Objective evaluation of ERCP procedures: a simple grading scale for evaluating technical difficulty

K Ragunath, L A Thomas, W Y Cheung, P D Duane, D G Richards

Background and objective: Endoscopic retrograde cholangiopancreatography (ERCP) is a technically demanding endoscopic procedure that varies from a simple diagnostic to a highly complex therapeutic procedure. Simple outcome measures such as success and complication rates do not reflect the competence of the operator or endoscopy unit, as case mix is not taken into account. A grading scale to assess the technical difficulty of ERCP can improve the objectivity of outcome data.

Methods: A I to IV technical difficulty grading scale was constructed and applied prospectively to all ERCPs over a 12 month period at a single centre. The procedures were performed by two senior trainees and two experienced consultants (trainers). The grading scale was validated for construct validity and inter-rater reliability at the end of the study using the $\chi^2$ test and $\kappa$ statistics.

Results: There were 305 ERCPs in 259 patients over the 12 months study period (males: 112, females: 143, age range 17–97, mean 70.3 years). There was overall success in 244 (80%) procedures with complications in 13 (4%): bleeding in five (1.6%), cholangitis in one (0.3%), pancreatitis in five (1.6%), and perforation in two (0.7%). Success rate was highest for grade I, 49/55 (89%), compared with grade IV procedures, 8/11 (73%). There was a significant linear trend towards a lower success rate from grade I to IV ($p<0.001$) for trainees, but not for trainers. Complications were low in grade I, II, and III procedures, 12/295(4%), compared with grade IV procedures, 1/11(9%). The inter-rater reliability for the grading scale was good with a substantial agreement between the raters ($\kappa=0.68$, $p<0.001$).

Conclusion: Success and complications of ERCP by trainees are influenced by the technical difficulty of the procedure. Outcome data incorporating a grading scale can give accurate information when auditing the qualitative outcomes. This can provide a platform for structured objective evaluation.

Endoscopic retrograde cholangiopancreatography (ERCP) is an advanced endoscopic procedure that is not only technically challenging, but also associated with a risk of serious complications. The procedure itself varies from a simple, straightforward diagnostic ERCP to a highly specialised therapeutic ERCP. A recent survey in the UK showed wide differences between endoscopists in success rates (from 76% to 95%) and serious complications (0% to 16%). It is far from acceptable to have such a variation in practice and outcomes in any health care system. Studies have shown that skill and experience have a huge impact on the outcomes of ERCP.

Success and complications depend not only on the endoscopist’s experience, but also on the technical difficulty. Auditing ERCP procedures by simple outcome measures such as complication and success, without taking into account the experience of the unit and technical difficulty, does not reflect the actual competence of an individual endoscopist. Units dealing with advanced, therapeutic ERCP procedures may not be achieving the same success rates as units dealing with simple, straightforward ERCPs. There is a pressing need for objective evaluation of medical interventions. It is essential that we have yardsticks to measure the outcomes in a more objective manner. In a consensus conference, ERCP complication definitions have been standardised by Cotton et al. In addition, Fleischer et al have developed a system for classifying and grading such complications by quantifying their negative repercussions. Similarly, Howell et al have developed a system to assess technical competence and experience in ERCP procedures. A grading scale to assess the technical difficulty would complement this to achieve an overall objective evaluation. This will allow centre-to-centre and endoscopist-to-endoscopist comparisons. Our aim was to introduce and validate a technical difficulty grading scale for ERCP procedures that can be used as a simple tool to objectively evaluate the procedures.

METHOD

Demographic and procedural data relating to all ERCP procedures performed in Morriston Hospital over a 12 month period from September 1999 to August 2000 were collected prospectively. All the procedures were performed by two experienced consultants (trainers) with experience of over 1000 ERCPs and two senior trainees with experience of 100–125 ERCPs. Whenever the trainee failed, a trainer took over and completed the procedure and the final outcome was recorded. Trainees were allowed a maximum of 15 minutes; where there was no adequate progress in the procedure it was deemed a failure by the trainee. Only when the entire intended procedure (for example, sphincterotomy, stent insertion) was completed without the trainer’s assistance was the procedure classed as trainee success. Overall success was defined as completion of the intended procedure (for example, sphincterotomy, stent insertion) and not just cannulating the desired duct. Technical difficulty of the procedure was recorded according to a grading system devised by us (table 1). Complications (within 30 days) were collected from the hospital records as well as from the general practitioner records and were graded according to a grading system introduced by Cotton et al. All procedural data were collected, using specially designed forms, by the endoscopist at the end of each procedure, and finally entered.

Abbreviations: ERCP, endoscopic retrograde cholangiopancreatography; MRCP, magnetic resonance cholangiopancreatography
The grading system by Cotton [4] comprised pancreatitis in five (1.6%), bleeding (mild) in five (1.6%), and cholangitis in one (0.3%). All the complications were graded as mild according to the grading scale we have introduced is an easily understandable scenario. The success and complication rates can vary. The grading system by Cotton et al. is a summary of ERCP procedures in the various grades including the success and complication rates are given in Table 2.

Regarding hypothesis (a) our findings indicate a significant linear trend towards a lower success rate from grade I to IV (p=0.021) for trainees (table 3). Success rate for trainees decreased from 85% for grade I to 50% for grade IV, whereas this trend was not seen for trainers where success was equally good for all grades (table 3). For trainees, the difference in success rate between grade II and grade III procedures (6%) was smaller than that between grade I and grade II (14%) or grade III and grade IV (15%) (table 3).

The overall complication rate was low (4%) (table 2). There were insufficient data to test hypothesis (b) directly. However, complications were low in grade I, II, and III procedures, 12/295 (4%), compared with grade IV procedures, 1/11 (9%) (table 2). This was in agreement with the hypothesis that the greater the technical difficulty, the higher the complication rate, though the difference was not statistically significant. The overall 30 day mortality was four deaths out of 259 patients (1.5%). Three patients succumbed to their hepatobiliary or pancreatic malignancy and one patient died of respiratory failure secondary to chronic lung disease. There were no procedure related deaths.

The inter-rater reliability was good. Agreement between the two raters for grade I procedures was 93%, 86% for grade II procedures, 68% for grade III procedures, and 91% for grade IV procedures (table 4). There was substantial agreement between the raters ($\kappa=0.68$, $p<0.001$).

### DISCUSSION

Clinical effectiveness, risk management, outcomes, and audit are now a major influence on any health care provider. Undoubtedly ERCP is one of the most difficult and technically challenging endoscopic procedures. Apart from being diagnostic it can provide vital therapeutic options that can prevent or complement open surgery on the pancreatic and hepatobiliary tract. It is imperative that we have robust measures to objectively assess these procedures. The focus of attention in this study was to introduce a technical difficulty grading scale that can be easily understood and applied in everyday practice to facilitate audit and outcomes research. The success rate in ERCP not only depends on the endoscopist but also on the technical difficulty of the procedure. It is well known that not all endoscopists or endoscopy units undertake difficult therapeutic ERCP. It is not ideal to compare all endoscopists or endoscopy units as one cohort because of this case mix scenario. The success and complication rates can vary. The grading scale we have introduced is an easily understandable I to IV scale to qualitatively assess ERCP procedures, which in turn will give an overall view of the type of work done in an endoscopy unit or by an individual endoscopist.

The grading scale was able to differentiate the simplest (grade I) from the most difficult (grade IV) procedures based on the success rate, although this was not the case for grade II and III procedures. Success for trainees was highest with grade I with a significant linear trend towards lower success for higher grade procedures. However, this was not the case for the trainers. This may be due to the limited number of procedures resulting in a type II error or the scale is truly specific for the trainees. Though not statistically significant, it is interesting to note that complications were low in grade I, II, and III procedures (4%) compared with grade IV procedures (9%). This again, seems to complement the grading scale. A larger study population should be able to overcome this shortfall.

The inter-rater reliability tested as part of the validation process showed a high degree of agreement between the raters. The discriminatory power between grade II and grade III procedures of our proposed scale was weaker and there was

### Table 1: Morriston Hospital ERCP grading scale

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic ERCP</td>
<td>I</td>
</tr>
<tr>
<td>Biliary sphincterotomy (bowstring sphincterotomy), balloon sphincteroplasty, removal of extrahepatic stones ≤1 cm using basket and/or balloon</td>
<td>II</td>
</tr>
<tr>
<td>Precut sphincterotomy, large stones removal (&gt;1 cm), intrahepatic stone removal, mechanical lithotripsy, stricture dilatation, cytology, stent insertion, and nasobiliary drain</td>
<td>III</td>
</tr>
<tr>
<td>Sphincter of Oddi manometry, diagnostic and therapeutic ERCP after Billroth II surgery, minor papilla sphincterotomy, endoscopic ampullectomy and all pancreatic duct therapeutic procedures. Cholangiocopy, laser lithotripsy, electrohydraulic lithotripsy, combined procedures (PTC and ERCP), and other advanced bile duct therapeutic procedures</td>
<td>IV</td>
</tr>
</tbody>
</table>

### Table 2: Summary of all ERCP procedures according to the grading scale

<table>
<thead>
<tr>
<th>Procedure</th>
<th>No (%)</th>
<th>Success (%)</th>
<th>Complication (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade I 55 (18)</td>
<td>49 (87)</td>
<td>2 (4)</td>
<td></td>
</tr>
<tr>
<td>Grade II 86 (28)</td>
<td>65 (75.5)</td>
<td>3 (3.5)</td>
<td></td>
</tr>
<tr>
<td>Grade III 153 (50)</td>
<td>122 (80)</td>
<td>7 (5)</td>
<td></td>
</tr>
<tr>
<td>Grade IV 11 (4)</td>
<td>8 (63)</td>
<td>1 (9)</td>
<td></td>
</tr>
<tr>
<td>Total 305</td>
<td>244 (80)</td>
<td>13 (4)</td>
<td></td>
</tr>
</tbody>
</table>

### References

Cotton [4].
a lower level of agreement between raters for grade III procedures, which indicated the need to further examine the differences between grade II and grade III procedures. There was 32% disagreement for the grade III procedures. The second rater graded the procedures retrospectively and blindly based on the ERCP report, having not been present during the procedure. This could be due to the size of the stones when exact measurements of the stones are not mentioned in the report. We have not attempted to look into this since this was not anticipated and we have not incorporated this in our study protocol. Combining grade II and III procedures should further simplify the grading scale. Results of our tests on construct validity and inter-reliability of the scale showed enough evidence to justify further work to refine the scale.

We have started our grading scale with diagnostic ERCP as grade I, which may become a rarity in the future with the advent of other imaging modalities like magnetic resonance cholangiopancreatography (MRCP) and endoscopic ultrasound. MRCP and endoscopic ultrasound are not widely available, are more expensive, and need extra technical expertise. Also there are limitations—for example, endoscopic ultrasound cannot image the intrahepatic ducts and hilum with clarity when compared with ERCP. MRCP is not accurate in non-dilated ducts when compared with ERCP and can still miss small stones <5 mm. MRCP cannot be performed in claustrophobic patients and patients with pacemakers and other metal implants. MRCP and endoscopic ultrasound cannot totally replace ERCP since a select group of patients will still need diagnostic ERCP. Hence we would need a grade to classify diagnostic ERCP.

Factors such as size of papilla, position with respect to duodenum, and distortion of the duodenum are variables that can be inconsistent. Only periampullary duodenal diverticulum has been shown to increase the failure rate.7 The presence or absence of periampullary duodenal diverticulum can be added as a modifier in each grade of this grading scale, although this will increase the complexity of the grading scale.

We are aware that two centres in the United States have introduced technical difficulty grading scales. Schutz and Abbott were the first investigators to introduce a technical difficulty grading scale. They introduced a five point grading scale (I to V with increasing difficulty) and a “B” modifier to classify diagnostic ERCP. This could be a useful instrument to assist training and help training units to identify areas where individual trainee would require further support.

In conclusion, success and complications of ERCP in trainees are influenced by the technical difficulty of the procedure. Outcome data incorporating a grading scale can give accurate information when auditing the qualitative outcomes. We have made significant progress in this area. A consensus conference should be held to finalize a validated, internationally acceptable, easily understandable, reproducible and robust grading scale, which can be applied in everyday practice as a tool to objectively analyse ERCP procedures.

ACKNOWLEDGEMENT

Presented in part at the BSG meeting, March 2001, Glasgow, United Kingdom (Gut 2001;48(suppl 1):A96).

Authors’ affiliations

K Ragunath, L A Thomas, P D Duane, Department of Gastroenterology, Morriston Hospital, Swansea, UK
W Y Cheung, School of Postgraduate Studies in Medical and Health Care, University of Wales, Swansea, UK
D G Richards, Department of Radiology, Morriston Hospital, Swansea, UK

REFERENCES

1 Tanner AR. ERCP: present practice in a single region. Eur J Gastroenterol Hepatol 1996;8:145–8

Table 3 Summary of ERCP procedures performed by trainers and trainees according to the grading scale

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Trainer (%)</th>
<th>Complication (%)</th>
<th>Trainer (%)</th>
<th>Trainee (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade I</td>
<td>20/21 (95)</td>
<td>0/21 (0)</td>
<td>Trainee</td>
<td>Trainee</td>
</tr>
<tr>
<td>Grade II</td>
<td>30/37 (81)</td>
<td>1/37 (3)</td>
<td>3/49 (6)</td>
<td>4/74 (5)</td>
</tr>
<tr>
<td>Grade III</td>
<td>74/79 (94)</td>
<td>2/79 (3)</td>
<td>2/47 (5)</td>
<td>4/74 (5)</td>
</tr>
<tr>
<td>Grade IV</td>
<td>6/7 (86)</td>
<td>1/7 (14)</td>
<td>0/4 (0)</td>
<td>0/4 (0)</td>
</tr>
</tbody>
</table>

Table 4 Level of agreement between raters

<table>
<thead>
<tr>
<th>Grade I (%)</th>
<th>Grade II (%)</th>
<th>Grade III (%)</th>
<th>Grade IV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raters agree</td>
<td>51 (93)</td>
<td>74 (86)</td>
<td>104 (68)</td>
</tr>
<tr>
<td>Raters disagree</td>
<td>4 (7)</td>
<td>12 (14)</td>
<td>49 (32)</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>86</td>
<td>153</td>
</tr>
</tbody>
</table>
Objective evaluation of ERCP procedures: a simple grading scale for evaluating technical difficulty
K Ragunath, L A Thomas, W Y Cheung, P D Duane and D G Richards

Postgrad Med J 2003 79: 467-470
doi: 10.1136/pmj.79.934.467

Updated information and services can be found at:
http://pmj.bmj.com/content/79/934/467

These include:

References
This article cites 12 articles, 0 of which you can access for free at:
http://pmj.bmj.com/content/79/934/467#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Clinical diagnostic tests (395)
General surgery (168)
Surgical diagnostic tests (164)
Pancreas and biliary tract (66)
Pancreatits (18)
Radiology (418)
Radiology (diagnostics) (291)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/