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 success and complication rates do not reflect the actual 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 in 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 χ² test and κ statistics.

Results: There were 305 ERCPs in 259 patients over the 12 months study period (males: 112, females: 147, age range 17–90, 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, 1/11 (9%). There was a significant linear trend towards a lower success rate from grade I to IV (p=0.021) 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 (κ=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.
A total of 305 ERCPs were performed in 259 patients over the 12 month study period. There were 112 males and 147 females, age range 17–97 years (mean 70.3 years). The overall success was 244 (80%) with complications in 13 (4%). All the complications were graded as mild according to the grading system by Cotton (0.3%).

A total of 305 ERCPs were performed in 259 patients over the 12 month study period. There were 112 males and 147 females, age range 17–97 years (mean 70.3 years). The overall success was 244 (80%) with complications in 13 (4%). All the complications were graded as mild according to the grading system by Cotton (0.3%).

The grading scale was able to differentiate the simplest (grade I) from the most difficult (grade IV) procedures based on the success 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 tested as part of the validation study population should be able to overcome this shortfall. This again, seems to complement the grading scale. A larger study population should be able to overcome this shortfall.


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 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, though 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>
<thead>
<tr>
<th>Table 1</th>
<th>Morriston Hospital ERCP grading scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure</td>
<td>Grade</td>
</tr>
<tr>
<td>Diagnostic ERCP</td>
<td>I</td>
</tr>
<tr>
<td>Biliary sphincterotomy (bowstring sphincterotome), 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 (≥ 1 cm), intrahepatic stone removal, mechanical lithotripsy, stricture dilatation, cystology, 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, Cholangioscopy, laser lithotripsy, electrohydrolc lithotripsy, combined procedures (PTC and ERCP), and other advanced bile duct therapeutic procedures</td>
<td>IV</td>
</tr>
</tbody>
</table>

PTC, percutaneous transhepatic cholangiography.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Summary of all ERCP procedures according to the grading scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure</td>
<td>No (%)</td>
</tr>
<tr>
<td>Grade I</td>
<td>55 (18)</td>
</tr>
<tr>
<td>Grade II</td>
<td>86 (28)</td>
</tr>
<tr>
<td>Grade III</td>
<td>153 (50)</td>
</tr>
<tr>
<td>Grade IV</td>
<td>11 (4)</td>
</tr>
<tr>
<td>Total</td>
<td>305</td>
</tr>
</tbody>
</table>

into our computer database. The construct validity of the grading scale was assessed by testing the following hypothesis: (a) the greater the technical difficulty of the procedure, the lower the success, (b) the greater the technical difficulty of the procedure, the higher the complication rate. Inter-rater reliability was tested at the end of the study as follows: (i) patient details and endoscopist identification were deleted on all ERCP reports, (ii) reports were initially graded by the endoscopists who performed the procedure and then, were randomly allocated to the endoscopists for regrading without knowledge of the previous grading or the endoscopist. The statistician involved in the study (WYC) did this blinded report review. Statistical analyses were done using SPSS 10 software (SPSS Inc, Chicago, IL, USA). The \( \chi^2 \) test was used to test the hypothesis; \( \kappa \) statistics were calculated to measure the inter-rater reliability.

RESULTS
A total of 305 ERCPs were performed in 259 patients over the 12 month study period. There were 112 males and 147 females, age range 17–97 years (mean 70.3 years). The overall success was 244 (80%) with complications in 13 (4%) comprising pancreatitis in five (1.6%), bleeding (mild) in five (1.6%), perforation in two (0.7%), and cholangitis in one (0.3%). All the complications were graded as mild according to the grading system by Cotton et al. 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.
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 ultrasonound. 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. 

Schutz and Madhotra have introduced technical difficulty grading scales. Schutz and Madhotra et al have introduced a modified degree of difficulty grading scale and applied this to data from seven North American centres using the GI Trac database. They applied a three tier grading scale to 8000 ERCP procedures and concluded that success rates were lower in higher grade procedures. The grading scale introduced by the above two groups is slightly more complicated than ours, and has not been validated. Our grading scale is another preliminary attempt to grade the technical difficulty in ERCP procedures. It is limited by the number of procedures and can be further simplified by combining the grade II and III procedures. It is also noticeable that our proposed scale is particularly sensitive to the performances of trainees. Trainees could start training in grade I procedures and gradually progress up the scale as competence is gained. It 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 addressed 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(supp 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>
<thead>
<tr>
<th>Table 3</th>
<th>Summary of ERCP procedures performed by trainers and trainees according to the grading scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure</td>
<td>Trainer (%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade I</td>
<td>20/21 (95)</td>
</tr>
<tr>
<td>Grade II</td>
<td>30/37 (81)</td>
</tr>
<tr>
<td>Grade III</td>
<td>74/79 (94)</td>
</tr>
<tr>
<td>Grade IV</td>
<td>6/7 (86)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Level of agreement between raters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade I (%)</td>
</tr>
<tr>
<td>Raters disagree</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Raters agree</td>
<td>51 (93)</td>
</tr>
<tr>
<td>Total</td>
<td>55</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)
- Pancreatitis (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/