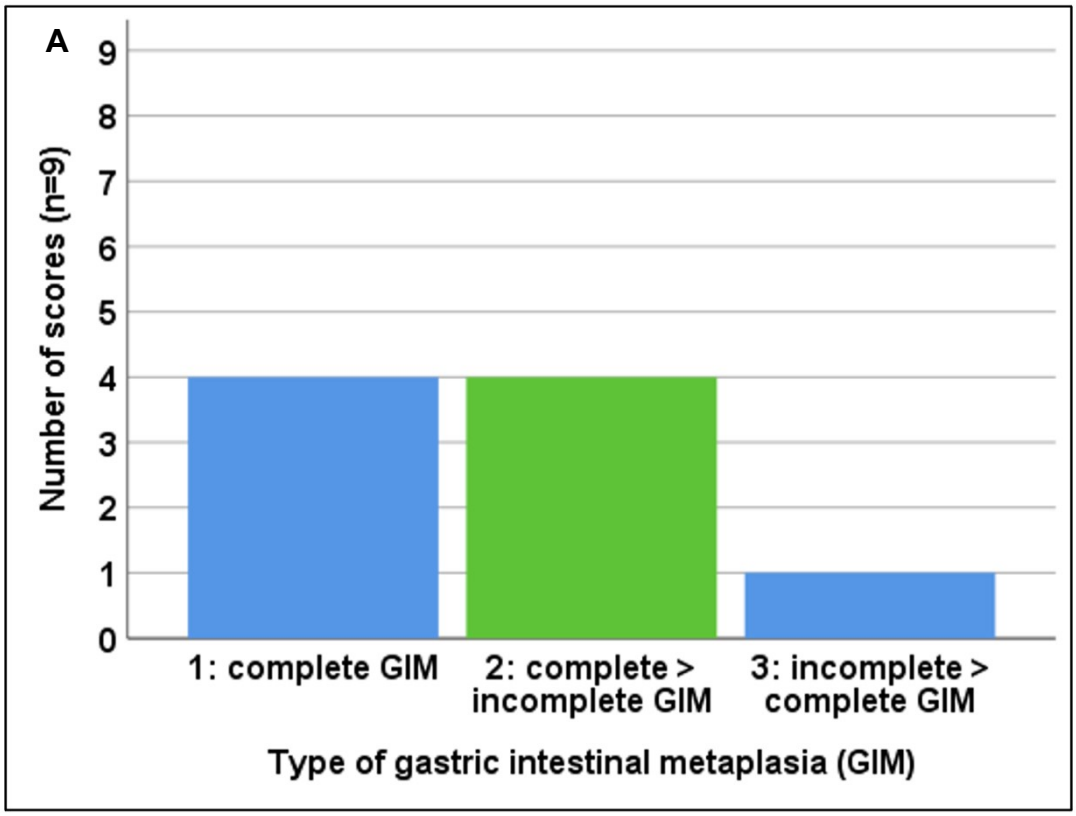
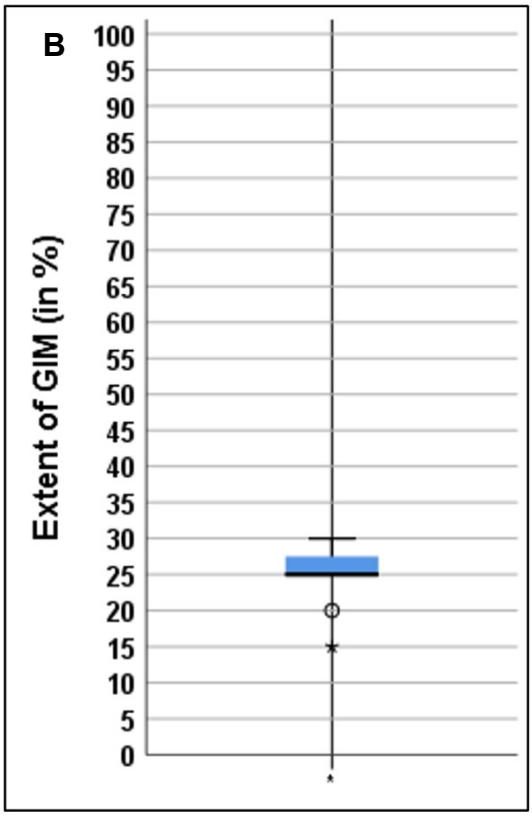
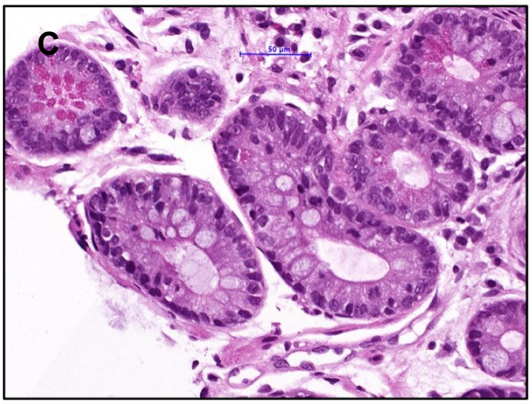


Figure 30: Case 23. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







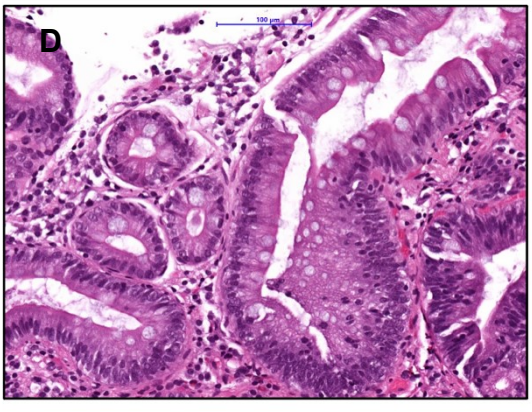


Figure 31: Case 24. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

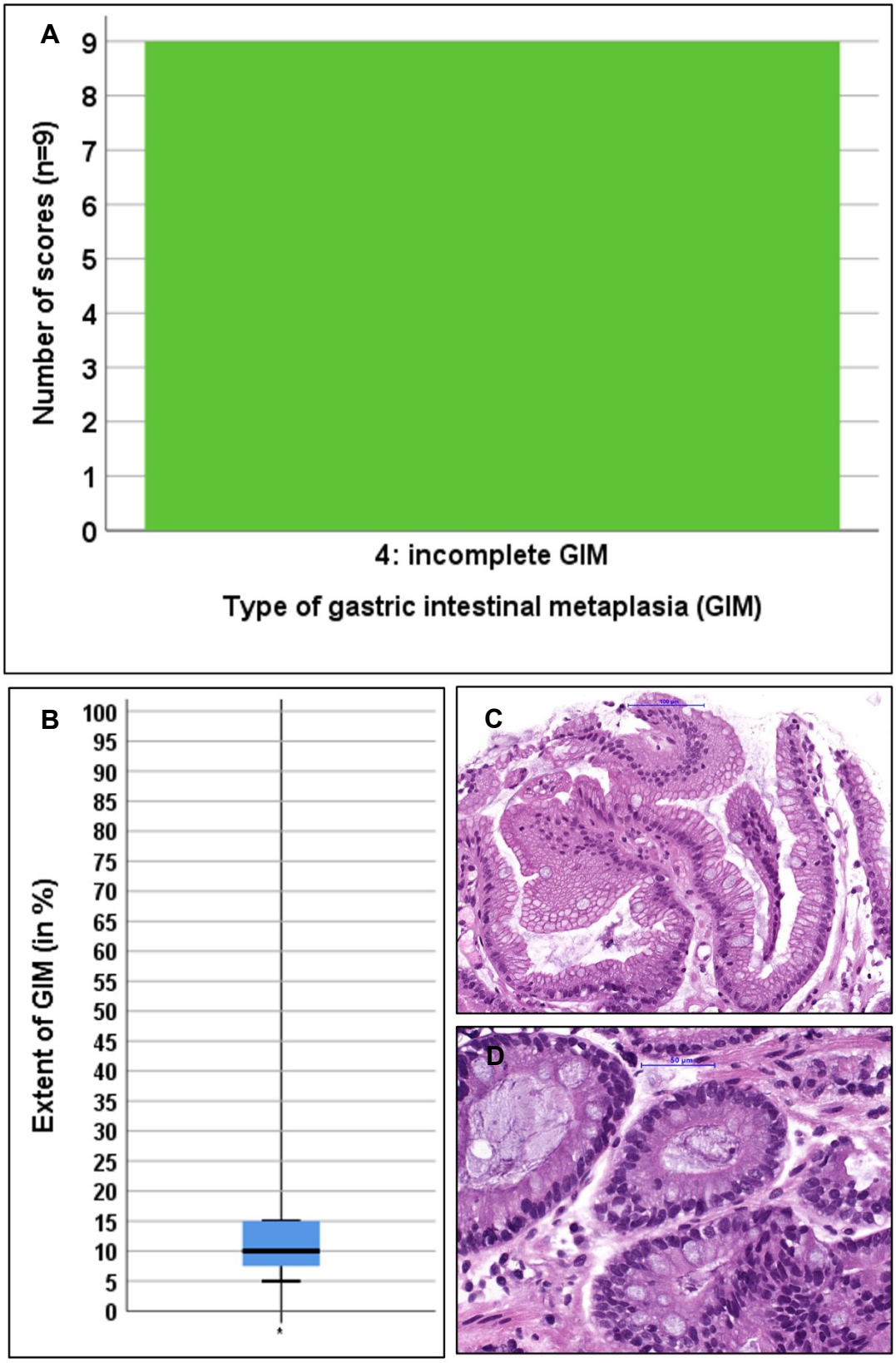


Figure 32: Case 25. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

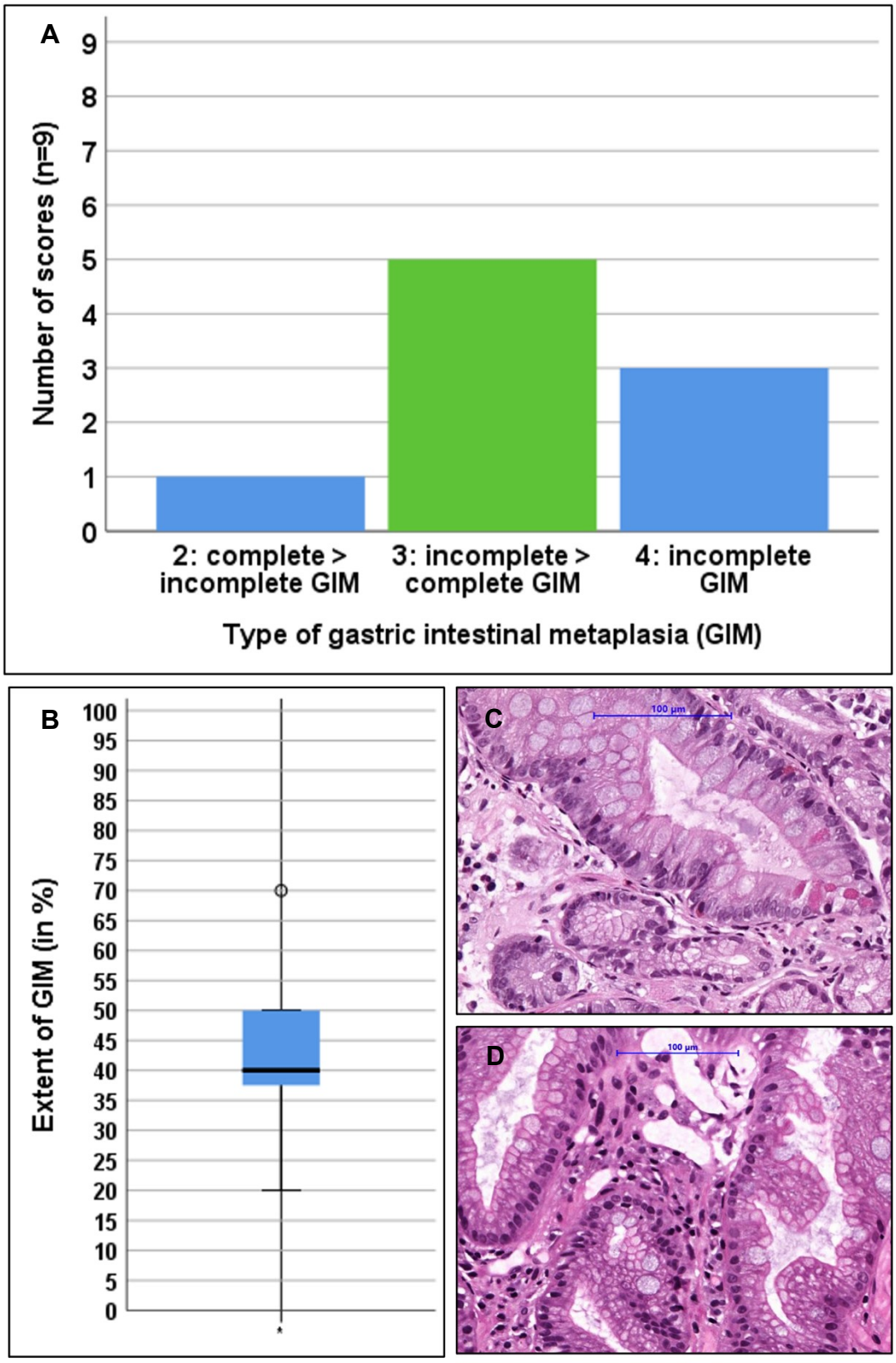
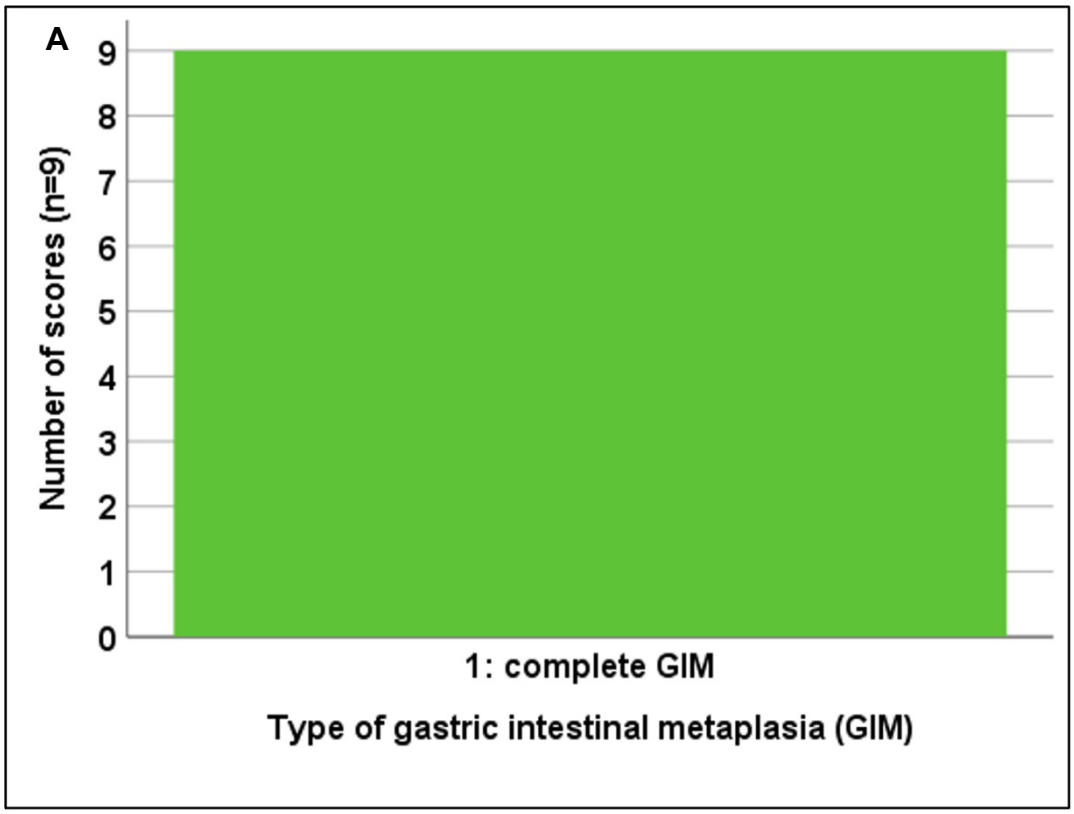
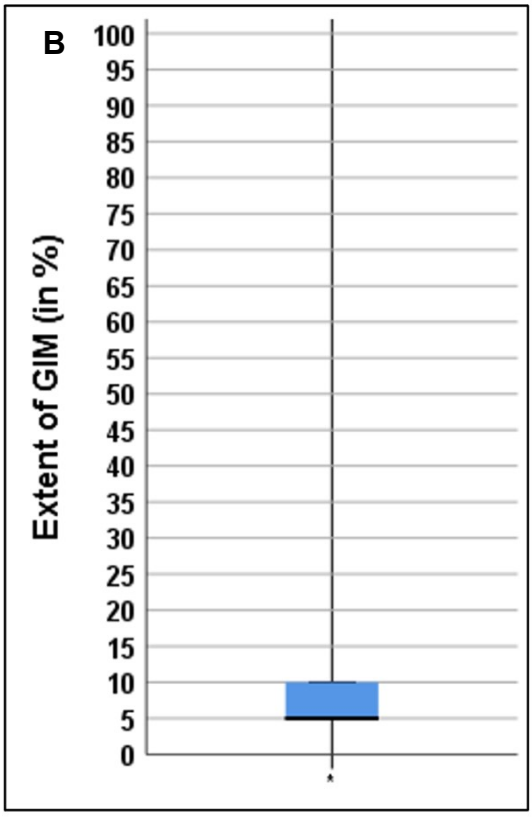
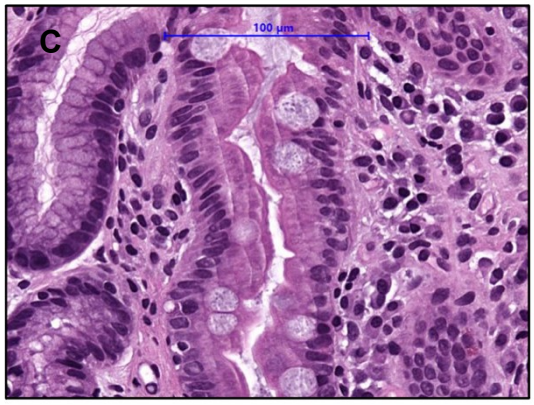


Figure 33: Case 26. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







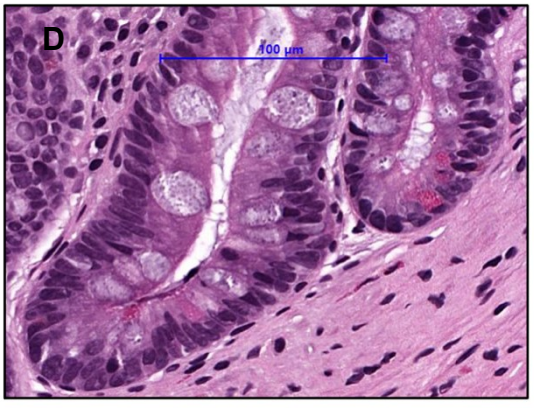


Figure 34: Case 27. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

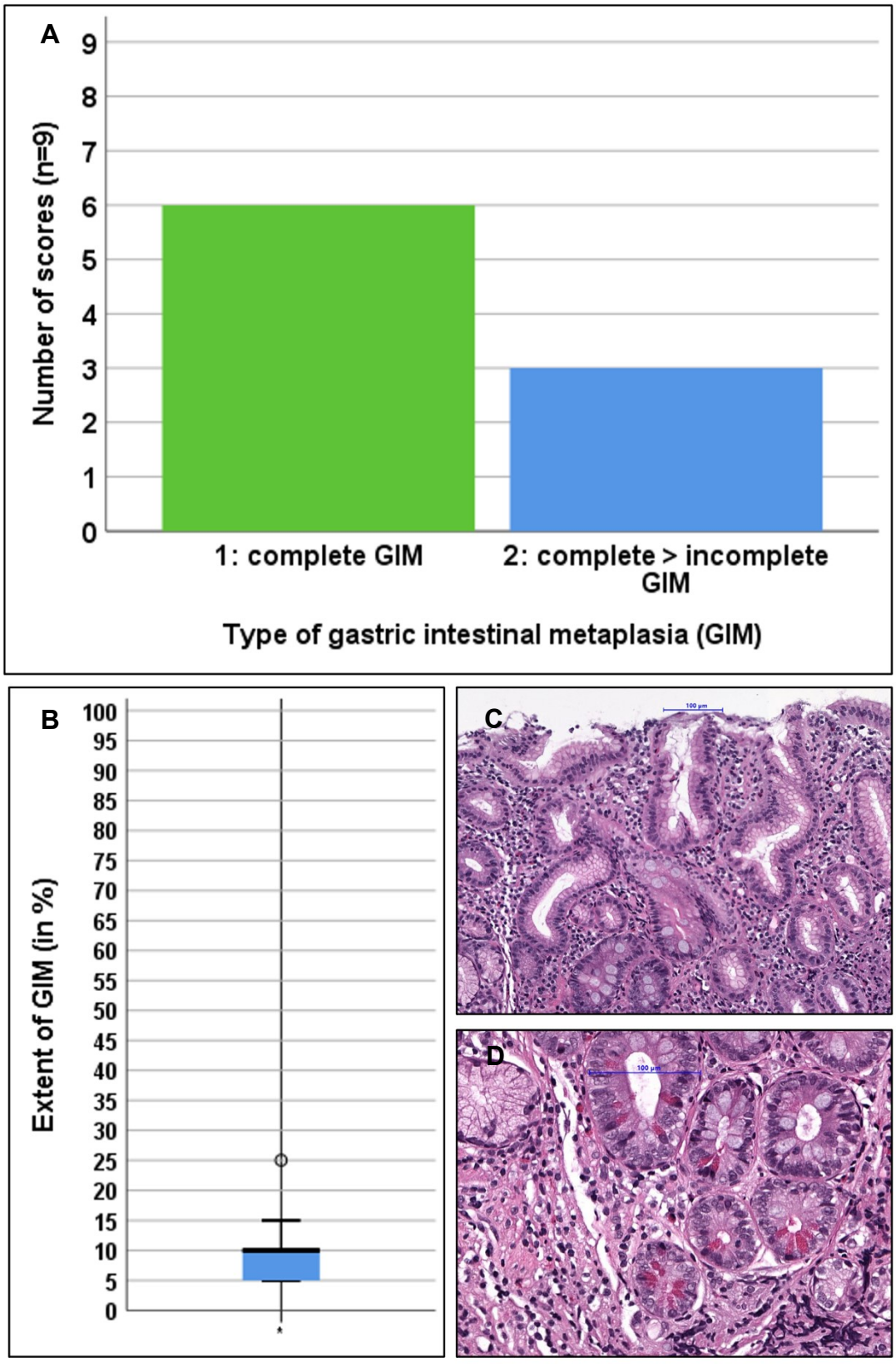


Figure 35: Case 28. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

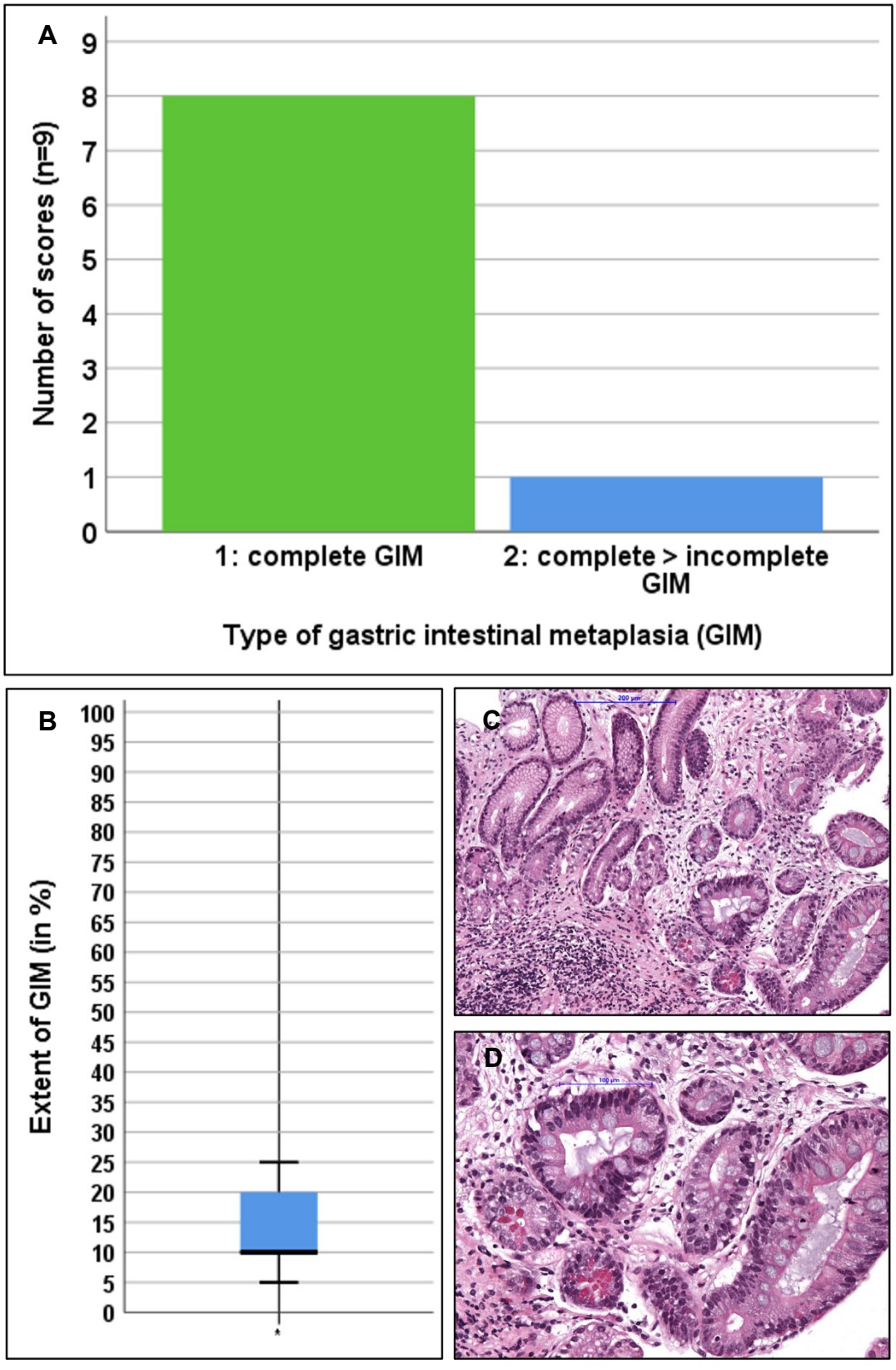
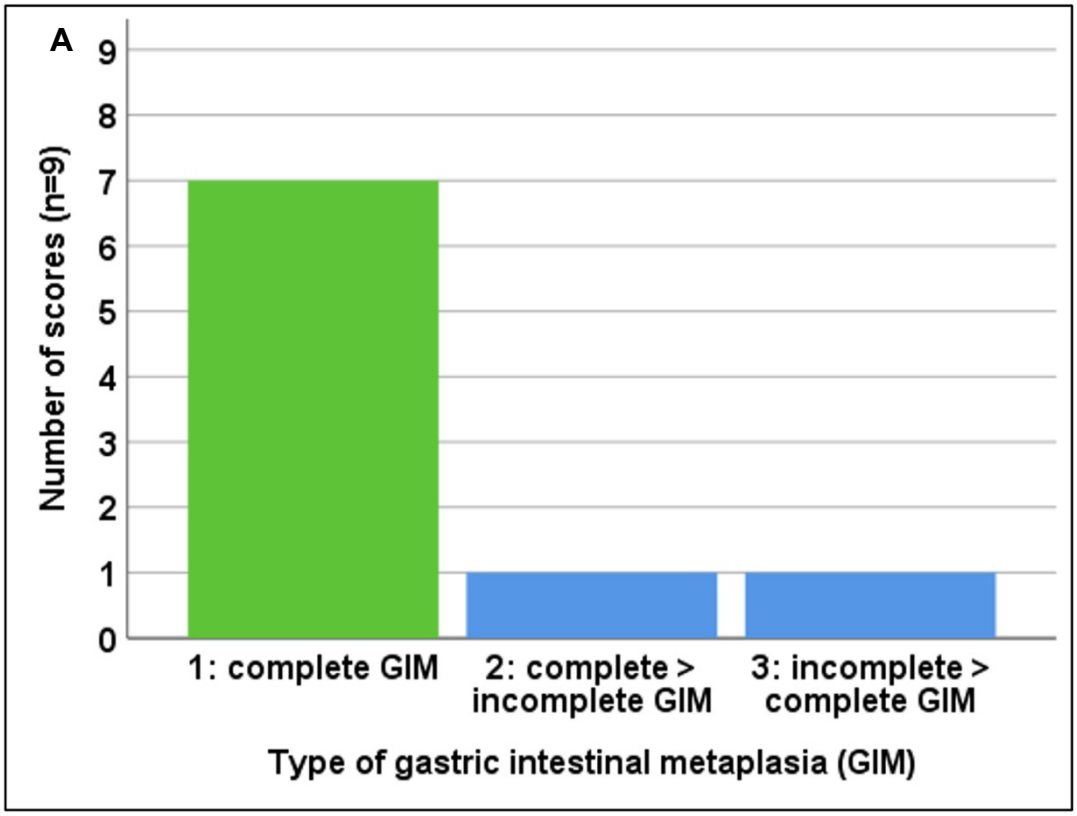
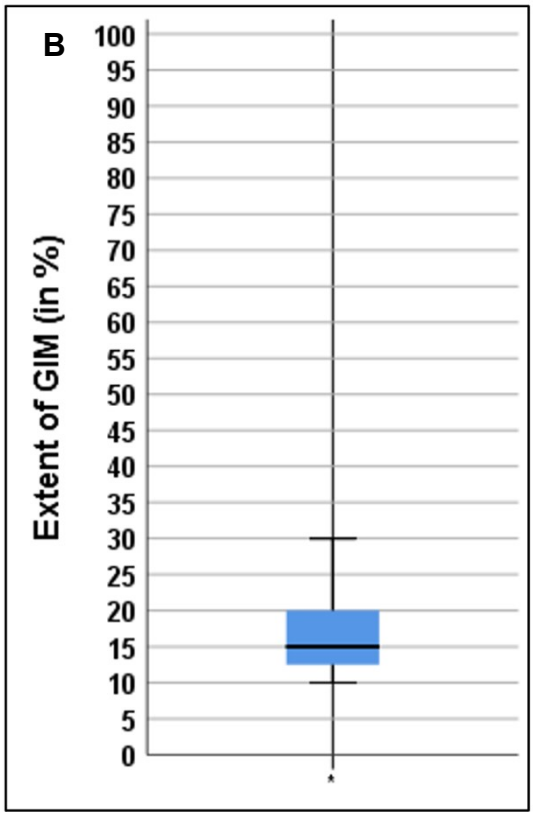
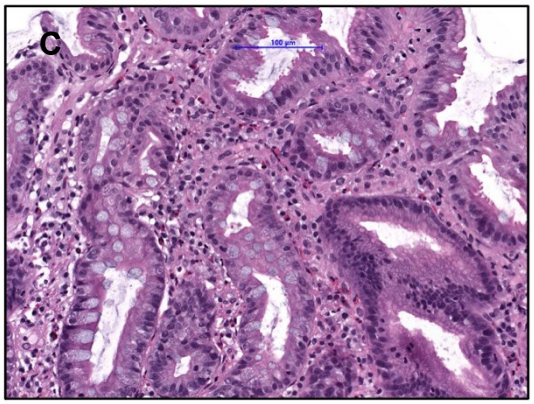


Figure 36: Case 29. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







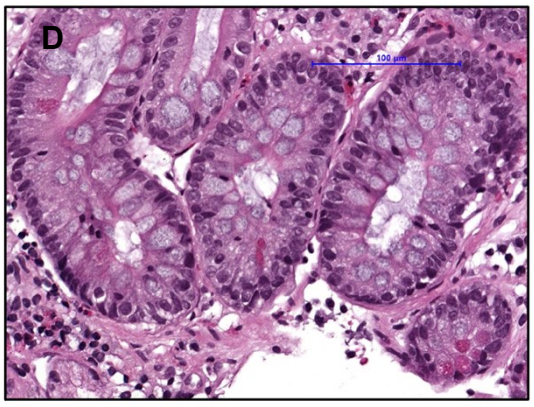
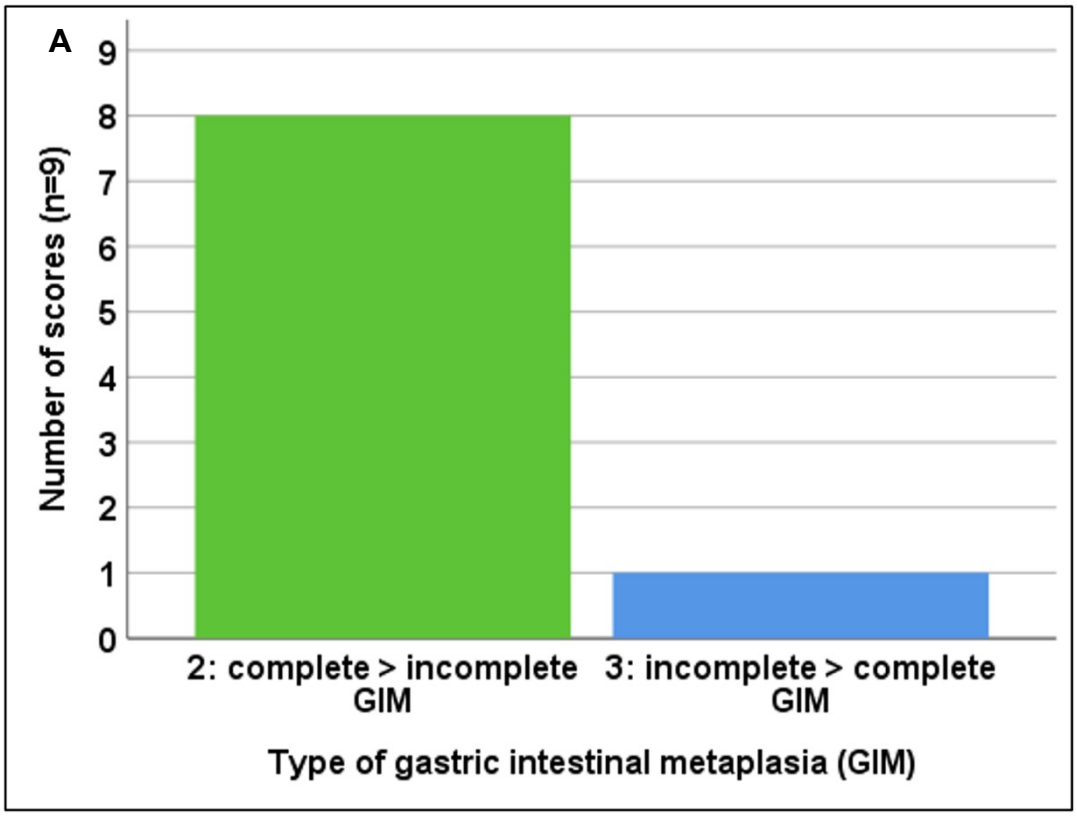
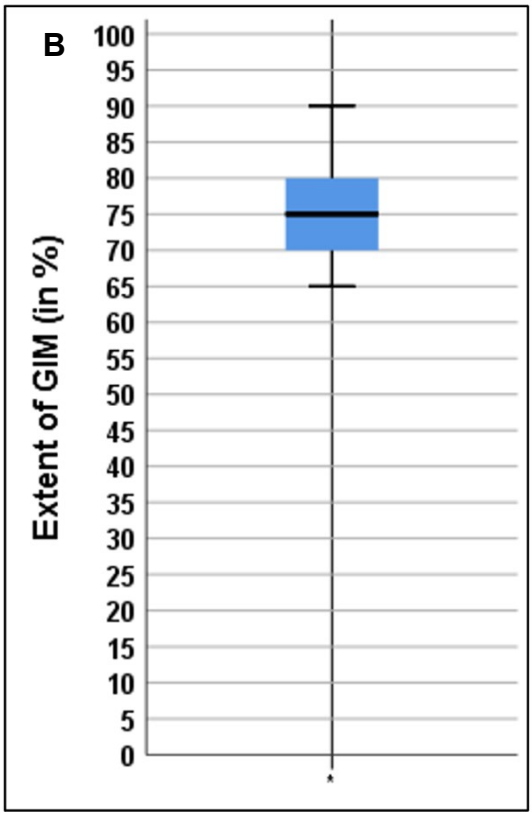
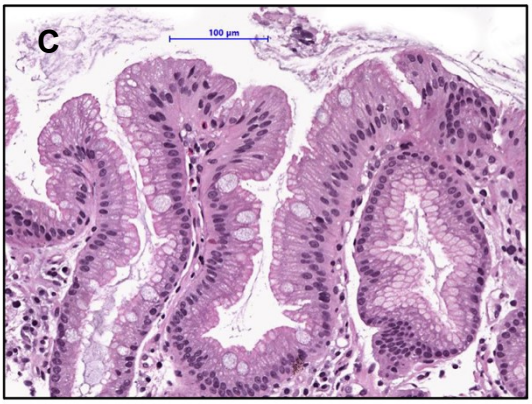


Figure 37: Case 30. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







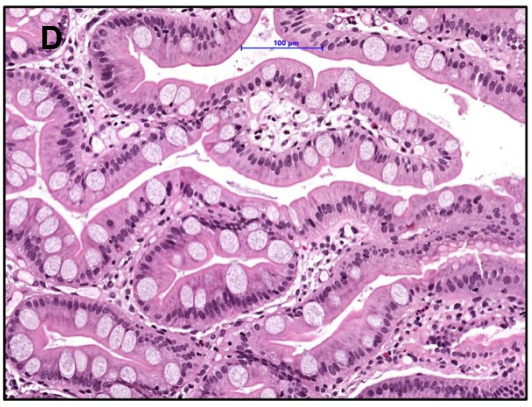


Figure 38: Case 31. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

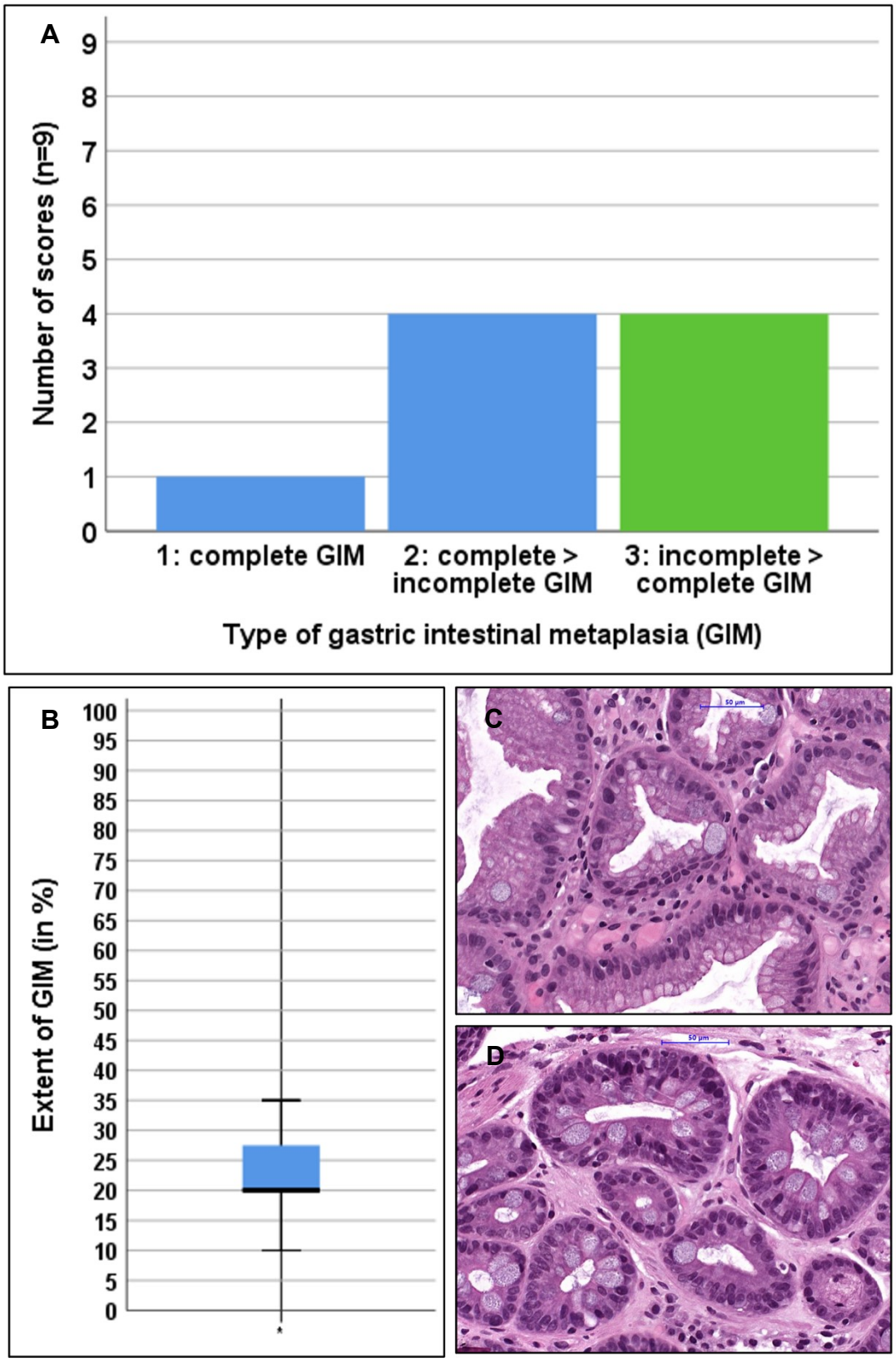


Figure 39: Case 32. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

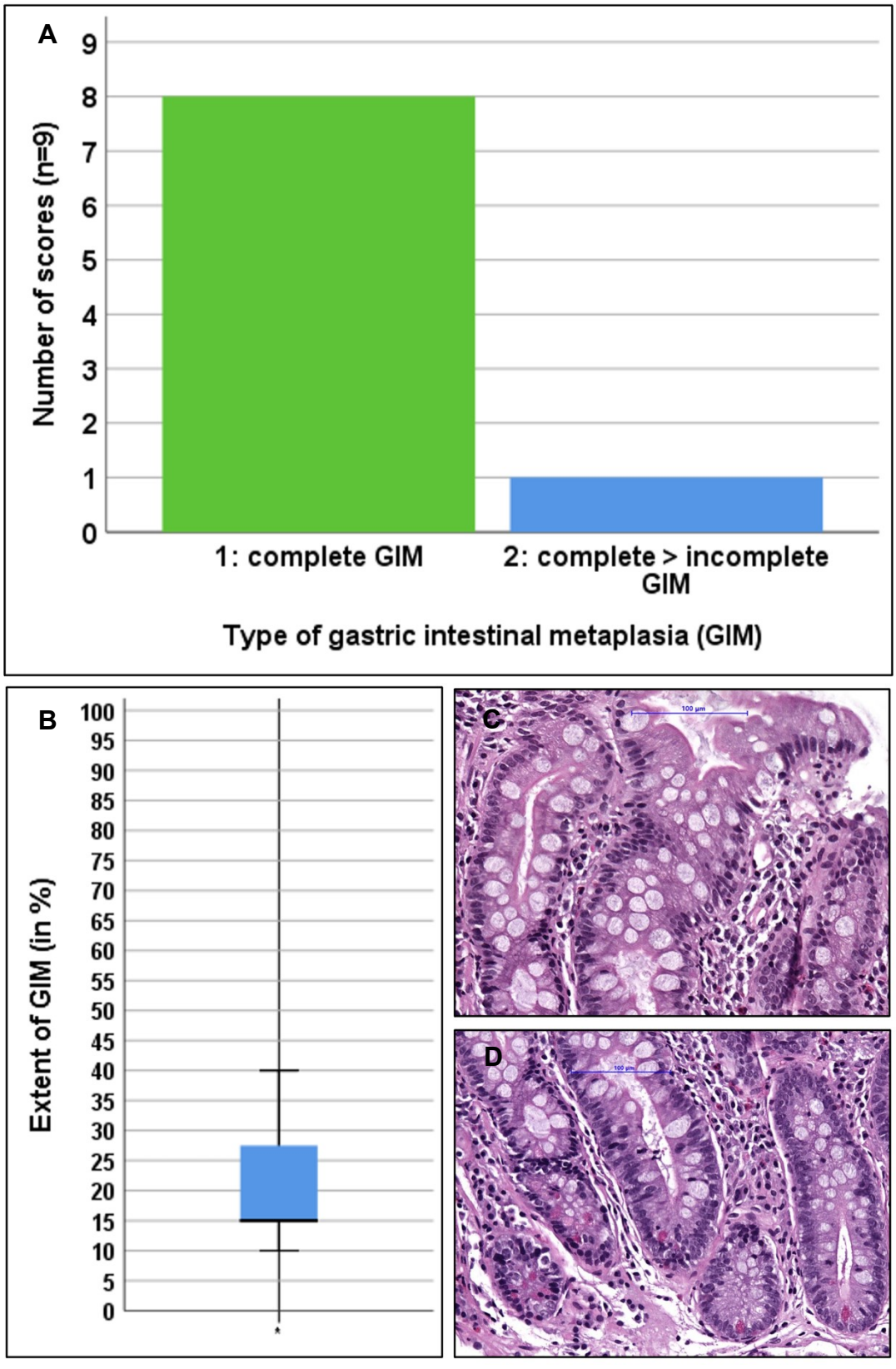
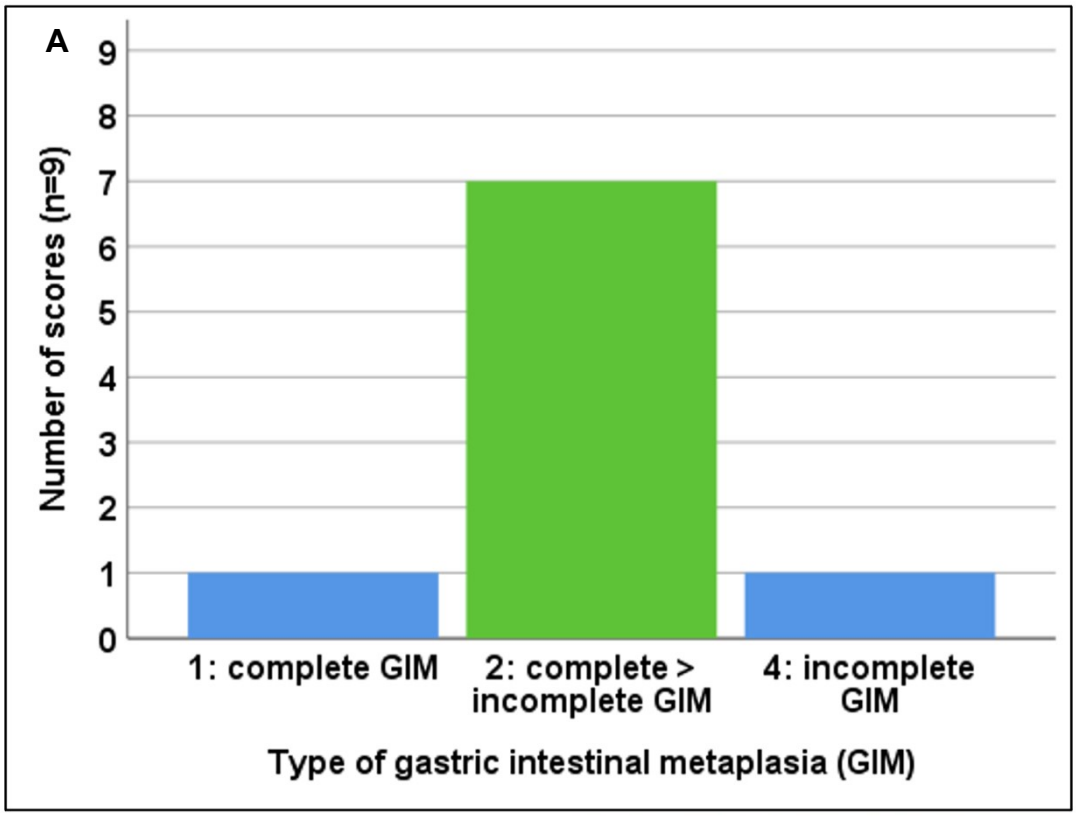
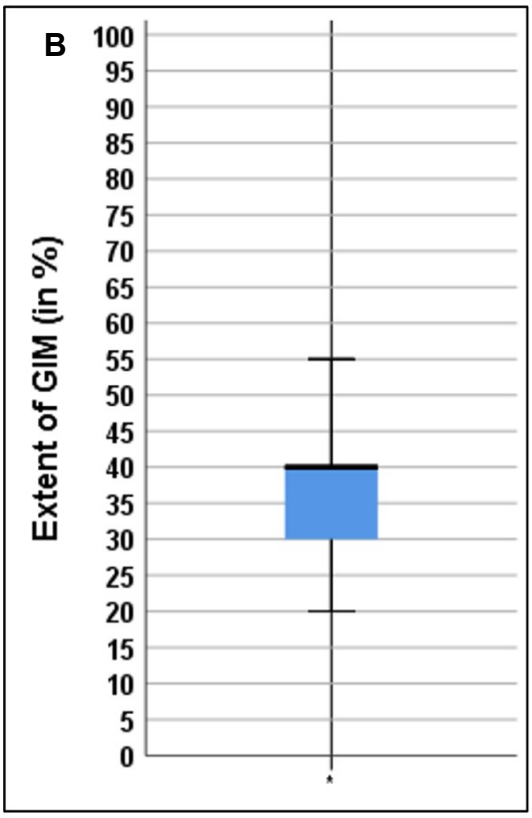
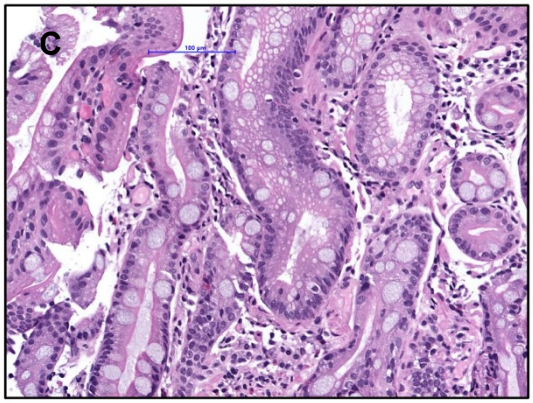


Figure 40: Case 33. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







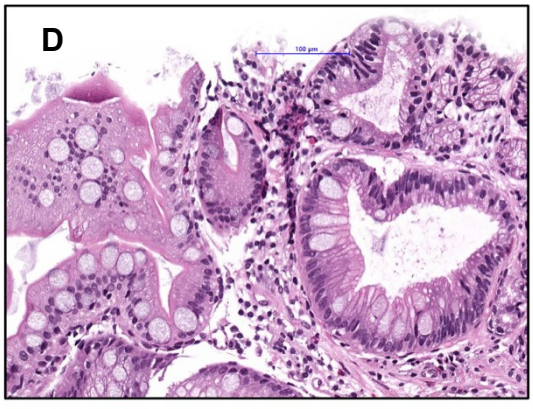
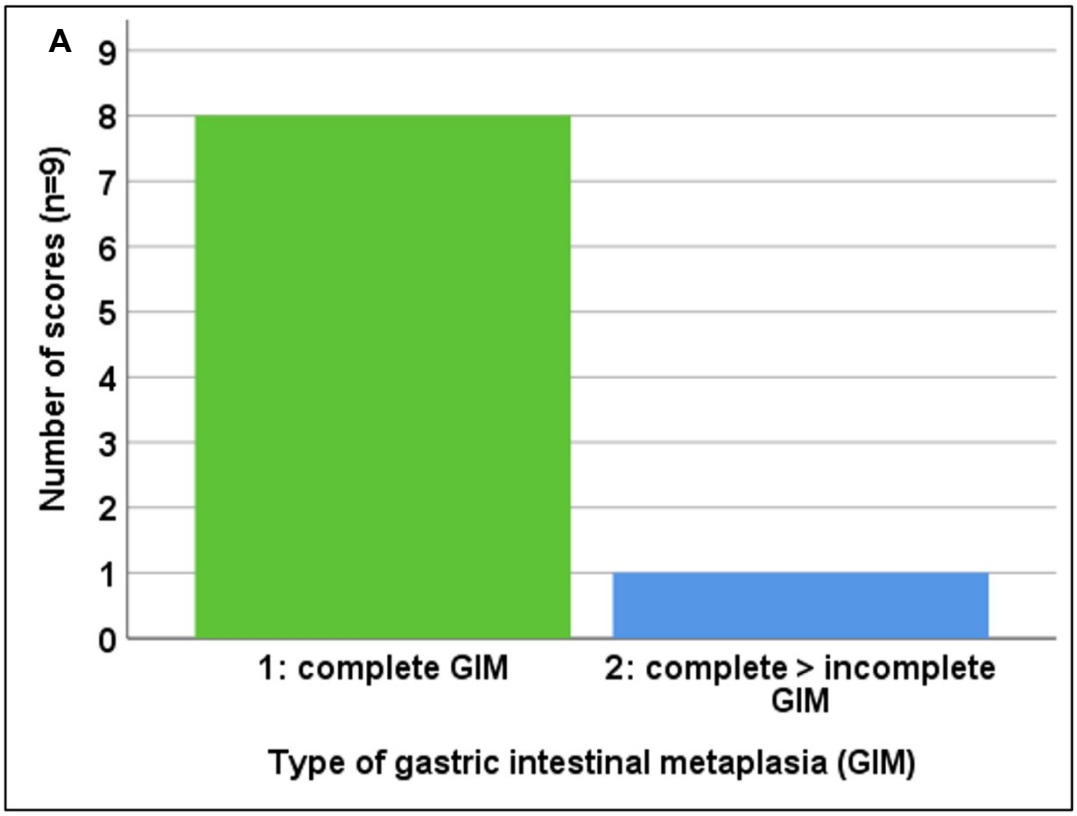
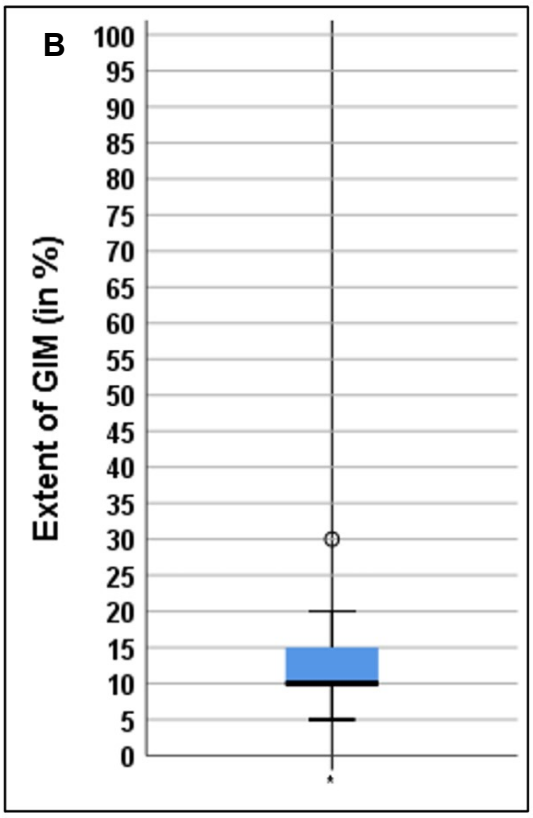
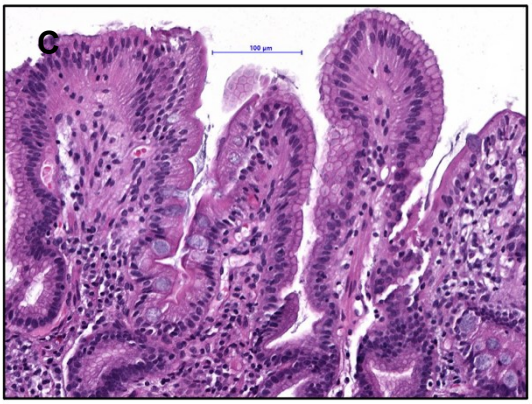


Figure 41: Case 34. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







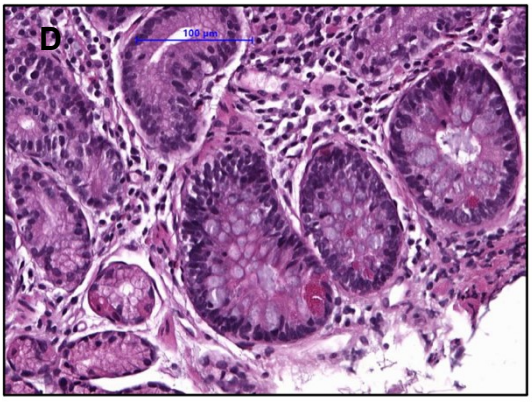
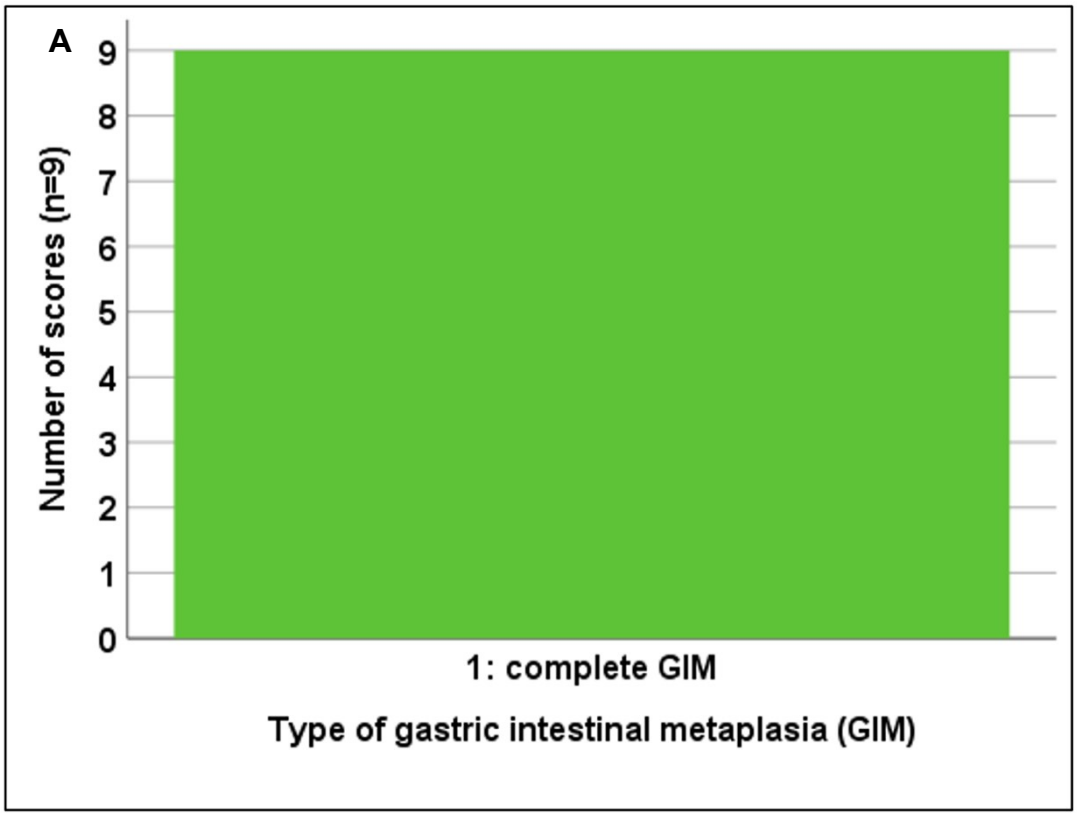
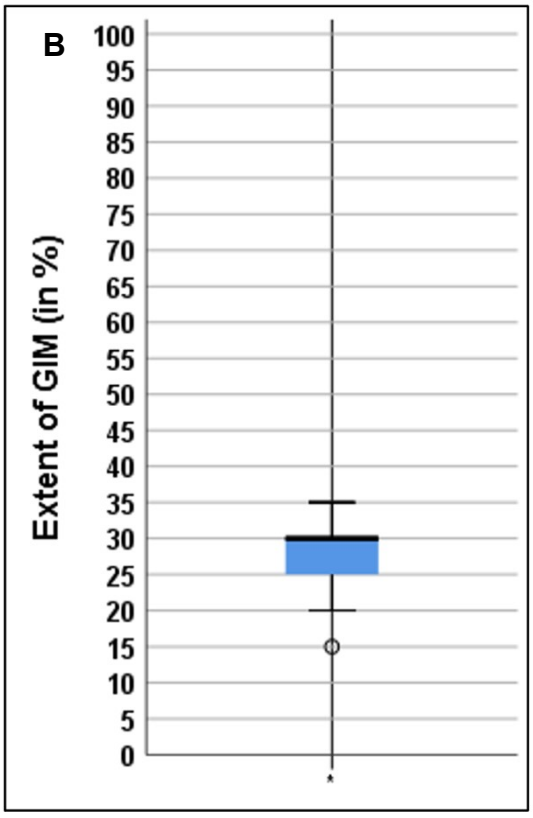
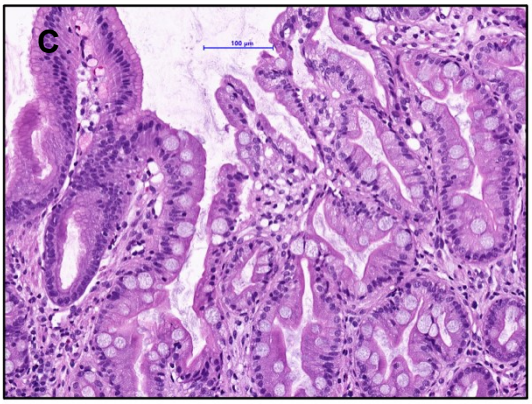


Figure 42: Case 35. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







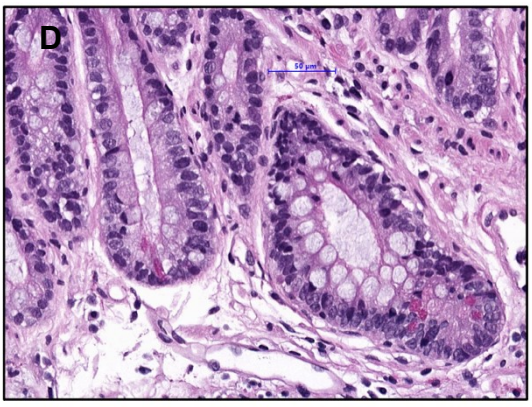


Figure 43: Case 36. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

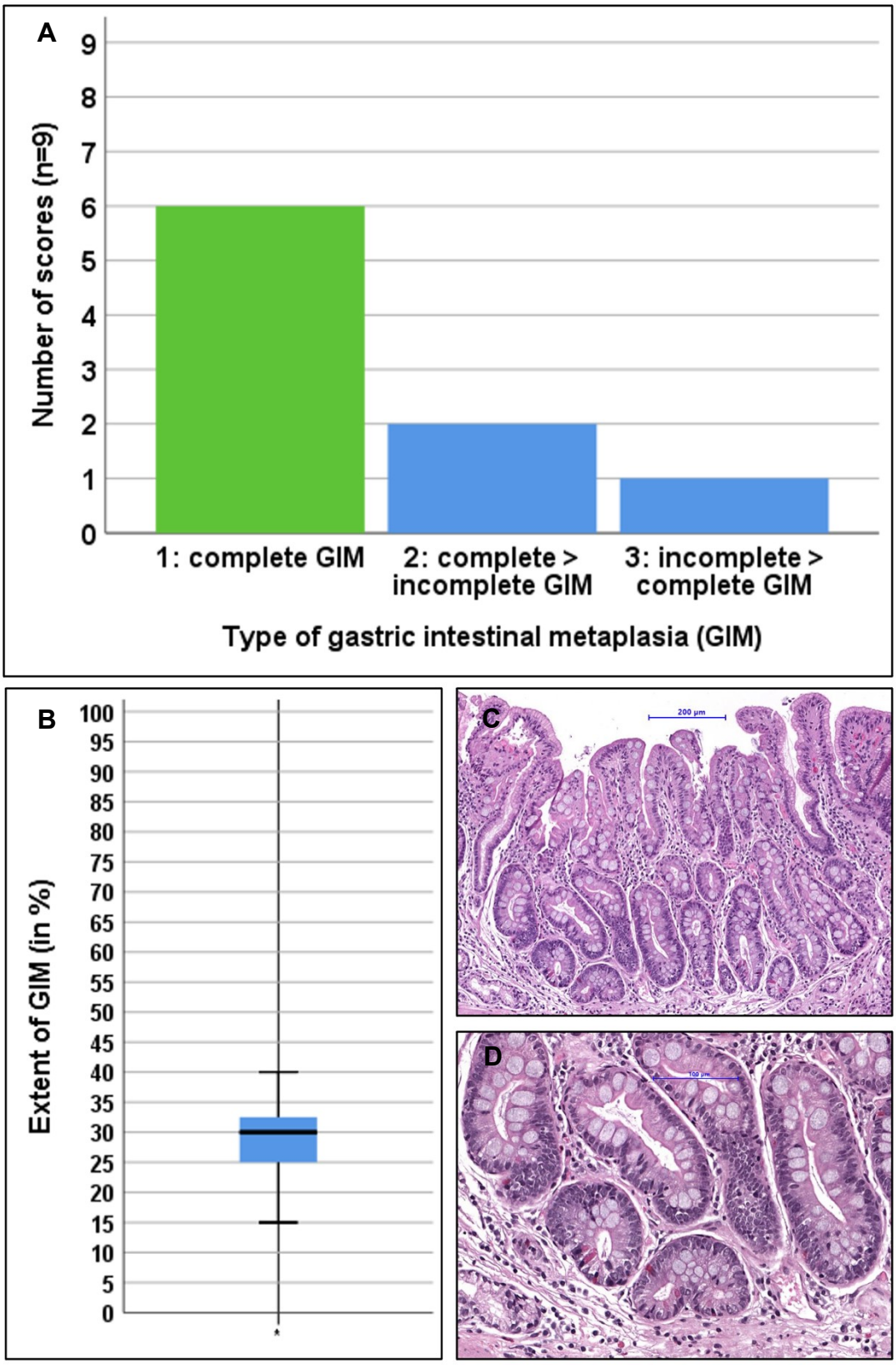


Figure 44: Case 37. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

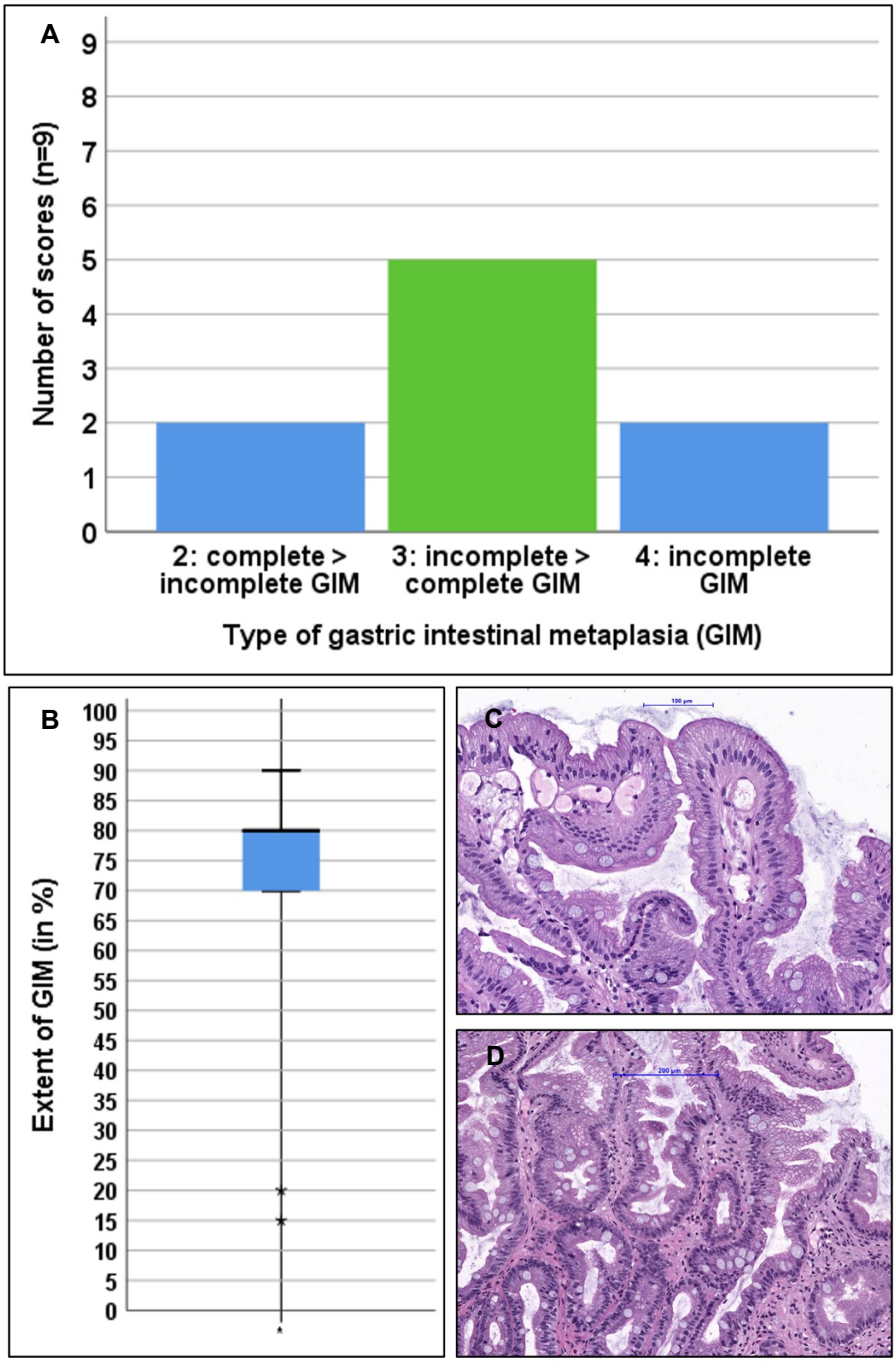


Figure 45: Case 38. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

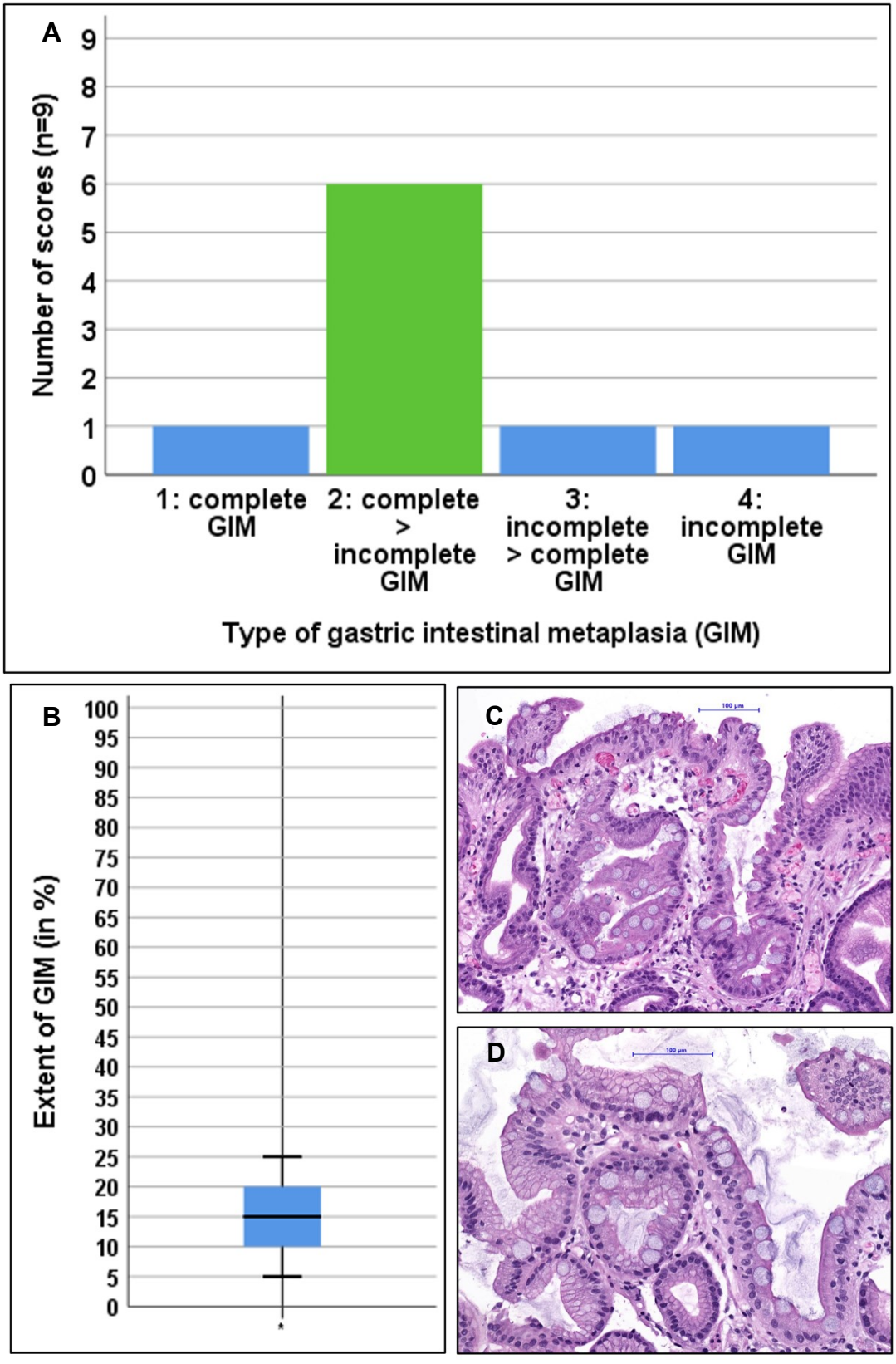
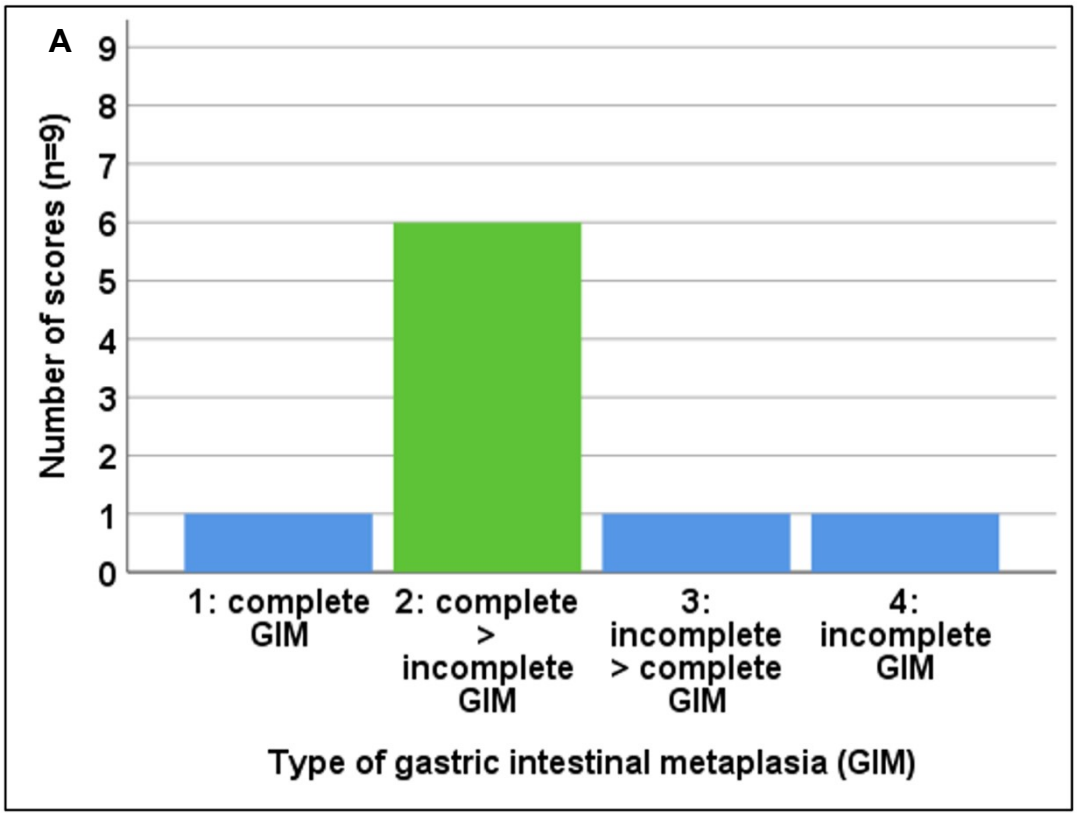
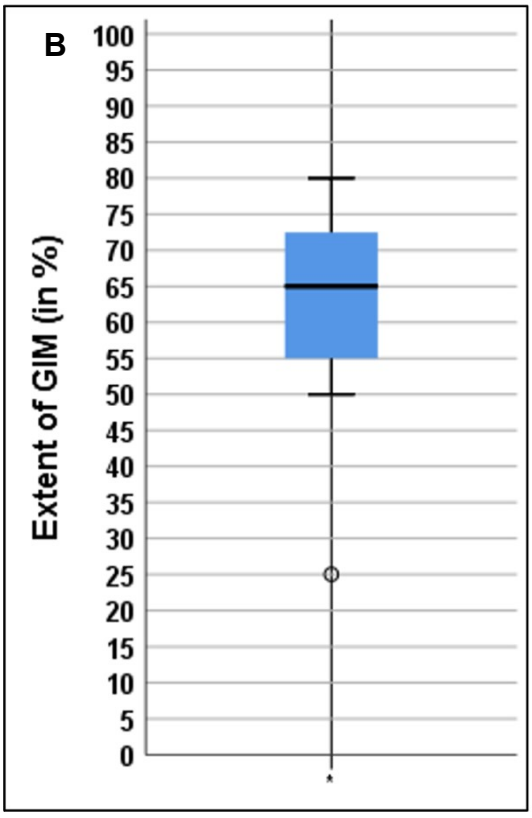
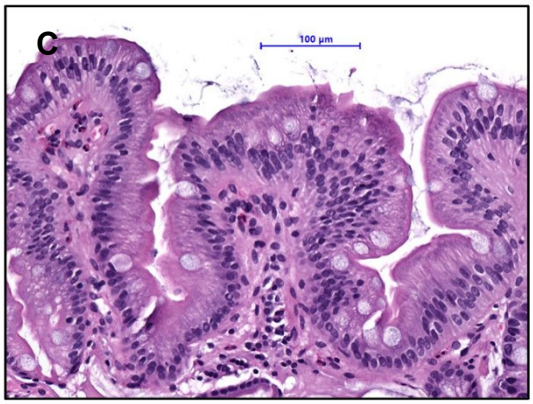


Figure 46: Case 39. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







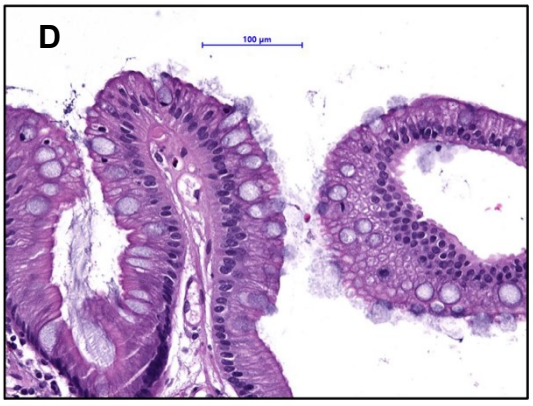
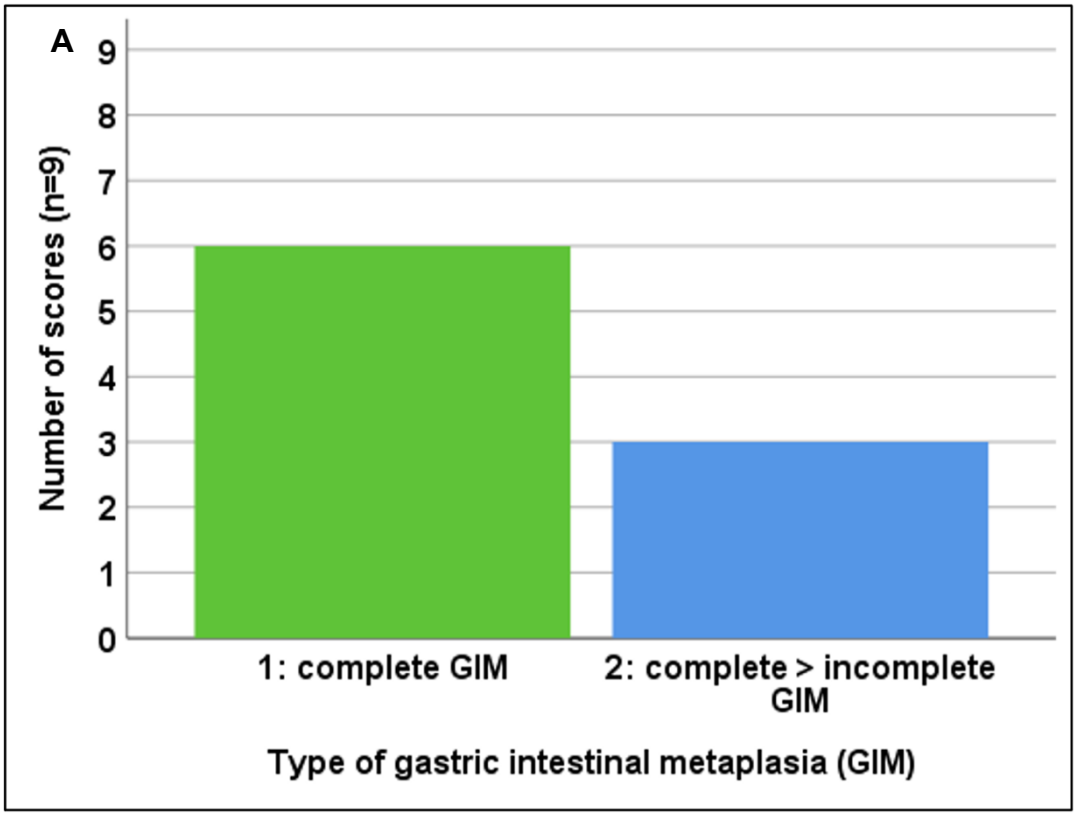
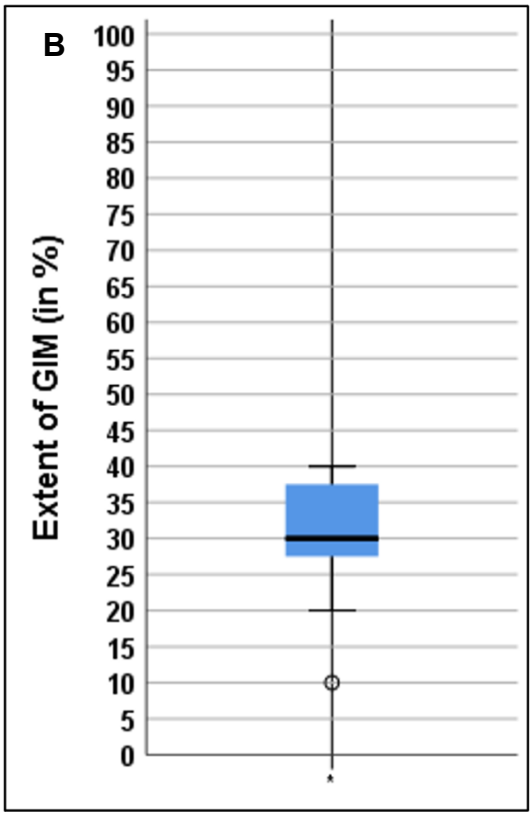
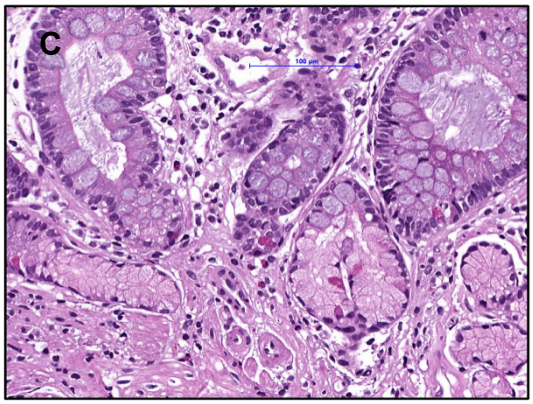


Figure 47: Case 40. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







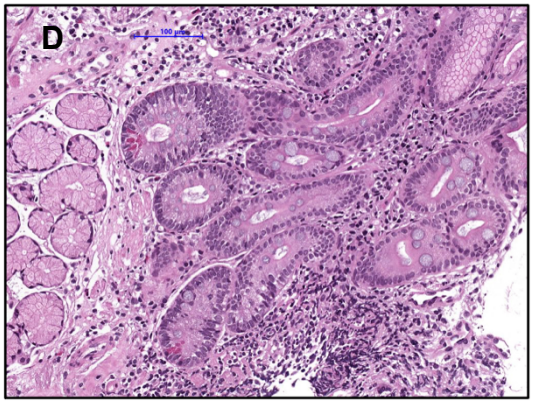
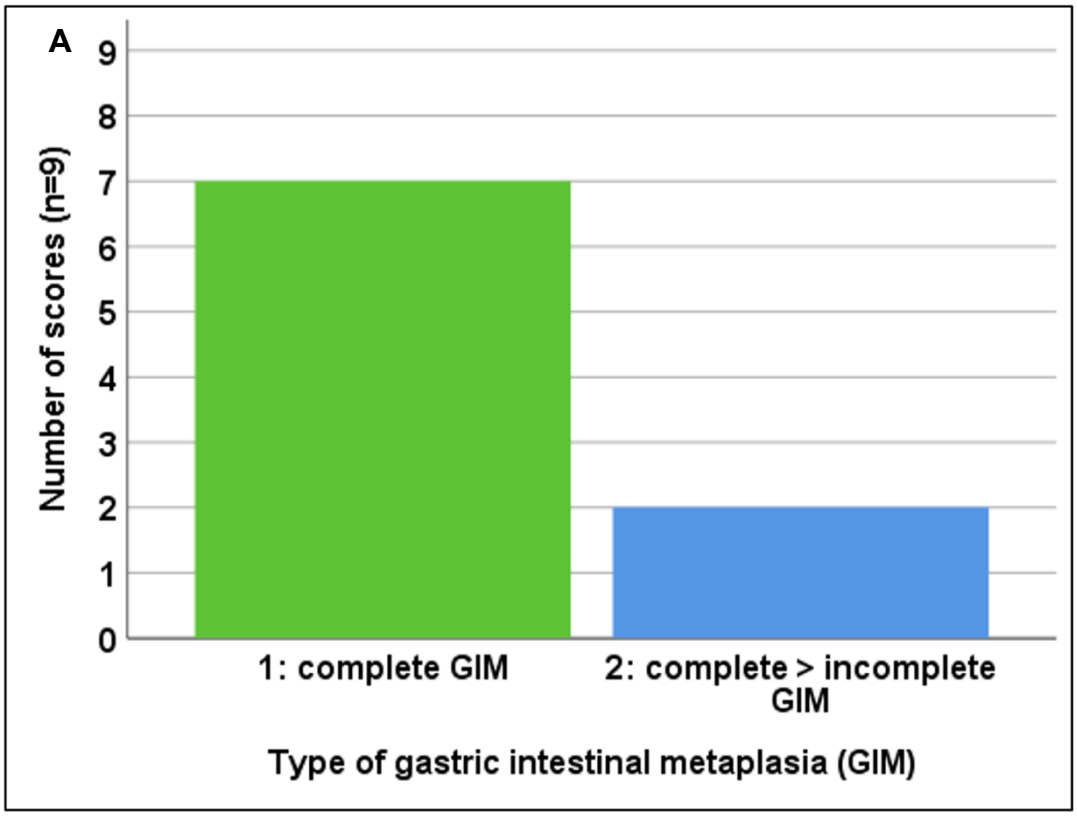
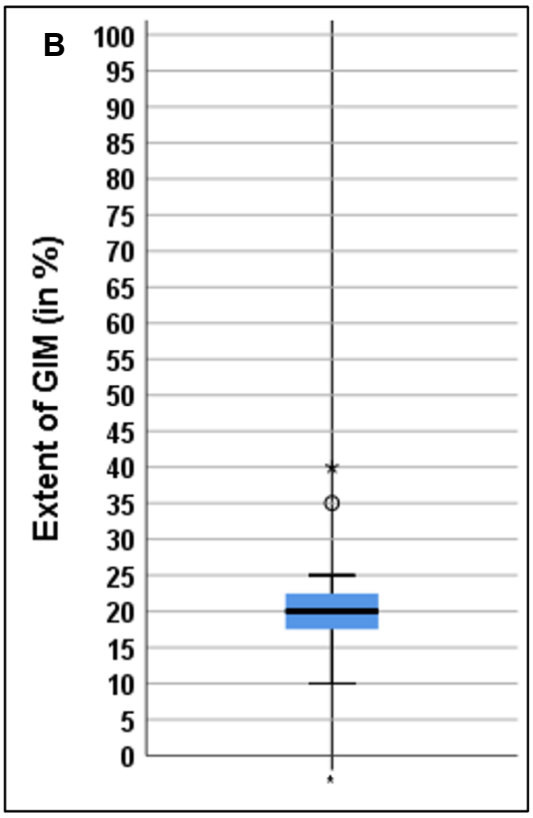
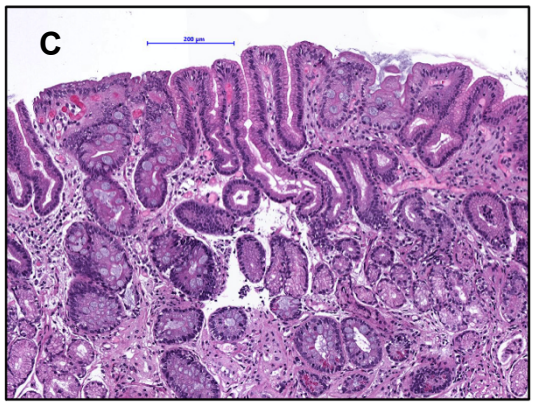


Figure 48: Case 41. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







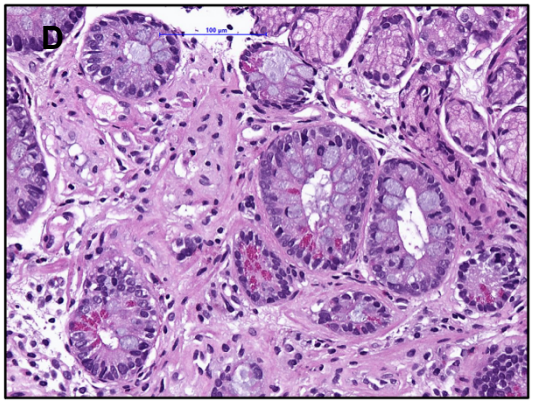
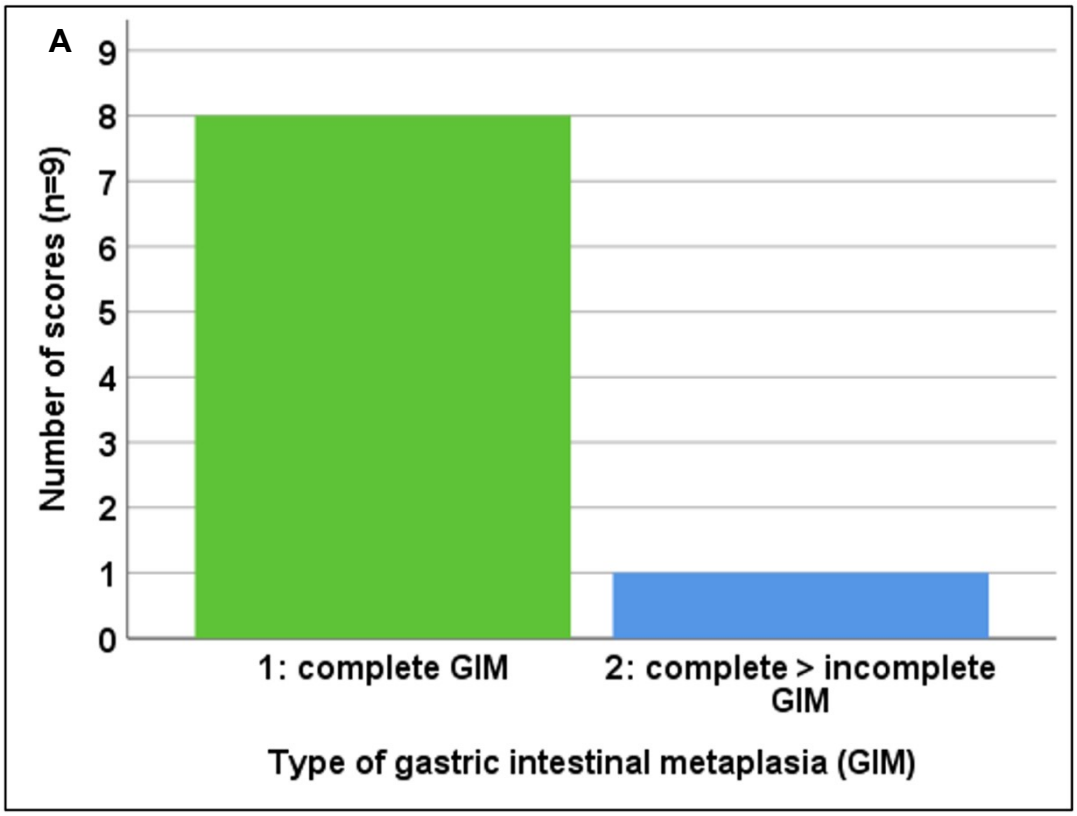
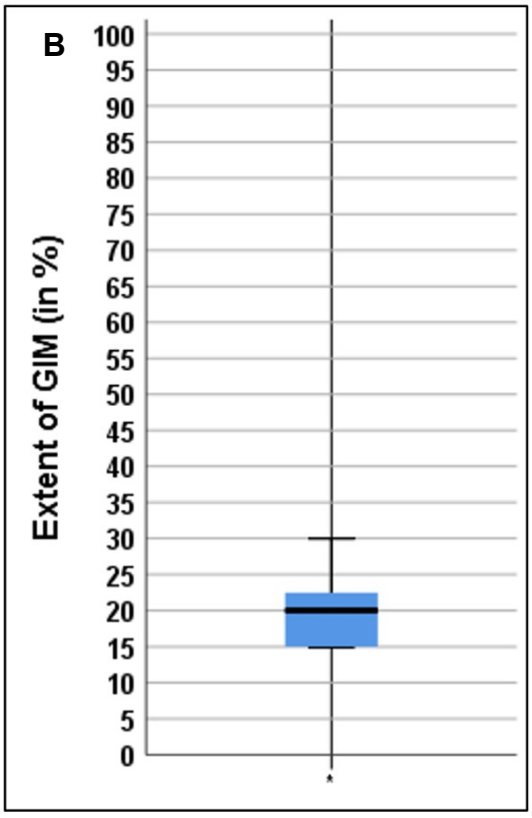
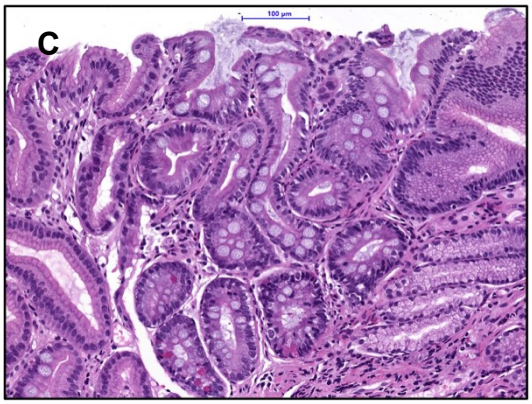


Figure 49: Case 42. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







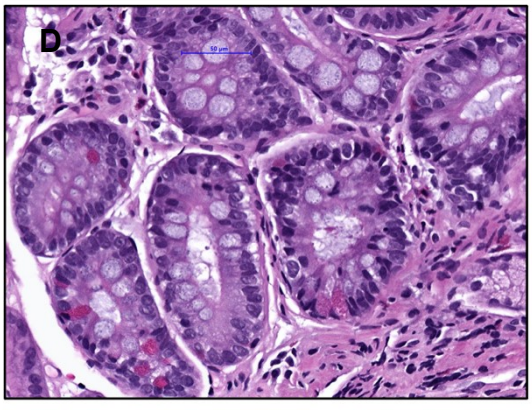


Figure 50: Case 43. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

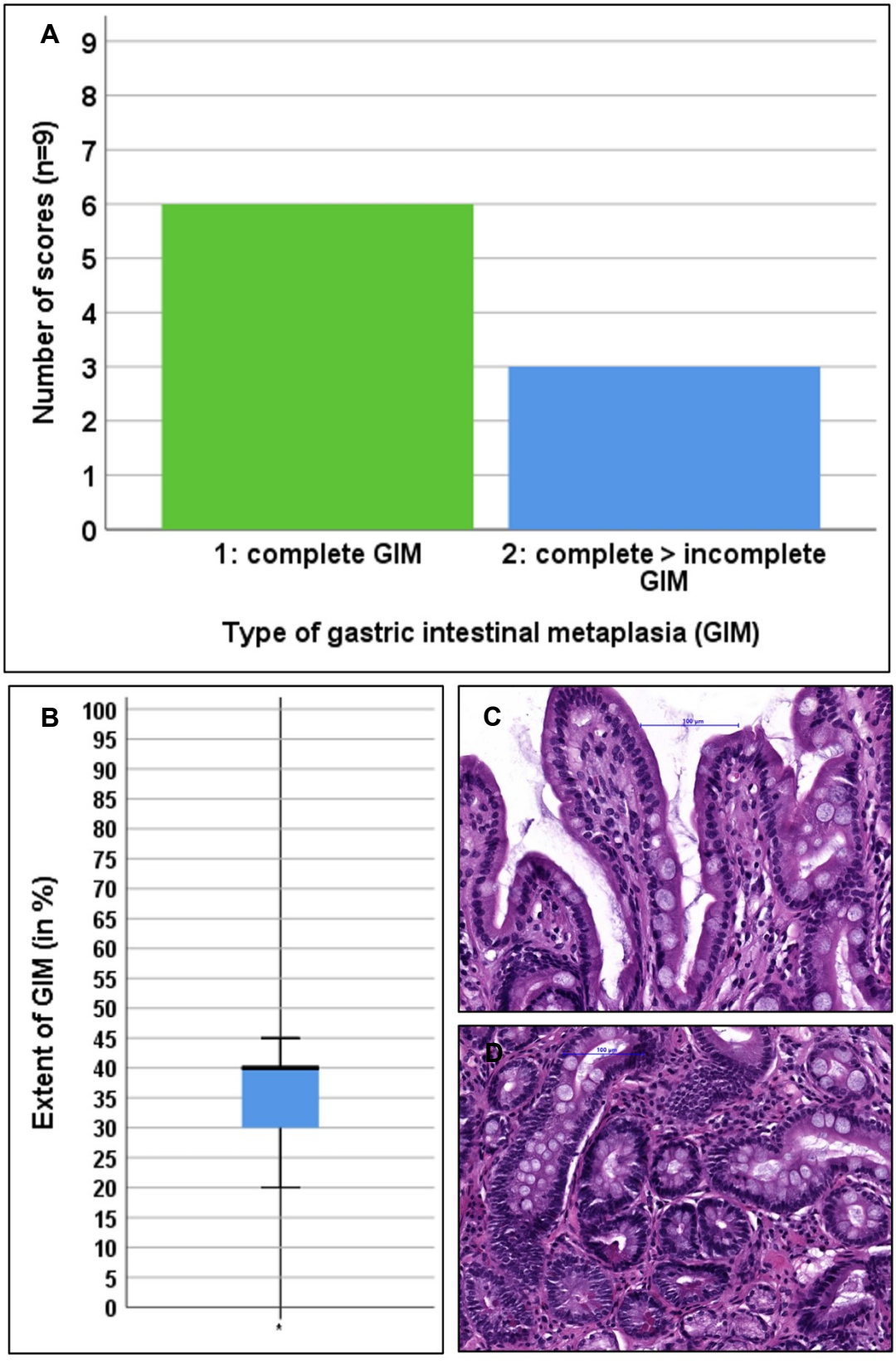


Figure 51: Case 44. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.

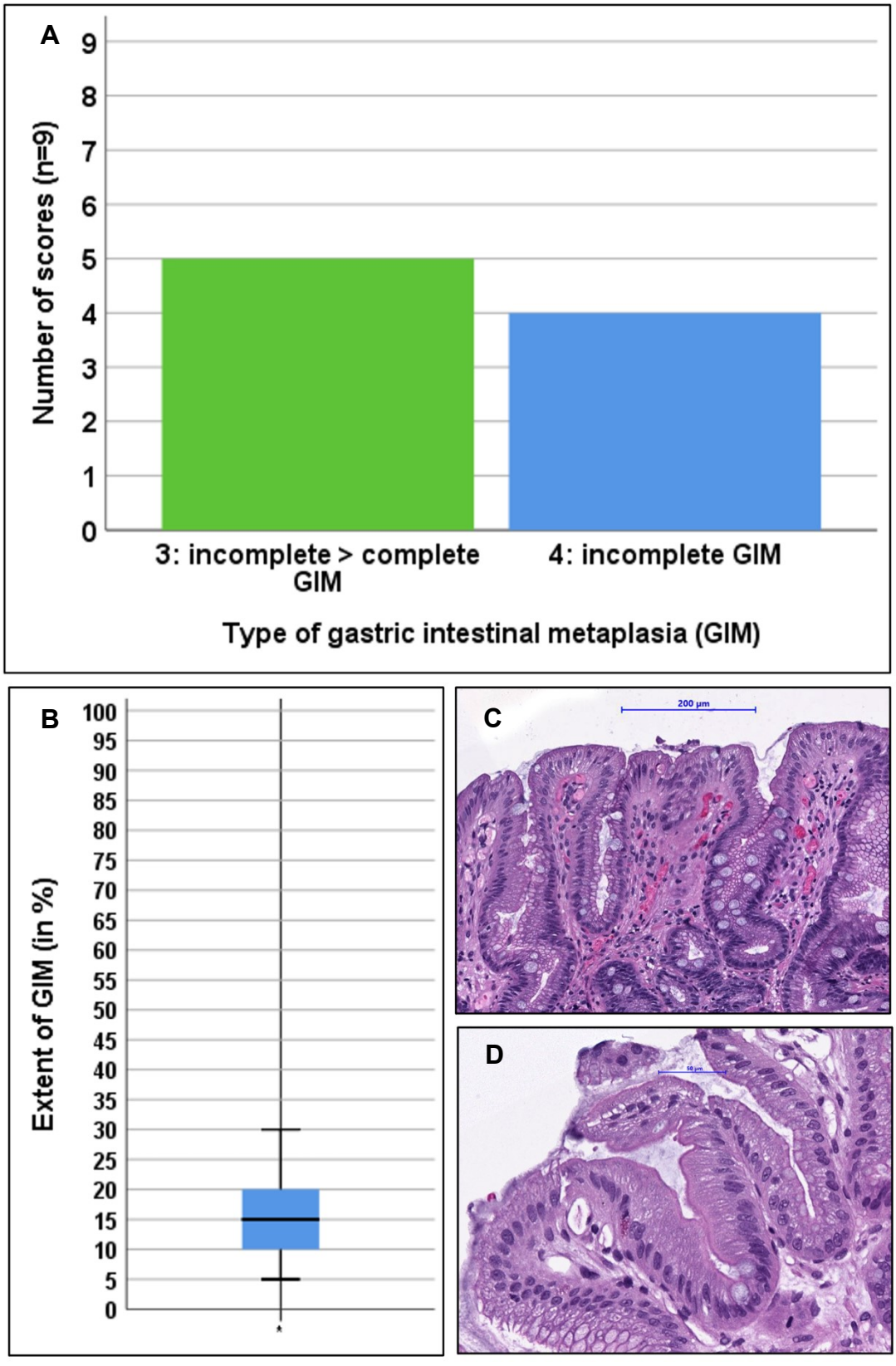
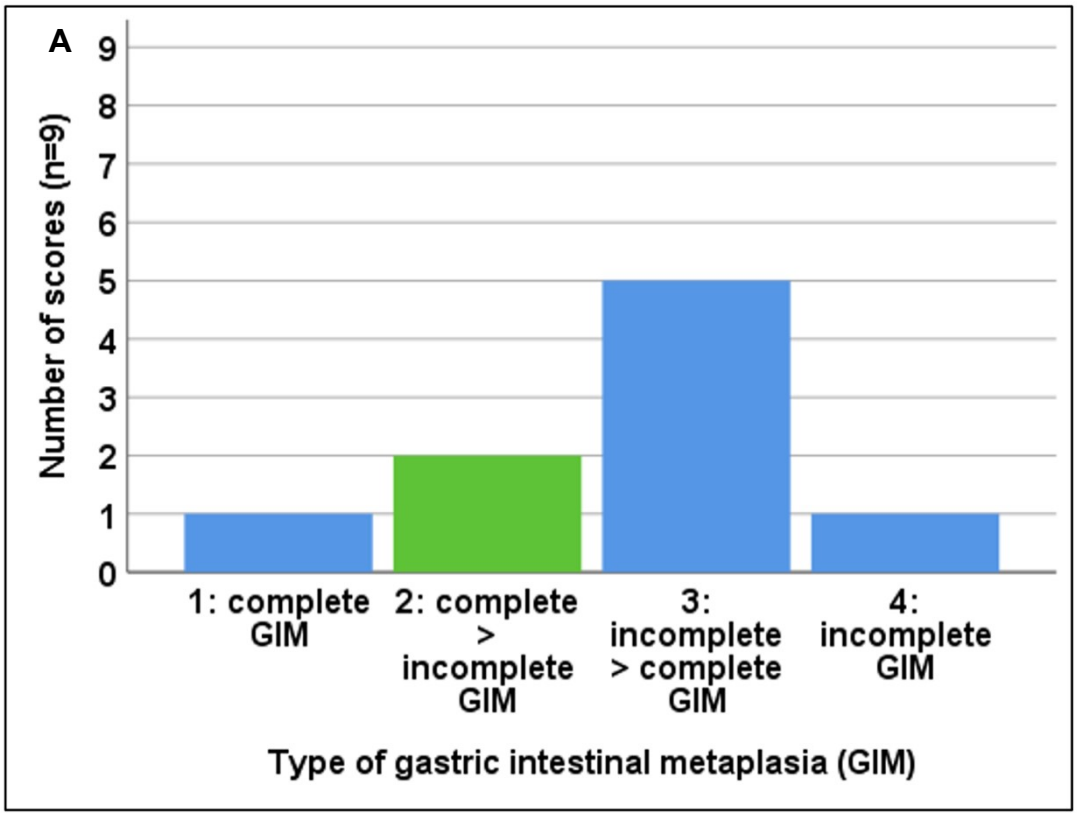
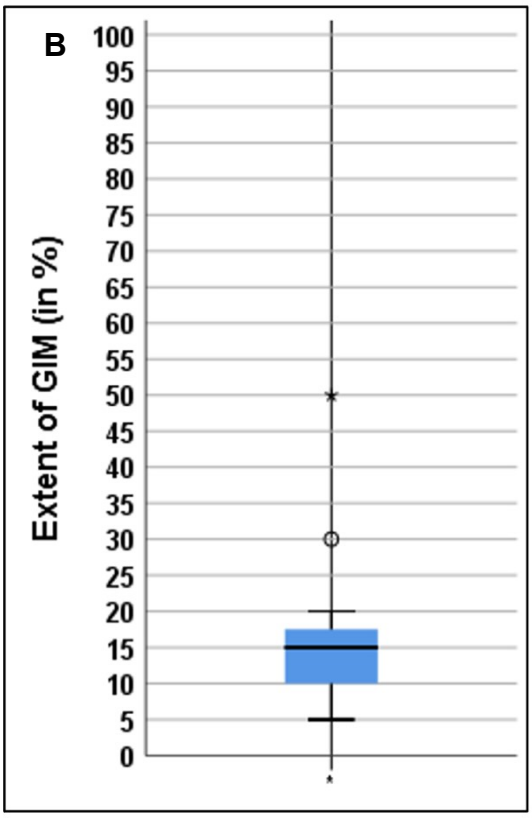
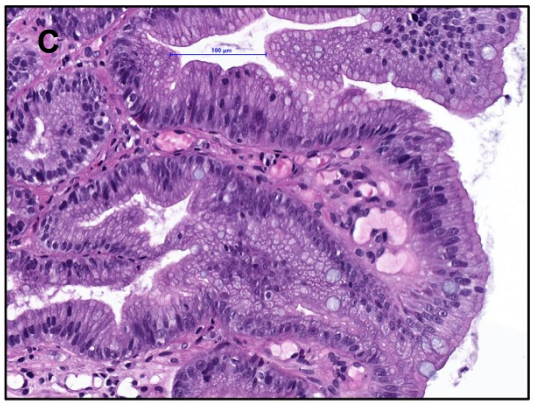


Figure 52: Case 45. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.







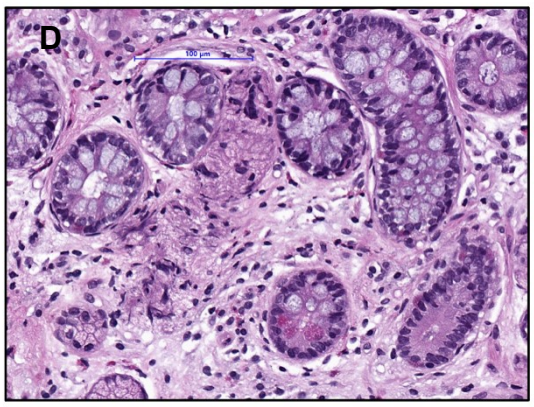
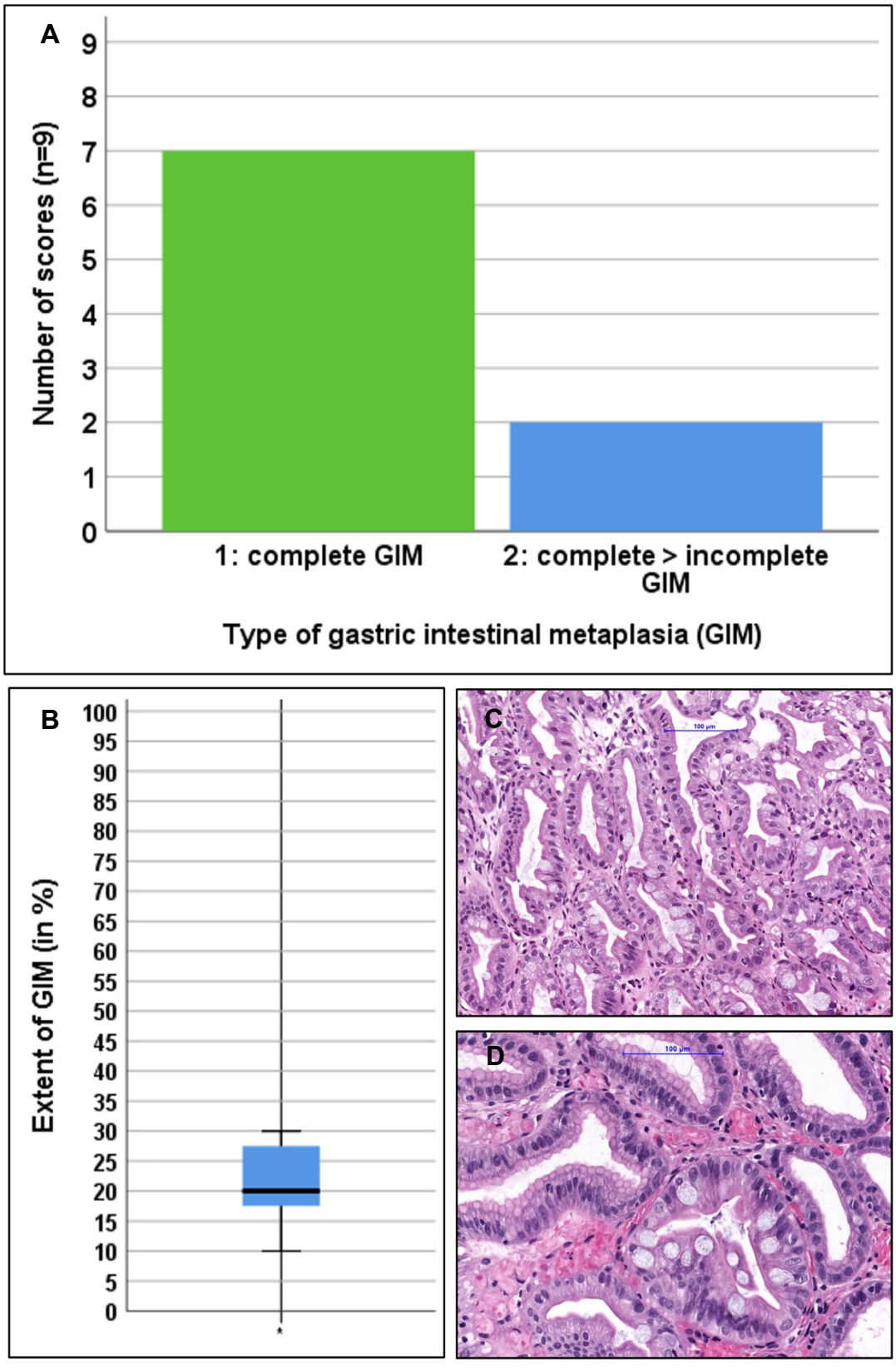


Figure 53: Case 46. (A) Nine observers assessed the gastric intestinal metaplasia (GIM) subtype. The consensus diagnosis is highlighted in green. **(B)** Eleven observers assessed the percental extent of GIM. **(C)** GIM at lower magnification **(D)** GIM at higher magnification.



4. Discussion

4.1 Interobserver variability in histological assessment

The present study demonstrates that eleven gastrointestinal expert pathologists were able to differentiate complete and incomplete GIM subtypes with substantial overall agreement (kappa=0.716, 95% CI: 0.677-0.755). Applying four categories of GIM (1: pure complete GIM *versus 2:* complete > incomplete GIM *versus 3:* incomplete > complete GIM *versus* 4: pure incomplete GIM), agreement slightly decreased with eleven pathologists reaching moderate overall agreement (kappa=0.447, 95%: CI 0.423-0.471). Agreement was significantly higher in cases with pure GIM types compared to cases with mixed GIM subtypes (p=0.010). Pathologists who apply subtyping in their daily routine performed better than those who do not (p=0.040). Misclassification with impact on clinical management decisions occurred in 5.7% (19 out of 333) of histological diagnoses. Additionally, this diploma thesis demonstrates that eleven gastrointestinal expert pathologists were able to estimate the percental extent of GIM with very good agreement (ICC 0.983, 95% CI: 0.975-0.990). Cases with a higher amount of metaplastic epithelium demonstrated a higher standard deviation (rs=0.644; p<0.01), suggesting lower diagnostic accuracy in cases with extensive GIM. .

4.2 Comparison with other interobserver variability studies

This diploma thesis is the first to systematically evaluate the interobserver variability in subtyping GIM on H&E-stained biopsy sections. To date, several studies assessed the interobserver variability in the diagnosis of atrophy and GIM (16,69,71,72), but only two studies included the assessment of GIM subtypes (84,88).

Cassaro et al. (88) classified 50 cases into complete or incomplete GIM based on staining with AB/HID. Cases where the incomplete subtype accounted for more than 30% of the total GIM were classified as incomplete GIM (88). The authors reported an absolute interobserver agreement of 93% with a kappa value of 0.91, suggesting almost perfect agreement (88). However, they did not render any data on the number of pathologists that were involved in this calculation. Presumed from the authors list, only two pathologists were included in this study (88). Even though their absolute agreement was higher than the initial absolute agreement of the two

pathologists providing the consensus diagnosis in our study, they only tested for two categories, namely complete versus incomplete GIM (88). The initial agreement of 83% (38 out of 46 cases) in our study was tested for four GIM categories and on H&E-stained slides. It may be of note that seven of the eight initially discordant cases were caused by either missing detection of a minor proportion of the second subtype or by divergent determination of the predominant subtype in mixed cases that showed a nearly even distribution of complete and incomplete GIM.

Isajevs et al. (84) invited four experienced pathologists to classify GIM into the traditional three types, that are, type I (complete GIM), type II (incomplete GIM) and type III GIM (incomplete GIM) (58), based on HID-PAS-stained samples of 166 patients from a high-risk population of Kazakhstan. Mixed cases were classified as incomplete GIM (84). The pathologists reached kappa values of 0.96 for type I, 0.88 for type II and 0.82 for type III GIM (84). They showed higher kappa values compared to our study, however, they included only four pathologists in this calculation, with three of them belonging to the same academic laboratory.

The above-described studies (84,88) provided data on the interobserver variability in GIM subtyping on specimens that were stained with AB/HID and HID-PAS. These stains are not used in routine pathological practice, they are technically challenging and contain potentially harmful and toxic iron diamine which led to their limited availability in many laboratories worldwide (57,85). Therefore, a satisfactory level of interobserver agreement has to be provided on H&E-stained slides, which are commonly used, do not afford additional costs and can be assessed in minimal due time (57,85).

Other interobserver variability studies in the field analysed agreement in the diagnosis of atrophy and GIM (16,69,71,72). In contrast to non-metaplastic gastric atrophy, interobserver agreement for the diagnosis of GIM was high, with kappa values ranging from 0.52 to 0.90 among gastrointestinal expert pathologists (16,69,71,72). Values for general pathologists were slightly lower (69,71,72). The study by Isajevs et al. (69) included two gastrointestinal expert pathologists and two trained pathologists who assessed 100 randomly selected H&E- and AB-stained tissue sections from patients with dyspeptic symptoms according to the OLGA and OLGIM staging systems. Agreement for atrophy in the antrum was moderate for expert pathologists (kappa=0.53) and slight for general pathologists (kappa=0.38) (69). Agreement for GIM in the antrum ranged from almost perfect agreement

(kappa=0.81) among expert pathologists to substantial agreement (kappa=0.68) among general pathologists (69).

Capelle et al. (16) investigated the interobserver agreement among three gastrointestinal expert pathologists in the grading atrophy, GIM and dysplasia according to the visual analogue scales of the Sydney classification. This study included 125 patients with the previous diagnosis of GIM or dysplasia and 20 patients with the diagnosis of GC (16). Consensus diagnosis was based on the majority diagnosis or on the mean score if all three pathologists disagreed (16). Interobserver agreement was slight for low-grade dysplasia (kappa=0.18), moderate for high-grade dysplasia (kappa=0.55), substantial for atrophic gastritis (kappa=0.64) and almost perfect for GIM (kappa=0.87) (16). Agreement regarding the OLGA system ranged from a poor to moderate level of agreement (kappa=0.19-0.56) (16). Agreement regarding the OLGIM system improved to a slight to almost perfect agreement level of agreement (kappa=0.31-0.88) (16).

Similar to the study within this diploma thesis, Leja et al. (72) invited two experts to reach consensus in the diagnosis of atrophy and GIM on H&E-stained slides. In contrast to the present study, agreement with the consensus diagnosis was evaluated by including the two expert pathologists and one experienced general pathologist (72). Kappa values for the detection of atrophy in the antrum ranged from 0.06 to 0.54 (poor to moderate level of agreement) and for the detection of GIM from 0.69 to 0.85 (substantial to almost perfect level of agreement), respectively (72). The general pathologist reached sufficient agreement with the consensus in detecting GIM, but not in detecting atrophy (72). This can be explained by the two expert pathologists' identical educational background and approach to assessing biopsy sections (72). The general pathologist used different criteria to define atrophy based on his experience and his interpretation of grading (72). While the expert pathologists regarded GIM as a phase within the evolution of atrophy, the general pathologist did not (72). The authors expected results closer to the consensus if the same criteria for the definition of atrophy would have been used (72).

A similar observation was found in a study by Kim et al. (71), where interobserver agreement for atrophy and GIM on H&E-stained slides significantly increased after the pathologists took part in a consensus process. Initial kappa values of 0.52 for GIM and 0.19 for atrophy suggested a moderate to poor level of agreement among five observers (71). After the consensus process, kappa values

increased to 0.68 (p=0.006) for GIM and 0.43 (p<0.001) for atrophy, suggesting a substantial and moderate level of agreement (71). The main responsible factors for the improvement of agreement were the definition of atrophy and the usage of percental cut-off values (30% and 60%) instead of using visual analogue scales for the grading of GIM (71).

The study of Guarner et al. (100) addressed interobserver variability in the application of the Sydney classification on H&E-stained slides. Two general pathologists assessed 74 cases and reached kappa values of 0.22 for acute inflammation, 0.60 for goblet cells, 0.53 for brush border cells, 0.81 for Paneth cells and 0.04 for atrophy, respectively (100). After the first assessment a joint review of discrepant cases took place (100). Assessment of another set of 76 cases improved agreement to kappa values of 0.50 for acute inflammation, 0.68 for goblet cells, 0.79 for brush border cells and 0.64 for atrophy, respectively (100). Agreement for Paneth cells decreased (kappa=0.71) (100). The authors compared this level of agreement to kappa values obtained by gastrointestinal expert pathologists (100). They concluded that the application of the Sydney classification would lead to minimal differences in interpretation and in kappa values greater than 0.80 (100). Due to the fact that pathology is eventually based on subjective interpretation, absolute agreement, i.e., kappa values of 1.0, cannot be reached (100). The authors regarded joint review of cases and the establishment of numerical parameters as the best strategies in improving diagnostic agreement (100). The demand for the use of percental cut-offs instead of visual analogue scales for the grading of GIM (71) and the wish for establishing numerical parameters in the diagnosis of GIM (100) meets the secondary objective of this study, investigating the interobserver variability in the percental assessment of GIM.

Subtyping GIM can be achieved with a similarly high level of agreement compared to other interobserver variability studies in the field. Although all study pathologists included in this diploma thesis were international gastrointestinal expert pathologists, only two of them reported to apply GIM subtyping in their daily routine practice. The two pathologists who routinely apply GIM subtyping achieved the highest level of agreement and the highest kappa values. This observation supports the study by Kim et al. (71) and Guarner at al. (100) who demonstrated that agreement in the diagnosis of GIM improves with increasing experience of pathologists.

4.3 Implications for clinical and pathological practice

4.3.1 Management and risk stratification according to the GIM subtype

The histologic subtype and extent of GIM play a major role in the risk stratification of patients with chronic atrophic gastritis (7,56). Cassaro et al. (88) concluded that if extensive GIM is present, no subtyping is needed, because the patient would be considered for surveillance due to his extent, irrespective of his subtype of GIM. The authors also emphasized that the presence of foci of incomplete GIM may be a marker for extensive GIM in cases with limited biopsy pieces or inconsistent mapping (88). This association supports the perception that the presence of incomplete GIM conveys a similar prognosis compared to advanced stages of OLGA/OLGIM, diminishing the importance of subtyping by replacing it with the staging systems OLGA/OLGIM as risk-stratifying tools (57).

Although the OLGA/OLGIM staging systems (15,16) are well established for the diagnosis of atrophy and GIM, management decisions that purely rely on these systems are not able to sufficiently detect individuals at risk of GC (84). A study by Isajevs et al. (84) investigated the prevalence of incomplete GIM among 166 participants from a high-risk population of Kazakhstan. In 12.6% of cases extensive GIM was detected (84). The overall prevalence of GIM was 45.8%, with 52.7% accounting for incomplete and 47.4% accounting for complete GIM (84). Even though it is known that incomplete GIM is associated with extensive GIM (5,57,85,88), this study proves that not only advanced stages of atrophic gastritis with extensive GIM, i.e., OLGIM III and IV, but also low risk stages, i.e., OLGIM I and II, can be affected by the incomplete subtype (84). Although Kazakhstan is considered as a high-risk GC country, most patients in this study were staged at OLGIM I or II and therefore would not have been scheduled for surveillance. However, the prevalence of incomplete GIM was 54.5% in cases with OLGIM I (n=66) and 37.5% in cases with OLGIM II (n=8) (84). While no patient with OLGIM IV was included in this study, the prevalence of the incomplete subtype in patients with OLGIM III was 50% (n=2) (84).

Currently, clinical guidelines by the ESGE warrant surveillance for patients with advanced stages of gastritis, i.e., OLGA/OLGIM III/IV, but not for patients with OLGA/OLGIM I or II (4). The BSG relies on the extent of GIM for management decisions instead of identifying incomplete GIM (89). The AGA has no specific

recommendations for patients with GIM, but does acknowledge patients with extensive GIM, incomplete GIM, a family history of GC or ethnic background at increased risk, potentially benefiting from short-time surveillance (90). According to these recommendations, the vast majority of patients with incomplete GIM in the study by Isajevs et al. (84) would have not been scheduled for surveillance.

In our study, all patients showed GIM restricted to the antrum (limited GIM). No surveillance would have been scheduled based upon GIM extent. However, in 13 out of 46 patients (28%) incomplete GIM is present alone or as predominant component. These patients would have been lost to follow-up, since in this subgroup the GIM subtype represents the sole determining factor for necessary surveillance endoscopy. Hence, subtyping is inevitable for detecting individuals at risk, especially in patients with GIM localized in the antrum and with no other present risk factors for GC except for the incomplete histologic subtype (57,84,88).

While the identification of incomplete GIM marks patients at risk of GC (8,57,84,85), it is currently unclear, if the absence of the incomplete subtype can be used for risk stratification in the same way. Therefore, the absence of incomplete GIM in patients with other risk factors present should not be used to “downstage” patients in surveillance strategy (57,77,90).

4.3.2 Relevance of mixed GIM subtypes

As incomplete GIM may develop *de novo* or from complete GIM (56), both types may coexist and are frequently found within one biopsy specimen (85). According to the ESGE (4) and AGA guidelines (90), patients with complete GIM limited to the antrum do not require surveillance, while patients with incomplete GIM do. It appears that the high proportion of mixed GIM cases is not taken into account. The ESGE refers to the publications by Gonzalez et al. (82,83,85) who established the prognostic impact for patients with pure incomplete or predominant incomplete GIM but did not consider patients with a minor incomplete component (4). The AGA states that patients with a higher risk for GC include those with “at least partial” incomplete GIM (90). For our study analysing interobserver variability and potential clinical impact of misclassification, we did not consider cases with a minor incomplete component as this situation warrants further scientific efforts (57,101).

A study on the prognostic importance of GIM extension, topography, and type (5) suggested, that a prognostic valuable cut-off value for the extent of incomplete

GIM can be either set at 10% or 20%. Patients who showed ≥10% compared to <10% incomplete GIM at baseline endoscopy had a hazard ratio (HR) of 3.10 (95% CI: 1.31-7.34) and a risk rate of 2.44 (95% CI: 1.78-3.36) per 100 persons per year for neoplastic change (5). Patients that showed ≥20% compared to <20% incomplete GIM at baseline endoscopy had a HR of 2.82 (95% CI: 1.45-5.47) and a risk rate of 2.85 (95% CI: 2.02-4.04) per 100 persons per year for neoplastic change (5). However, the authors of this study argue that subtyping GIM would require additional histochemical stains and costs and that its focality may lead to sampling errors (5). Therefore, Tava et al. (5) recommend using a cut-off value of ≥20% total GIM extent to identify patients at risk of GC.

In addition to the histologic prognostic value of incomplete GIM, molecular profiling was able to reveal factors that differentiate incomplete from complete GIM and may explain its increased malignant potential (102). Huang et al. (102) investigated 138 GIMs of GC-free individuals. Molecular profiling revealed that patients with incomplete GIM had shortened telomeres and chromosomal alterations that have been associated with subsequent progression to dysplasia and GC (102). Still, no statistically significant difference for the shortened telomeres, an increase in the mutational burden, HP density, or DNA methylation levels in patients with incomplete GIM (p=0.16-0.17) compared to patients with complete GIM was noted (102). The lack of statistical significance, however, could be related to insufficient power in this secondary analysis (57). Somatic copy number alterations (SCNA) were detected in 13% of cases with incomplete GIM, whereas none were detected in cases with complete GIM or in cases with a normal gastric epithelium (102). SCNAs can lead to deletion and amplification of tumour suppressors and oncogenes (45). Consistent to these findings, Tan et al. (45) outlined that SCNAs in genes such as KRAS contribute to the progression from incomplete GIM to dysplasia and that telomerase activation and telomerase reverse transcriptase (TERT) expression takes place within this step of the precancerous cascade (45,103,104).

4.3.3 Implementation of GIM subtyping in routine practice

While GIM subtyping is routinely performed in several European countries, others, such as the US, have not yet implemented GIM subtyping in daily routine practice (57). This observation has given rise to concerns as to whether the histologic

subtype of GIM can be utilized as part of risk stratification without a substantial educational initiative for pathologists (90). This diploma thesis and the studies by Kim et al. (71) and Guarner et al. (100) indicate that the histopathological knowledge required for diagnosis on H&E-stained slides can be gained in due time and applied within the routine diagnostic work-up without additional costs. This notion complies with the recently published guidelines by ESGE and AGA who both recognized incomplete GIM as an important risk-stratifying feature but contrasts with the BSG guidelines that have not yet implemented GIM subtyping in decision-making (4,89,90).

Even though some authors may argue that subtyping GIM might require not commonly used immunohistochemistry techniques (85), this and previous studies showed that GIM cases can be classified on H&E-stained slides (57,84,85). González et al. (85) do not recommend identifying sulfomucins with the HID-stain, because the morphologic criteria on H&E-stained specimens work accurately enough to classify the subtypes of GIM and iron diamine is considered as a potentially harmful and toxic agent (85). It is a fact, that comparison studies addressing the sensibility and sensitivity of methods and techniques for the diagnosis of GIM are currently lacking (85). Still, the underutilization of GIM subtyping might result from a lack of awareness among pathologists and clinicians and the misperception that subtyping might be time consuming, costly or afford additional expertise that is withheld to specialists (57).

Albeit proven evidence for the usefulness of GIM subtyping in routine gastric biopsies exists, barriers to its implementation on a global scale still need to be overcome (57). In this context, a recently published overview on the topic by authors from the USA (57) states the following target points that may help to overcome these barriers:

(1) closer collaboration among pathologists and clinicians (57).
(2) increased awareness on the prognostic value of GIM subtyping and its relevance for the management of patients with gastric precancerous lesions (57).
(3) collaborative efforts in research to generate further evidence and standardized protocols with defined criteria for endoscopists and pathologists, e.g., percental cut-offs for the diagnosis of mixed, complete, or incomplete GIM (57).

This diploma thesis demonstrates not only that subtyping GIM can be achieved with sufficient interobserver agreement, especially when differentiating predominantly complete from predominantly incomplete GIM cases, but also that the rate of misclassification according to current clinical guidelines (4) is low with 5.7% of case ratings. In 14 out of 333 case ratings, patients with predominantly incomplete or pure incomplete GIM were “underreported” when misclassified as pure GIM and therefore might have been lost to follow-up. On the other hand, in 5 out of 333 ratings, patients with pure complete GIM were “overreported” when misclassified as predominantly incomplete or pure incomplete GIM and therefore might have induced unnecessary follow-up investigations. In summary, “underreporting” occurred more often than “overreporting”.

This diploma thesis aimed at identifying factors associated with a higher risk of misclassification. The number of biopsy pieces, the portion of mucosal surface involved by GIM and the pattern of GIM (pure GIM *versus* mixed GIM) might have impact on the pathologists’ ratings and thereby on the level of agreement. The pattern of GIM (pure GIM versus mixed-type GIM) differed significantly between cases with high and low agreement (p=0.010), while the number of biopsy pieces per sample and the portion of mucosal surface involved by GIM did not show a statistically significant difference. One observer reported in the comment section, that he found it particularly challenging to distinguish between predominantly complete or predominantly incomplete GIM within mixed cases. While not only the prognostic value of mixed GIM cases is currently unclear, also interobserver agreement within these cases might be reduced. In our study, Fleiss’ kappa values for mixed GIM categories were lower in comparison to values for pure GIM categories (Table 10). Studies have noted that the likelihood of detecting GIM increases with the number of gastric biopsies obtained which has given rise to concerns that inconsistent sampling thereby might impair the assessment (87,104,105). In our study, no statistically significant difference in the level agreement concerning the number of biopsy pieces obtained was noted.

4.3.4 Management and risk stratification according to the extent of GIM

The extent of GIM in patients with chronic gastritis has prognostic relevance and identifies patients at risk of GC who could benefit from endoscopic surveillance (5,7,77,105,106). For individuals with extensive GIM, an approximately 2-fold

increased risk of neoplastic progression compared to individuals with limited GIM (relative risk 2.07, 95% CI: 0.97-4.42) was reported in a meta-analysis based on two studies (77). Risk stratification based upon the extent of GIM is regarded as cost-effective in low-to intermediate risk regions (7). Current guidelines by the ESGE (4), the BSG (89), and the AGA (90) classify GIM as “extensive”, when it affects both antrum and corpus, requiring, at minimum, biopsies from both regions of the stomach. For patients with extensive GIM endoscopic surveillance is recommended, while patients with GIM limited to the antrum do not need follow-up (4,89,90).

Furthermore, the OLGA/OLGIM staging systems can be used to determine the extent of atrophy and GIM and their associated GC risk. A multicentre cohort study (7) included 290 patients from a population at low- to intermediate-risk of GC for screening endoscopy and surveillance endoscopy at year three and year five. Individuals with OLGIM stage III and IV (12.3% of participants with GIM) were at greatest risk of early gastric neoplasia (adjusted HR 20.7, 95%: CI 5.04-85.6; $p<0.01$) (7). More than 50% of patients staged OLGIM III and IV at baseline endoscopy developed early gastric neoplasia within two years, prompting a two-year surveillance interval (7). Median time intervals between baseline endoscopy and the appearance of early gastric neoplasia were 22.7 months in patients with OLGIM III and IV and 50.7 months in patients with OLGIM II (19.1% of participants with GIM) (7). Therefore, patients with OLGIM II were denominated as patients at intermediate-risk of GC (adjusted HR 7.34, 95% CI: 1.6 0-3.7; $p=0.02$), potentially benefiting from surveillance every five years (7). The vast majority of study participants were staged at OLGIM 0 or I (68.6% of participants with GIM) and do not warrant surveillance endoscopy, because they did not show a significantly increased risk for developing early gastric cancer compared to patients without GIM (7). Still, 12.3% of patients with OLGIM I progressed to OLGIM II-IV (7). At current date, the ESGE recommends endoscopic surveillance in patients with OLGA/OLGIM III and IV (4).

It may be of note, that the number of biopsies may affect the histological diagnosis, that is, the yield of GIM may increase when multiple biopsies are obtained (106). A study by de Vries et al. (106) demonstrated, that the likelihood to be classified as a “patient at risk”, increases with the number of biopsies obtained. In this study, a biopsy protocol with five biopsies sampled according to the Sydney classification (11) was able to detect 71 out of 79 of patients with GIM (90%) while

a biopsy protocol with nine non-targeted biopsies was able to detect 77 out of 79 patients of GIM (97%) (106). Based on this study, de Vries et al. recommend taking at least nine non-targeted biopsies from the cardia, the lesser curvature of corpus, the incisura angularis and the antrum in a GC low-risk population (106).

In clinical practice, heterogenous biopsy protocols and reporting practices are used, *inter alia* due to varying recommendations from heterogenous literature evidence (7). Moreover, biopsy specimens may be involved by GIM in varying quantities, with some showing only tiny foci of GIM, whereas others may show total replacement of the original mucosa. Consequently, some authors suggested alternative scoring methods, such as the relative number of biopsies involved by GIM (88) or the percentage of mucosa involved by GIM (5,105,107). For instance, Cassaro et al. (88) classified GIM as extensive if more than 4 out of 12 biopsy specimens were involved by GIM. Tava et al. (5) investigated the prognostic importance of GIM extent, topography, and type among 471 dyspeptic patients who showed either a GIM extent of ≥10% or lesions that were indefinite-for-dysplasia. The study participants were repeatedly examined over a median period of 52 months (5). Amongst other factors, the GIM extent proved to be an important predictive factor for patients at risk of GC (5). Specifically, a GIM extent of ≥20% at baseline endoscopy was identified as a sensitive cut-off for patients at increased neoplastic risk (5). Individuals with GIM extent of ≥20% compared to <20% had the highest HR (HR=9.25, 95% CI: 1.27-67.19) within this study and a risk rate of 2.24 (95% CI: 1.66-3.02) per 100 persons per year for neoplastic change (5). Similar to our study, the authors used an approach to quantify the extent of GIM that works independently of the number of biopsy pieces obtained, that is, percental estimation at 10% increments (5). Our study shows that eleven gastrointestinal expert pathologists can estimate the extent of GIM in 5% increments with very good agreement (ICC=0.983, 95% CI: 0.975-0.990). It is of note, however, that the standard deviation increased with the amount of GIM, suggesting lower diagnostic accuracy and lower agreement in cases with extensive GIM. High amounts of GIM, i.e., involvement of several biopsy pieces within one sample and/or multiple foci within a single biopsy piece, may impede the estimation and may thereby have a negative impact on the quality of assessment.

4.4 Strengths and limitations

This study has several strengths and limitations. Strengths include the systematic approach involving an international group of gastrointestinal expert pathologists who analysed a large set of biopsies representing all potential patterns of GIM including mixed cases and different quantities of GIM. The pathologists had varying routine experience in GIM subtyping and originated from different parts of the world, with different educational backgrounds and different diagnostic approaches. Still, the lack of general pathologists in this study may be regarded as a limitation.

The restriction to H&E-stained slides could likewise be regarded as a limitation. However, H&E histology has proven prognostic impact in prior publications (82,83,85). Furthermore, enzyme-histochemical and/or immunohistochemical staining methods are not generally applied and have no proven additional prognostic impact. Subtyping GIM on H&E-stained slides is regarded as a minimal time intensive, cheap, and widely available (57,85).

In addition, some may regard the lack of an independent “gold standard”, e.g., provided by morphometric image analysis, as a limitation when estimating the percental extent of GIM. We regarded this, however, as outside the scope of our project, in particular since the routine assessment of GIM is done by usual light microscopy and not by morphometry or comparable tools.

Another limitation might be the use of virtual microscopy, which bears specific technical challenges: pathologists may find it harder to move around all biopsy specimens with the same ease as they do on a microscope. In addition, the evaluation of scanned slides does not allow the assessment of more than one level, thereby potentially hampering the identification of brush borders and Paneth cells. However, the findings in our study are still relevant in view of the expected increase in the use of virtual diagnostics in the future.

It may be of note, that this diploma thesis does not aim at evaluating the GC risk of individuals with GIM, but rather the pathologists’ ability to subtype GIM. GIM has similar features among low-risk and high-risk populations and the GC risk within the Austrian population would not have any impact on the results nor the validity of this study. Therefore, all cases were selected at a single centre, namely at the Medical University of Graz, Research and Diagnostic Institute of Pathology. In this context, one limitation might be that staining protocols and techniques might differ

slightly in other parts of the world, potentially affecting the histological assessment. Nonetheless, other interobserver studies discussed within this diploma thesis, likewise included biopsy samples embedded and stained at a single diagnostic institute. The asymmetrical distribution of cases in different categories of GIM reflects the proportion of subtypes in Austria and may even be an advantage within this study, as the observers participating in a study like this usually expect an even distribution.

4.5 Evidence gaps and considerations for future research

Subtyping GIM is not yet fully incorporated in all clinical guidelines (89) nor included in the routine assessment of gastric biopsies by many pathologists, particularly outside of Europe (7,57). In order to encourage the implementation of subtyping into clinical and pathological routine, further scientific efforts have to be undertaken and current knowledge gaps need to be closed.

Firstly, to date, no RCTs have assessed the effect of GIM surveillance on the detection of dysplasia, early gastric cancer, neuroendocrine tumours (NETs) and survival (4,7,8,57,90,101). Current recommendations for surveillance intervals are widely based on expert opinion (101). Robust evidence from RCTs is needed in order to optimise timing of surveillance in low- to high-risk GC populations (4,90) and in patients with present risk factors such as incomplete (57,101) or extensive GIM (4,5,90,105). Although observational studies in regions with a varying risk of GC have been conducted, these studies used different definitions and showed heterogenous effects which might limit their worldwide application (8,57). Future studies need to precisely define their outcomes by applying standardized protocols for sampling and histological assessment (7,77) and specifying risk factors for progression (108).

Secondly, the cost-effectiveness of management strategies that are built upon GIM subtyping has not been investigated yet. The ESGE regards endoscopic surveillance at a three-year interval in patients with gastric precancerous lesion in countries with intermediate risk of GC as cost-effective based on three studies (109-111) but welcomes further economic research for the optimal surveillance interval (4). Endoscopic screening is currently only recommended in high-risk countries such as Japan and Korea (4,112). A Markov cost-utility analysis from Portugal found that combining screening upper endoscopy at the time of screening colonoscopy is cost-

effective within this intermediate- to high-risk population (4,113). The ESGE regards this strategy as cost-effective within European countries with a GC risk of ≥10 persons per 100.000 inhabitants (4).

Thirdly, the definitive clinical significance of mixed GIM cases, specifically cases with a minor component of GIM (101), still needs to be elucidated. Clearly defined percental cut-offs for the classification of the incomplete subtype have not become established yet (57). The ESGE (4) refers to studies conducted by Gonzalez et al. (82,83,85) who defined a higher GC risk for “predominant incomplete” GIM. Tava et al. (5) proposed a cut-off of either 10% or 20% incomplete GIM identifying patients at risk for GC.

Fourthly, it is still unclear if also general pathologists are able to subtype GIM with sufficient agreement. Studies showed that interobserver agreement in the diagnosis of atrophy and GIM is slightly lower among general pathologists compared to gastrointestinal expert pathologists (69,71,72). Future interobserver variability studies need to be conducted at a larger scale and involve also general pathologists. Previous studies (71,100) and this diploma thesis demonstrated that the knowledge for subtyping GIM can acquired in minimal due time and that the interobserver agreement improves with increasing experience of pathologists. Accordingly, Shah et al. demanded the implementation of online training modules and educational sessions at conferences of the respective clinical and pathological societies (57). The BSG suggested that knowledge on GIM subtyping should be incorporated in the national curricula for gastroenterologists, surgeons, and pathologists (89).

4.6 Conclusion

In conclusion, subtyping and estimating the percental extent of GIM on H&E-stained slides can be achieved with high interobserver agreement among international gastrointestinal expert pathologists. The implementation of GIM subtyping as a risk-stratifying tool in current practice guidelines by the ESGE and the AGA carried a low rate of misclassification within this study. This diploma thesis provides the basis for future research in the field, e.g., by expanding the evaluation to general pathologists in a nation-wide setting, and for the potential implementation of percental GIM assessment in the respective guidelines on gastric precancerous lesions.

References

(1) Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021 May;71(3):209-249.

(2) Correa P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res 1992 Dec 15;52(24):6735-6740.

(3) Pimentel-Nunes P, Dinis-Ribeiro M, Ponchon T, Repici A, Vieth M, De Ceglie A, et al. Endoscopic submucosal dissection: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy 2015 Sep;47(9):829-854.

(4) Pimentel-Nunes P, Libânio D, Marcos-Pinto R, Areia M, Leja M, Esposito G, et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019. Endoscopy 2019 Apr;51(4):365-388.

(5) Tava F, Luinetti O, Ghigna MR, Alvisi C, Perego M, Trespi E, et al. Type or extension of intestinal metaplasia and immature/atypical "indefinite-for-dysplasia" lesions as predictors of gastric neoplasia. Hum Pathol 2006 Nov;37(11):1489-1497.

(6) Yeh JM, Hur C, Kuntz KM, Ezzati M, Goldie SJ. Cost-effectiveness of treatment and endoscopic surveillance of precancerous lesions to prevent gastric cancer. Cancer 2010 Jun 15;116(12):2941-2953.

(7) Lee JWJ, Zhu F, Srivastava S, Tsao SK, Khor C, Ho KY, et al. Severity of gastric intestinal metaplasia predicts the risk of gastric cancer: a prospective multicentre cohort study (GCEP). Gut 2021 May 11;gutjnl-2021-324057.

(8) Matysiak-Budnik T, Camargo MC, Piazuelo MB, Leja M. Recent Guidelines on the Management of Patients with Gastric Atrophy: Common Points and Controversies. Dig Dis Sci 2020 Jul;65(7):1899-1903.

(9) World Health Organization. International classification of diseases for mortality and morbidity statistics. 2018; Available at: https://icd.who.int/browse11/l-m/en. Accessed Aug 6, 2021.

(10) Sugano K, Tack J, Kuipers EJ, Graham DY, El-Omar EM, Miura S, et al. Kyoto global consensus report on Helicobacter pylori gastritis. Gut 2015 Sep;64(9):1353-1367.

(11) Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol 1996 Oct;20(10):1161-1181.

(12) Rugge M, Genta RM. Staging and grading of chronic gastritis. Hum Pathol 2005 Mar;36(3):228-233.

(13) Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, et al. Helicobacter pylori infection and the development of gastric cancer. N Engl J Med 2001 Sep 13;345(11):784-789.

(14) Rugge M, Correa P, Dixon MF, Fiocca R, Hattori T, Lechago J, et al. Gastric mucosal atrophy: interobserver consistency using new criteria for classification and grading. Aliment Pharmacol Ther 2002 Jul;16(7):1249-1259.

(15) Rugge M, Meggio A, Pennelli G, Piscioli F, Giacomelli L, De Pretis G, et al. Gastritis staging in clinical practice: the OLGA staging system. Gut 2007 May;56(5):631-636.

(16) Capelle LG, de Vries AC, Haringsma J, Ter Borg F, de Vries RA, Bruno MJ, et al. The staging of gastritis with the OLGA system by using intestinal metaplasia as an accurate alternative for atrophic gastritis. Gastrointest Endosc 2010 Jun;71(7):1150-1158.

(17) Bettington M, Brown I. Autoimmune gastritis: novel clues to histological diagnosis. Pathology 2013 Feb;45(2):145-149.

(18) Toh BH, Chan J, Kyaw T, Alderuccio F. Cutting edge issues in autoimmune gastritis. Clin Rev Allergy Immunol 2012 Jun;42(3):269-278.

(19) Chmiela M, Gonciarz W. Molecular mimicry in Helicobacter pylori infections. World J Gastroenterol 2017 Jun 14;23(22):3964-3977.

(20) Neumann WL, Coss E, Rugge M, Genta RM. Autoimmune atrophic gastritis--pathogenesis, pathology and management. Nat Rev Gastroenterol Hepatol 2013 Sep;10(9):529-541.

(21) Owen DA. Gastritis and carditis. Mod Pathol 2003 Apr;16(4):325-341.

(22) Sepulveda AR, Patil M. Practical approach to the pathologic diagnosis of gastritis. Arch Pathol Lab Med 2008 Oct;132(10):1586-1593.

(23) Torbenson M, Abraham SC, Boitnott J, Yardley JH, Wu TT. Autoimmune gastritis: distinct histological and immunohistochemical findings before complete loss of oxyntic glands. Mod Pathol 2002 Feb;15(2):102-109.

(24) Wolf EM, Plieschnegger W, Geppert M, Wigginghaus B, Höss GM, Eherer A, et al. Changing prevalence patterns in endoscopic and histological diagnosis of gastritis? Data from a cross-sectional Central European multicentre study. Dig Liver Dis 2014 May;46(5):412-418.

(25) Jhala NC, Montemor M, Jhala D, Lu L, Talley L, Haber MM, et al. Pancreatic acinar cell metaplasia in autoimmune gastritis. Arch Pathol Lab Med 2003 Jul;127(7):854-857.

(26) Malfertheiner P, Chan FK, McColl KE. Peptic ulcer disease. Lancet 2009 Oct 24;374(9699):1449-1461.

(27) Malfertheiner P, Megraud F, O'Morain CA, Atherton J, Axon AT, Bazzoli F, et al. Management of Helicobacter pylori infection--the Maastricht IV/ Florence Consensus Report. Gut 2012 May;61(5):646-664.

(28) Fock KM, Graham DY, Malfertheiner P. Helicobacter pylori research: historical insights and future directions. Nat Rev Gastroenterol Hepatol 2013 Aug;10(8):495-500.

(29) Zamani M, Ebrahimtabar F, Zamani V, Miller WH, Alizadeh-Navaei R, Shokri-Shirvani J, et al. Systematic review with meta-analysis: the worldwide prevalence of Helicobacter pylori infection. Aliment Pharmacol Ther 2018 Apr;47(7):868-876.

(30) Tang CT, Zeng L, Yang J, Zeng C, Chen Y. Analysis of the Incidence and Survival of Gastric Cancer Based on the Lauren Classification: A Large Population-Based Study Using SEER. Front Oncol 2020 Aug 3;10:1212.

(31) Yamamichi N, Yamaji Y, Shimamoto T, Takahashi Y, Majima K, Wada R, et al. Inverse time trends of peptic ulcer and reflux esophagitis show significant association with reduced prevalence of Helicobacter pylori infection. Ann Med 2020 Dec;52(8):506-514.

(32) Wolf EM, Plieschnegger W, Schmack B, Bordel H, Höfler B, Eherer A, et al. Evolving patterns in the diagnosis of reactive gastropathy: data from a prospective Central European multicenter study with proposal of a new histologic scoring system. Pathol Res Pract 2014 Dec;210(12):847-854.

(33) Maguilnik I, Neumann WL, Sonnenberg A, Genta RM. Reactive gastropathy is associated with inflammatory conditions throughout the gastrointestinal tract. Aliment Pharmacol Ther 2012 Oct;36(8):736-743.

(34) Dixon MF, O'Connor HJ, Axon AT, King RF, Johnston D. Reflux gastritis: distinct histopathological entity? J Clin Pathol 1986 May;39(5):524-530.

(35) Carneiro F, Fukayama M, Grabsch HI, Yasui W. Gastric adenocarcinoma. In: WHO Classification of Tumours Editorial Board, editor. WHO Classification of Tumours Editorial Board. Digestive System Tumours. Lyon (France): International Agency for Research on Cancer; 2019. p. 85-95.

(36) LAUREN P. The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. an Attempt at a Histo-Clinical Classification. Acta Pathol Microbiol Scand 1965;64:31-49.

(37) Ma J, Shen H, Kapesa L, Zeng S. Lauren classification and individualized chemotherapy in gastric cancer. Oncol Lett 2016 May;11(5):2959-2964.

(38) Arnold M, Park JY, Camargo MC, Lunet N, Forman D, Soerjomataram I. Is gastric cancer becoming a rare disease? A global assessment of predicted incidence trends to 2035. Gut 2020 May;69(5):823-829.

(39) Petryszyn P, Chapelle N, Matysiak-Budnik T. Gastric Cancer: Where Are We Heading? Dig Dis 2020;38(4):280-285.

(40) Sung H, Siegel RL, Rosenberg PS, Jemal A. Emerging cancer trends among young adults in the USA: analysis of a population-based cancer registry. Lancet Public Health 2019 Mar;4(3):e137-e147.

(41) Ibrahim A, Morais S, Ferro A, Lunet N, Peleteiro B. Sex-differences in the prevalence of Helicobacter pylori infection in pediatric and adult populations: Systematic review and meta-analysis of 244 studies. Dig Liver Dis 2017 Jul;49(7):742-749.

(42) Díaz Del Arco C, Ortega Medina L, Estrada Muñoz L, García Gómez de Las Heras, S., Fernández Aceñero MJ. Is there still a place for conventional histopathology in the age of molecular medicine? Laurén classification, inflammatory infiltration and other current topics in gastric cancer diagnosis and prognosis. Histol Histopathol 2021 Jun;36(6):505-514.

(43) Correa P, Piazuelo MB. The gastric precancerous cascade. J Dig Dis 2012 Jan;13(1):2-9.

(44) Correa P. A human model of gastric carcinogenesis. Cancer Res 1988 Jul 1;48(13):3554-3560.

(45) Tan P, Yeoh KG. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. Gastroenterology 2015 Oct;149(5):1153-1162.e3.

(46) Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014 Sep 11;513(7517):202-209.

(47) Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015 May;21(5):449-456.

(48) Allum W, Lordick F, Alsina M, Andritsch E, Ba-Ssalamah A, Beishon M, et al. ECCO essential requirements for quality cancer care: Oesophageal and gastric cancer. Crit Rev Oncol Hematol 2018 Feb;122:179-193.

(49) Sumiyama K. Past and current trends in endoscopic diagnosis for early stage gastric cancer in Japan. Gastric Cancer 2017 Mar;20(Suppl 1):20-27.

(50) Ono H, Yao K, Fujishiro M, Oda I, Nimura S, Yahagi N, et al. Guidelines for endoscopic submucosal dissection and endoscopic mucosal resection for early gastric cancer. Dig Endosc 2016 Jan;28(1):3-15.

(51) Noffsinger A, Fenoglio-Preiser C, Maru D, Gilinsky N. Gastrointestinal diseases - Atlas of nontumor pathology. 1st ed. Washington, DC: The American Registry of Pathology; 2007.

(52) Kushima R, Lauwers GY, Rugge M. Gastric dysplasia. In: WHO Classification of Tumours Editorial Board, editor. WHO Classification of Tumours Editorial Board. Digestive System Tumours Lyon (France): International Agency for Research on Cancer; 2019. p. 71-75.

(53) Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma: 3rd English edition. Gastric Cancer 2011 Jun;14(2):101-112.

(54) Rugge M. Gastritis and metaplasia: precursors of gastric neoplasms. In: WHO Classification of Tumours Editorial Board, editor. WHO Classification of Tumours Editorial Board. Digestive system tumors. 5th ed. ed. Lyon (France): International Agency for Research on Cancer; 2019. p. 65-66.

(55) Faller G, Kirchner T. Immunological and morphogenic basis of gastric mucosa atrophy and metaplasia. Virchows Arch 2005 Jan;446(1):1-9.

(56) Correa P, Piazuelo MB, Wilson KT. Pathology of gastric intestinal metaplasia: clinical implications. Am J Gastroenterol 2010 Mar;105(3):493-498.

(57) Shah SC, Gawron AJ, Mustafa RA, Piazuelo MB. Histologic Subtyping of Gastric Intestinal Metaplasia: Overview and Considerations for Clinical Practice. Gastroenterology 2020 Feb;158(3):745-750.

(58) Jass JR, Filipe MI. The mucin profiles of normal gastric mucosa, intestinal metaplasia and its variants and gastric carcinoma. Histochem J 1981 Nov;13(6):931-939.

(59) Rugge M, Cassaro M, Di Mario F, Leo G, Leandro G, Russo VM, et al. The long-term outcome of gastric non-invasive neoplasia. Gut 2003 Aug;52(8):1111-1116.

(60) Zheng H, Takahashi H, Murai Y, Cui Z, Nomoto K, Miwa S, et al. Pathobiological characteristics of intestinal and diffuse-type gastric carcinoma in Japan: an immunostaining study on the tissue microarray. J Clin Pathol 2007 Mar;60(3):273-277.

(61) Qiu MZ, Cai MY, Zhang DS, Wang ZQ, Wang DS, Li YH, et al. Clinicopathological characteristics and prognostic analysis of Lauren classification in gastric adenocarcinoma in China. J Transl Med 2013 Mar 6;11:58-58.

(62) Cisło M, Filip AA, Arnold Offerhaus GJ, Ciseł B, Rawicz-Pruszyński K, Skierucha M, et al. Distinct molecular subtypes of gastric cancer: from Laurén to molecular pathology. Oncotarget 2018 Apr 10;9(27):19427-19442.

(63) Chen YC, Fang WL, Wang RF, Liu CA, Yang MH, Lo SS, et al. Clinicopathological Variation of Lauren Classification in Gastric Cancer. Pathol Oncol Res 2016 Jan;22(1):197-202.

(64) Union for International Cancer Control. TNM classification of malignant tumours. 8th ed. ed. Oxford (UK): Wiley Blackwell; 2017.

(65) Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D, et al. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2016 Sep;27(suppl 5):v38-v49.

(66) Waddell T, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D, et al. Gastric cancer: ESMO-ESSO-ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2013 Oct;24 Suppl 6:vi57-63.

(67) Sano T, Coit DG, Kim HH, Roviello F, Kassab P, Wittekind C, et al. Proposal of a new stage grouping of gastric cancer for TNM classification: International Gastric Cancer Association staging project. Gastric Cancer 2017 Mar;20(2):217-225.

(68) Hamashima C, Systematic Review Group and Guideline Development Group for Gastric Cancer Screening Guidelines. Update version of the Japanese Guidelines for Gastric Cancer Screening. Jpn J Clin Oncol 2018 Jul 1;48(7):673-683.

(69) Isajevs S, Liepniece-Karele I, Janciauskas D, Moisejevs G, Putnins V, Funka K, et al. Gastritis staging: interobserver agreement by applying OLGA and OLGIM systems. Virchows Arch 2014 Apr;464(4):403-407.

(70) Yue H, Shan L, Bin L. The significance of OLGA and OLGIM staging systems in the risk assessment of gastric cancer: a systematic review and meta-analysis. Gastric Cancer 2018 Jul;21(4):579-587.

(71) Kim SS, Kook MC, Shin OR, Kim HS, Bae HI, Seo AN, et al. Factors to improve the interobserver agreement for gastric atrophy and intestinal metaplasia: consensus of definition and criteria. Histopathology 2018 Apr;72(5):838-845.

(72) Leja M, Funka K, Janciauskas D, Putnins V, Ruskule A, Kikuste I, et al. Interobserver variation in assessment of gastric premalignant lesions: higher agreement for intestinal metaplasia than for atrophy. Eur J Gastroenterol Hepatol 2013 Jun;25(6):694-699.

(73) de Vries AC, van Grieken NC, Looman CW, Casparie MK, de Vries E, Meijer GA, et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. Gastroenterology 2008 Apr;134(4):945-952.

(74) Aumpan N, Vilaichone RK, Nunanan P, Chonprasertsuk S, Siramolpiwat S, Bhanthumkomol P, et al. Predictors for development of complete and incomplete intestinal metaplasia (IM) associated with H. pylori infection: A large-scale study from low prevalence area of gastric cancer (IM-HP trial). PLoS One 2020 Oct 1;15(10):e0239434.

(75) Song H, Ekheden IG, Zheng Z, Ericsson J, Nyrén O, Ye W. Incidence of gastric cancer among patients with gastric precancerous lesions: observational cohort study in a low risk Western population. BMJ 2015 Jul 27;351:h3867.

(76) Li D, Bautista MC, Jiang SF, Daryani P, Brackett M, Armstrong MA, et al. Risks and Predictors of Gastric Adenocarcinoma in Patients with Gastric Intestinal Metaplasia and Dysplasia: A Population-Based Study. Am J Gastroenterol 2016 Aug;111(8):1104-1113.

(77) Gawron AJ, Shah SC, Altayar O, Davitkov P, Morgan D, Turner K, et al. AGA Technical Review on Gastric Intestinal Metaplasia-Natural History and Clinical Outcomes. Gastroenterology 2020 Feb;158(3):705-731.e5.

(78) Wei N, Zhou M, Lei S, Zhong Z, Shi R. A meta-analysis and systematic review on subtypes of gastric intestinal metaplasia and neoplasia risk. Cancer Cell Int 2021 Mar 17;21(1):173.

(79) Chapelle N, Péron M, Quénéhervé L, Bourget A, Leroy M, Touchefeu Y, et al. Long-Term Follow-up of Gastric Precancerous Lesions in a Low GC Incidence Area. Clin Transl Gastroenterol 2020 Dec;11(12):e00237.

(80) Dinis-Ribeiro M, Lopes C, da Costa-Pereira A, Guilherme M, Barbosa J, Lomba-Viana H, et al. A follow up model for patients with atrophic chronic gastritis and intestinal metaplasia. J Clin Pathol 2004 Feb;57(2):177-182.

(81) Mera RM, Bravo LE, Camargo MC, Bravo JC, Delgado AG, Romero-Gallo J, et al. Dynamics of Helicobacter pylori infection as a determinant of progression of gastric precancerous lesions: 16-year follow-up of an eradication trial. Gut 2018 Jul;67(7):1239-1246.

(82) González CA, Sanz-Anquela JM, Companioni O, Bonet C, Berdasco M, López C, et al. Incomplete type of intestinal metaplasia has the highest risk to progress to gastric cancer: results of the Spanish follow-up multicenter study. J Gastroenterol Hepatol 2016 May;31(5):953-958.

(83) González CA, Pardo ML, Liso JM, Alonso P, Bonet C, Garcia RM, et al. Gastric cancer occurrence in preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. Int J Cancer 2010 Dec 1;127(11):2654-2660.

(84) Isajevs S, Savcenko S, Liepniece-Karele I, Piazuelo MB, Kikuste I, Tolmanis I, et al. High-risk individuals for gastric cancer would be missed for surveillance without subtyping of intestinal metaplasia. Virchows Arch 2021 May 14;479(4):679-686.

(85) González CA, Sanz-Anquela JM, Gisbert JP, Correa P. Utility of subtyping intestinal metaplasia as marker of gastric cancer risk. A review of the evidence. Int J Cancer 2013 Sep 1;133(5):1023-1032.

(86) Shao L, Li P, Ye J, Chen J, Han Y, Cai J, et al. Risk of gastric cancer among patients with gastric intestinal metaplasia. Int J Cancer 2018 Oct 1;143(7):1671-1677.

(87) Pittayanon R, Rerknimitr R, Klaikaew N, Sanpavat A, Chaithongrat S, Mahachai V, et al. The risk of gastric cancer in patients with gastric intestinal metaplasia in 5-year follow-up. Aliment Pharmacol Ther 2017 Jul;46(1):40-45.

(88) Cassaro M, Rugge M, Gutierrez O, Leandro G, Graham DY, Genta RM. Topographic patterns of intestinal metaplasia and gastric cancer. Am J Gastroenterol 2000 Jun;95(6):1431-1438.

(89) Banks M, Graham D, Jansen M, Gotoda T, Coda S, di Pietro M, et al. British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma. Gut 2019 Sep;68(9):1545-1575.

(90) Gupta S, Li D, El Serag HB, Davitkov P, Altayar O, Sultan S, et al. AGA Clinical Practice Guidelines on Management of Gastric Intestinal Metaplasia. Gastroenterology 2020 Feb;158(3):693-702.

(91) Dinis-Ribeiro M, Kuipers EJ. How to Manage a Patient with Gastric Intestinal Metaplasia: An International Perspective. Gastroenterology 2020 May;158(6):1534-1537.

(92) Gupta S, Tao L, Murphy JD, Camargo MC, Oren E, Valasek MA, et al. Race/Ethnicity-, Socioeconomic Status-, and Anatomic Subsite-Specific Risks for Gastric Cancer. Gastroenterology 2019 Jan;156(1):59-62.e4.

(93) Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, et al. Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. Gastroenterology 2017 Aug;153(2):420-429.

(94) Fleiss J. Statistical methods for rates and proportions. New York: Wiley; 1981.

(95) Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Measures of agreement. Perspect Clin Res 2017;8(4):187-191.

(96) Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174.

(97) Altman D. Practical statistics for medical research. London: Chapman and Hall; 1977.

(98) Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428.

(99) Hayes AF, Cai L. Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation. Behav Res Methods 2007 Nov;39(4):709-722.

(100) Guarner J, Herrera-Goepfert R, Mohar A, Sanchez L, Halperin D, Ley C, et al. Interobserver variability in application of the revised Sydney classification for gastritis. Hum Pathol 1999 Dec;30(12):1431-1434.

(101) Bhandari P, Longcroft-Wheaton G, Libanio D, Pimentel-Nunes P, Albeniz E, Pioche M, et al. Revising the European Society of Gastrointestinal Endoscopy (ESGE) research priorities: a research progress update. Endoscopy 2021 May;53(5):535-554.

(102) Huang KK, Ramnarayanan K, Zhu F, Srivastava S, Xu C, Tan ALK, et al. Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer. Cancer Cell 2018 Jan 8;33(1):137-150.e5.

(103) Yasui W, Oue N, Aung PP, Matsumura S, Shutoh M, Nakayama H. Molecular-pathological prognostic factors of gastric cancer: a review. Gastric Cancer 2005;8(2):86-94.

(104) Chen K, Yang D, Li X, Sun B, Song F, Cao W, et al. Mutational landscape of gastric adenocarcinoma in Chinese: implications for prognosis and therapy. Proc Natl Acad Sci U S A 2015 Jan 27;112(4):1107-1112.

(105) Zullo A, Hassan C, Romiti A, Giusto M, Guerriero C, Lorenzetti R, et al. Follow-up of intestinal metaplasia in the stomach: When, how and why. World J Gastrointest Oncol 2012 Mar 15;4(3):30-36.

(106) de Vries AC, Haringsma J, de Vries RA, Ter Borg F, van Grieken NC, Meijer GA, et al. Biopsy strategies for endoscopic surveillance of pre-malignant gastric lesions. Helicobacter 2010 Aug;15(4):259-264.

(107) Pennelli G, Grillo F, Galuppini F, Ingravallo G, Pilozzi E, Rugge M, et al. Gastritis: update on etiological features and histological practical approach. Pathologica 2020 Sep;112(3):153-165.

(108) Altayar O, Davitkov P, Shah SC, Gawron AJ, Morgan DR, Turner K, et al. AGA Technical Review on Gastric Intestinal Metaplasia-Epidemiology and Risk Factors. Gastroenterology 2020 Feb;158(3):732-744.e16.

(109) Areia M, Dinis-Ribeiro M, Rocha Gonçalves F. Cost-utility analysis of endoscopic surveillance of patients with gastric premalignant conditions. Helicobacter 2014 Dec;19(6):425-436.

(110) Zhou HJ, Dan YY, Naidoo N, Li SC, Yeoh KG. A cost-effectiveness analysis evaluating endoscopic surveillance for gastric cancer for populations with low to intermediate risk. PLoS One 2013 Dec 27;8(12):e83959.

(111) Wu JT, Zhou J, Naidoo N, Yang WY, Lin XC, Wang P, et al. Determining the cost-effectiveness of endoscopic surveillance for gastric cancer in patients with precancerous lesions. Asia Pac J Clin Oncol 2016 Dec;12(4):359-368.

(112) Areia M, Carvalho R, Cadime AT, Rocha Gonçalves F, Dinis-Ribeiro M. Screening for gastric cancer and surveillance of premalignant lesions: a systematic review of cost-effectiveness studies. Helicobacter 2013 Oct;18(5):325-337.

(113) Areia M, Spaander MC, Kuipers EJ, Dinis-Ribeiro M. Endoscopic screening for gastric cancer: A cost-utility analysis for countries with an intermediate gastric cancer risk. United European Gastroenterol J 2018 Mar;6(2):192-202.

Appendix

Figure 54: Evaluation sheet.

19.04.2021 | Version 01

Investigator: .....

In my daily routine work, I subtype IM:

Always .....

Sometimes .....

Never .....

For those colleagues, who are familiar with subtyping from daily routine service:

I usually do the subtyping on H&E stained slides .....

I usually perform histochemical stains .....

I usually perform immunohistochemistry .....

Case number + slide name	**Category 1:** Complete IM only	**Category 2:** Complete IM > incomplete IM	**Category 3:** Incomplete IM > complete IM	**Category 4:** Incomplete IM only	%	H.P. gastritis	Post(ex)-H.P. gastritis	Reactive gastropathy
Case #1								
Case #2								
Case #3								
Case #4								

Case #5								
Case #6								
Case #7								
Case #8								
Case #9								
Case #10								
Case #11								
Case #12								
Case #13								
Case #14								
Case #15								
Case #16								
Case #17								
Case #18								
Case #19								

Case #20								
Case #21								
Case #22								
Case #23								
Case #24								
Case #25								
Case #26								
Case #27								
Case #28								
Case #29								
Case #30								
Case #31								
Case #32								
Case #33								
Case #34								
Case #35								

Case #36								
Case #37								
Case #38								
Case #39								
Case #40								
Case #41								
Case #42								
Case #43								
Case #44								
Case #45								
Case #46								

Comments: