Supplementary 1. An example of prompts and responses in ChatGPT

Prompts 1 to 3 were not used in this study

Prompt 1

I am running an experiment on a case report to see how your diagnoses compare with those of human experts. I am going to give you part of a medical case. These have all been published in the medical journals. You are not trying to treat any patients. In this case, you are "Dr. GPT-4," an AI language model who is discussing the case along with human experts. A case report has several unspoken rules. The first is that there is most often a single definitive diagnosis (though rarely there may be more than one), and it is a diagnosis that is known today to exist in humans. The diagnosis is almost always confirmed by some sort of clinical pathology test or anatomic pathology test, though in rare cases when such a test does not exist for a diagnosis the diagnosis can instead be made using validated clinical criteria or very rarely just confirmed by expert opinion. You will be told at the end of the case description whether a diagnostic test/tests are being ordered, which you can assume will make the diagnosis/diagnoses. After you read the case, I want you to give two pieces of information. The first piece of information is your most likely diagnosis/diagnoses. You need to be as specific as possible -- the goal is to get the correct answer, not a broad category of answers. You do not need to explain your reasoning, just give the diagnosis/diagnoses. The second piece of information is to give a robust differential diagnosis, ranked by their probability so that the most likely diagnosis is at the top, and the least likely is at the bottom. There is no limit to the number of diagnoses on your differential. You can give as many diagnoses as you think are reasonable. You do not need to explain your reasoning, just list the diagnoses. Again, the goal is to be as specific as possible with each of the diagnoses. Do you have any questions, Dr. GPT-4?

Response 1

Understood, I'm ready to assist. Please present the case details, and I'll provide the two pieces of information as specified: the most likely diagnosis/diagnoses and a robust differential diagnosis ranked by probability.

Prompt 2

A 47-year-old woman presented to our emergency department one and a half years after she had undergone a laparoscopic Roux-en-Y gastric bypass. The laparoscopic

Roux-en-Y gastric bypass (LRYGB) was created with an antecolic alimentary limb of 150cm and a biliopancreatic limb of 50cm. The original report stated that possible hernia sites were closed during the primary surgical procedure by using sutures. At the time of presentation, the patient's body mass index (BMI) had declined from 36.1kg/m^2 to 26kg/m^2. The patient complained of intermitted abdominal pain with nausea and vomiting. Physical examination revealed no tachycardia or fever. The bowel sounds were normal. Tenderness was found in the upper abdomen without further signs of peritonitis. A laboratory test showed no abnormalities (CRP 2.0mg/L, white cell count 6.3 x 10^9 /L). A CT of the abdomen with oral contrast showed minimally dilated small bowel loops without signs of obstruction. Our differential diagnosis included an obstruction due to adhesions, internal herniation, or a gastrointestinal infection. The patient was admitted to our hospital for clinical observation. During admission, she continued having intermitted abdominal cramps. She developed diarrhea, which seemed to confirm the diagnosis of gastrointestinal infection. During the first 2 days of admission, our patient remained clinically stable without abnormalities on physical examination or blood tests. Three days after admission, she suffered from clinical deterioration. Physical examination of the abdomen showed the absence of bowel sounds and signs of generalized peritonitis. Laboratory tests showed an elevated CRP of 243 mg/L and a reduced white cell count of 1.1 x 10^9 /L. A CT of the abdomen showed free air within the peritoneal cavity and dilated small bowel loops.

Response 2

Most Likely Diagnosis:
Internal herniation leading to bowel ischemia and perforation.

Differential Diagnosis (ranked by probability):
Internal herniation.
Obstruction due to post-surgical adhesions.
Gastrointestinal infection leading to inflammation and subsequent perforation.
Anastomotic leak from the gastric bypass procedure.
Non-occlusive mesenteric ischemia.
Bowel volvulus.
Diverticular perforation (although less likely given the patient's surgical history).
Ischemic bowel disease.
Perforated peptic ulcer disease.

Mesenteric venous thrombosis.
Considering the progression of the symptoms and the CT findings, the most likely diagnosis would be internal herniation leading to bowel ischemia and perforation. This condition can occur after LRYGB, and even if hernia sites were sutured closed during the primary surgery, herniation can still sometimes occur.

Prompt 3
The correct diagnosis in this case is “Internal hernia (herniation of the small intestine distal to the jejunojejunostomy through Petersen’s space after Roux-en-Y gastric bypass)”. Do you think your most likely diagnosis is correct? Please indicate 1 if you think your most likely diagnosis is correct, and 0 if you think your most likely diagnosis is not correct.

Response 3
1
My most likely diagnosis was "Internal herniation leading to bowel ischemia and perforation," which aligns with the provided diagnosis.

Prompt 4
Next, I am going to give you the whole case description for this case. I want you to classify the case into one of the following four categories: typical presentation of common disease; atypical presentation of common disease; typical presentation of uncommon disease; and atypical presentation of uncommon disease. Uncommon disease is defined as a disease that affects not more than 1 person per 2000. Atypical presentation is defined as a shortage of prototypical features that are most frequently encountered in patients with the disease, features encountered in advanced presentations of the disease or simply features of the disease commonly listed in medical textbooks. Please provide your answer as an assigned number: 1 for typical presentation of common disease; 2 for atypical presentation of common disease; 3 for typical presentation of uncommon disease; and 4 for atypical presentation of uncommon disease. Do you have any questions, Dr. GPT-4?

Response 4
I understand the instructions. Please present the complete case description, and I'll classify it into one of the specified categories.

Prompt 5

A 47-year-old woman presented to our emergency department one and a half years after she had undergone a laparoscopic Roux-en-Y gastric bypass. The laparoscopic Roux-en-Y gastric bypass (LRYGB) was created with an antecolic alimentary limb of 150cm and a biliopancreatic limb of 50cm. The original report stated that possible hernia sites were closed during the primary surgical procedure by using sutures. At the time of presentation, the patient's body mass index (BMI) had declined from 36.1kg/m^2 to 26 kg/m^2. The patient complained of intermitted abdominal pain with nausea and vomiting. Physical examination revealed no tachycardia or fever. The bowel sounds were normal. Tenderness was found in the upper abdomen without further signs of peritonitis. A laboratory test showed no abnormalities (CRP 2.0mg/L, white cell count 6.3 x 10^9 /L). A CT of the abdomen with oral contrast showed minimally dilated small bowel loops without signs of obstruction. Our differential diagnosis included an obstruction due to adhesions, internal herniation, or a gastrointestinal infection. The patient was admitted to our hospital for clinical observation. During admission, she continued having intermitted abdominal cramps. She developed diarrhea, which seemed to confirm the diagnosis of gastrointestinal infection. During the first 2 days of admission, our patient remained clinically stable without abnormalities on physical examination or blood tests. Three days after admission, she suffered from clinical deterioration. Physical examination of the abdomen showed the absence of bowel sounds and signs of generalized peritonitis. Laboratory tests showed an elevated CRP of 243mg/L and a reduced white cell count of 1.1 x 10^9 /L. A CT of the abdomen showed free air within the peritoneal cavity and dilated small bowel loops. An emergency laparoscopy showed a herniation of the small intestine distal to the jejunojejunostomy through Petersen's space. By laparotomy, 30cm of strangulated, perforated jejunum was resected. No evidence was found that potential hernia sites were closed during the first operation. We closed all internal hernia sites by suturing. The postoperative course was complicated by a paralytic ileus and pneumonia. Our patient was discharged after 18 days.

Response 5

Considering the details provided:

The condition described is an "Internal hernia through Petersen's space after Roux-en-Y gastric bypass," which is a known complication after Roux-en-Y gastric bypass

surgeries. However, it's relatively rare, occurring in a minority of patients who have undergone this surgery.

Regarding the presentation, several prototypical features of internal herniation after LRYGB are present, such as intermittent abdominal pain, episodes of obstruction, and clinical findings of bowel ischemia and perforation. However, the presentation was somewhat delayed and initially masked by symptoms mimicking a gastrointestinal infection.

Given this information:

This case can be classified as:
3 - Typical presentation of uncommon disease.

Prompt 6
Next, I want you to judge whether the diagnostic errors occurred in this case. Diagnostic errors are defined as the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient. Please indicate 1 if you think diagnostic errors occurred, and 0 if you think diagnostic errors did not occur.

Response 6
Given the case details:

The patient initially presented with intermittent abdominal pain, nausea, and vomiting following a surgical procedure known to have a risk of complications like internal hernias. Despite the differential diagnosis including the possibility of an internal herniation, the initial clinical focus shifted towards a gastrointestinal infection due to the presence of diarrhea. The significant deterioration of the patient three days after admission, highlighted by the absence of bowel sounds, generalized peritonitis, and CT findings, only then prompted an intervention that revealed the true cause: an internal herniation.

Given the definition of diagnostic errors and the information provided:

1 - Diagnostic errors occurred.

Prompt 7

Next, I want you to code the contributing factors of diagnostic errors in this case by reading the whole case description. Coding should be based on Diagnosis Error Evaluation and Research (DEER), Reliable Diagnosis Challenges (RDC), and Generic Diagnostic Pitfalls (GDP) taxonomies, respectively. I am going to give you the details of the taxonomies. You should output the codes in the style presenting only a number and a character such as "1A," "2B," for DEER and RDC, and only a number such as "1," "2" for GDP. There is no limit to the number of codes. You can give as many codes as you think are reasonable. When no codes are applicable, output should be "unclear" for DEER, RDC, and GDP. After that, I want you to select the most important code in this case for DEER, RDC, and GDP taxonomies. You should output the codes in the style presenting only a number and a character such as "1A," "2B," for DEER and RDC, and only a number such as "1," "2" for GDP. When no codes are applicable, output should be "unclear" for DEER, RDC, and GDP. Here is the details of DEER, RDC, and GDP taxonomies:

DEER
1. Access/Presentation
A. Failure/delay in presentation
B. Failure/denied care access

2. History
A. Failure/delay in eliciting critical piece of history data
B. Inaccurate/misinterpretation
C. Failure in weighing
D. Failure/delay to follow-up

3. Physical Exam
A. Failure/delay in eliciting critical physical exam finding
B. Inaccurate/misinterpreted
C. Failure in weighing
D. Failure/delay to follow-up

4. Tests (Lab/Radiology)
Ordering

A. Failure/delay in ordering needed test(s)
B. Failure/delay in performing ordered test(s)
C. Error in test sequencing
D. Ordering of wrong test(s)
E. Test ordered wrong way

Performance
F. Sample mixup/mislabeled (eg, wrong patient/test)
G. Technical errors/poor processing of specimen/test
H. Erroneous lab/radiology reading of test
I. Failed/delayed reporting of result to clinician

Clinician Processing
J. Failed/delayed follow-up of (abnormal) test result
K. Error in clinician interpretation of test

5. Assessment Hypothesis Generation
A. Failure/delay in considering the diagnosis

Suboptimal Weighing/Prioritization
B. Too little consideration/weight given to the diagnosis
C. Too much weight on competing/coexisting diagnosis

Recognizing Urgency/Complications
D. Failure/delay to recognize/weigh urgency
E. Failure/delay to recognize/weigh complication(s)

6. Referral/Consultation
A. Failure/delay in ordering referral
B. Failure/delay obtaining/scheduling ordered referral
C. Error in diagnostic consultation performance
D. Failure/delayed communication/follow-up of consultation

7. Follow-up
A. Failure to refer patient to close/safe setting/monitoring
B. Failure/delay in timely follow-up/rechecking of patient

RDC

1. Challenging disease presentation

A. Atypical presentation

B. Non-specific symptoms and signs

C. Unfamiliar/outside specialty

D. Findings masking/mimicking another diagnosis

E. Red herring misleading findings

F. Rapidly progressive course

G. Slowly evolving blunting onset perception

H. Deceptively benign course

2. Patient factors

A. Language/communication barriers

B. Signal:noise -- patients with multiple other symptoms or diagnoses

C. Failure to share data (to be forthcoming with symptoms or their severity)

D. Failure to follow-up

3. Testing challenges

A. Test not available due geography, access, cost

B. Logistical issues in scheduling, performing

C. False positive/negative test limitations

D. Performance/interpretation failures

E. Equivocal results/interpretation

F. Test follow-up issues (e.g., tracking pending results)

4. Stressors

A. Time constraints for clinicians and patients

B. Discontinuities of care

C. Fragmentation of care

D. Memory reliance/challenges

5. Broader challenges

A. Recognition of acuity/severity

B. Diagnosis of complications

C. Recognition of failure to respond to therapy

D. Diagnosis of underlying etiologic cause
E. Recognizing misdiagnosis occurrence

GDP
1. Failure to follow-up
2. Limitations of a test or exam finding not appreciated
3. Disease A repeatedly mistaken for Disease B
4. Risk factors not adequately appreciated
5. Atypical presentation
6. Counter-diagnosis cues overlooked (e.g., red flags)
7. Communication failures between primary care physician and specialist
8. Issues surrounding referral
9. Urgency not fully appreciated
10. Chronic disease presumed to account for new symptoms
11. Miscommunication related to lab ordering
12. Evolving symptoms not monitored

Response 7
Based on the provided case and the taxonomies, the following codes appear applicable:

DEER Taxonomy:
2B Inaccurate/misinterpretation of history: Initially attributing symptoms to a gastrointestinal infection despite a history of surgery that is known to have complications like internal hernias.
4A Failure/delay in ordering needed test(s): The initial CT scan didn't capture the herniation, suggesting that further testing was needed.
5A Failure/delay in considering the diagnosis: The diagnosis of internal herniation post-gastric bypass is a well-recognized complication.
5B Too little consideration/weight given to the diagnosis: The initial symptoms and differential diagnosis included internal herniation, but this was sidelined by the diarrhea symptom.
5D Failure/delay to recognize/weigh urgency: The patient's clinical deterioration was significant, suggesting an urgent condition.

RDC Taxonomy:

1A Atypical presentation: Diarrhea misled the clinical team away from considering a surgical complication.
1D Findings masking/mimicking another diagnosis: The diarrhea masked the possibility of an internal hernia.
3D Performance/interpretation failures: The initial CT scan failed to capture the internal herniation.
5A Recognition of acuity/severity: The urgency of the situation was only recognized upon severe clinical deterioration.

GDP Taxonomy:
2 Limitations of a test or exam finding not appreciated: The initial CT didn't pick up the herniation.
5 Atypical presentation: The presence of diarrhea after a gastric bypass made the presentation atypical.
9 Urgency not fully appreciated: Significant clinical deterioration occurred before the urgency was recognized.

Most important code for each taxonomy:
DEER: 5A Failure/delay in considering the diagnosis
RDC: 1A Atypical presentation
GDP: 2 Limitations of a test or exam finding not appreciated

Do note that while the above codes are deduced from the provided case information and taxonomies, in real-life medical scenarios, the judgment can be more nuanced based on the entirety of the patient's presentation and clinical course.

Prompt 8
Further, please summarize the results as follows (utilizing the specific combination of numbers and capitalized characters I provided in the initial taxonomy): DEER: {insert DEER codes here} with the pivotal code being {insert most significant DEER code here}. RDC: {insert RDC codes here} with the pivotal code being {insert most significant RDC code here}. GDP: {insert GDP codes here} with the pivotal code being {insert most significant GDP code here}.

Response 8
Certainly!

DEER: 2B, 4A, 5A, 5B, 5D with the pivotal code being 5A.
RDC: 1A, 1D, 3D, 5A with the pivotal code being 1A.
GDP: 2, 5, 9 with the pivotal code being 2.