RESEARCH

ERCP cannulation success benchmarking: implications for certification and validation

D P Sheppard,¹ S J Craddock,¹ B D Warner,² M L Wilkinson^{2,3}

¹King's College London Medical School, London, UK

²Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, UK

³King's College London, London, UK

Correspondence to:

David Sheppard, King's College London Medical School London, SE1 3QD, UK; david.sheppard@kcl.ac.uk

Received 1 May 2014

Revised 22 August 2014

Accepted 27 August 2014

Published Online First

12 September 2014

ABSTRACT

Objective Investigate success rates of cannulating a 'virgin' papilla during endoscopic retrograde cholangiopancreatography (ERCP) at a tertiary referral centre; determine reasons for failure and propose learnings for consideration in future revision of success benchmarking.

Design Review of all ERCPs recorded on Endosoft database from 2006 to 2012 (n=1862). Specifically, 'virgin' papillae, defined as those with no evidence of prior surgical intervention, stents *in situ* or sphincterotomy (n=947). Virgin papillae present the most challenging target for endoscopists.

Setting Gastroenterology department, St Thomas' Hospital, London.

Patients All patients who underwent an ERCP recorded on Endosoft from 2006 to 2012 (n=1134). A proportion of these patients underwent repeat procedures, all considered virgin provided the aforementioned criteria were met.

Interventions None, retrospective audit and benchmarking exercise.

Main outcome measures Determine criteria for successful cannulation of a virgin papilla.

Results Overall success of cannulation of a virgin papilla at ERCP was 79.5%, 753 out of a total of 947 virgin papillae cases. Per patient with a virgin papilla, the success rate was 79.7%, 693 out of 869. Eliminating cases with features complicating cannulation increased success rates to 86% and 87%, respectively. Chronic pancreatitis was the single indication associated with a failed cannulation (OR=3.9, CI 2.1 to 7.1), while biliary stones were significantly associated with a successful cannulation (OR=0.3, CI 0.2 to 0.4). Reasons for failure included patient agitation (OR=27.1, CI 7.9 to 92.7), duodenal stricturing (OR=12.5, CI 5.5 to 28.5), previous anatomy-changing surgery (OR=12.2, CI 3.3 to 45.4), tumour impingement (OR=9.5, CI 4.1 to 22.3) and equipment failure (OR=7.9, CI 1.4 to 43.5).

Conclusions The Joint Advisory Group's 80% success rate for completion of therapeutic intent must be viewed in light of published difficulty rating scales, if fair comparisons and standards are to be met. This highlights the need for standardised success criterion for ERCP training and accreditation.

INTRODUCTION

Endoscopic retrograde cholangiopancreatography (ERCP) is one of the most challenging endoscopic techniques indicated in a range of pancreatic and biliary diseases. Measuring success of the procedure against an agreed benchmark is an evolving and debated topic, and one which is essential in providing a basis for assessing performance of both trainees and established practitioners.

St Thomas' Hospital was the first centre performing ERCP in the UK in the early 1970s and now serves as a tertiary referral centre performing a moderately high volume of procedures per year. The caseload is shared between three consultants, one of whom works for another Trust where he also performs ERCP.

The Joint Advisory Group (JAG) on Gastrointestinal Endoscopy published guidelines for the use of a Global Rating Scale in 2009,¹ which promotes the frequent audit of quality and safety of each mode of endoscopy. These guidelines echo the statement made in their previous endoscopy guidelines of 2007,² which advise that completion of the intended therapeutic procedure should be achieved in at least 80% of cases.

Due to caseload differences between endoscopy units, direct comparison of success rates requires standardising, with a wide variation nationally and

CrossMark

To cite: Sheppard DP, Craddock SJ, Warner BD, et al. *Frontline Gastroenterology* 2015;**6**:141–146.

internationally in both grading systems and success rates.^{3–6}

The intent of this paper is to define a straightforward and readily reproducible measure of successful cannulation of a virgin papilla, an interpretation of JAG's definition of 'initial attempt', where subsequent therapeutic or diagnostic success is treated as a separate category. Once cannulated, the therapeutic intent will be realised partially or fully, but separating the two elements of the procedure, as far as benchmarking is concerned, allows a clearer understanding of exactly what is occurring at each ERCP centre. The alternative approach of dealing with all cases on an intention-to-treat basis, and ignoring case mix, is undoubtedly simpler, but less precise.

In most proposed systems of assessment, a key differentiator is not accounted for, namely whether the papillae were virgin or otherwise prior to commencement. Virgin papillae have been defined as those unaltered by prior surgery, endoscopic sphincterotomy or stent insertion. Therefore, a patient's second ERCP attempt may again be counted among the virgin cases, provided that the ampulla has not been irreversibly altered at the first attempt. For the purpose of this study, the use of needle knife 'pre-cut' sphincterotomy is defined as an aid to a successful virgin cannulation if this is achieved at the same sitting; however, if cannulation is only achieved at a second procedure, when any oedema and/or bleeding from the pre-cut have subsided, the first procedure is defined as a failure, and the second as a successful non-virgin cannulation.

In this article, we review all ERCP procedures recorded on the Endosoft reporting software between 2006 and 2012 at St Thomas' and analyse whether particular indications or intraprocedural factors were implicated in cannulation failure. The results will hopefully aid in the future redefining of success benchmarks, with implications for national training and performance standards.

METHODS

ERCP was carried out in one of two dedicated X-ray rooms with an integrated C-arm, under sedation with Midazolam and Pethidine or Fentanyl. The use of Buscopan, glucagon and Xylocaine throat spray varied among consultants. Few patients had general anaesthesia or Propofol sedation. Endoscopes were Olympus TJF 240 or 260, and most procedures used short wire (Boston Scientific) cannulation systems.

The data were collected retrospectively from the in-house Endosoft database for all ERCPs entered ($n=1862$), of which 947 procedures were conducted on virgin papillae and the remainder were cannulations of non-virgin ducts, as outlined in figure 1. The variables analysed were selected for inclusion based on their appearance in the Endosoft database. For some Endosoft entries, there were multiple indications listed, so in an attempt to accurately reflect the raw

data and avoid any manipulation, the exact information given has been used. The results reflect this, as the number of indications outweighs the number of virgin cannulation procedures.

The data were initially analysed for 'all-comers', with success stipulating that only the desired duct was cannulated, and subsequently after removal of certain difficult categories associated with particularly high failure rates—previous failed cannulation; indications of previous Billroth II or other gastroduodenal surgery, tumour invasion of papilla or adjacent duodenal wall, duodenal stricture and pancreatic therapy. Data were analysed per procedure, per patient and individualised per consultant (A, B and C). To assess data for scan time and dose area product, a sample of 175 patients was taken from radiology records. Where a procedure failed, we recorded the reason given, if there was one stated.

The binary success data were analysed using OR, with a 95% CI as well as the probability that the difference between the means was down to chance. The OR was derived on the basis of a failed cannulation. In addition, a χ^2 analysis of the 2×2 contingency table was performed: success versus failure for each significant variable.

DISCUSSION

Indications and reasons for failure

Table 1 illustrates that chronic pancreatitis was the only indication for ERCP significantly associated with a failed cannulation (OR=3.9, CI 2.1 to 7.1), while biliary stones were significantly associated with a successful cannulation (OR=0.3, CI 0.2 to 0.4). This is in line with expectation. Pancreatitis has been highlighted as posing a high risk of cannulation failure,⁷ as ERCPs performed for acute or chronic pancreatitis management were deemed to be a 'Grade 5' standard of difficulty. By the same difficulty grading system, ERCPs for multiple stones (≥ 3) or large (≥ 1 cm) common bile duct stones are deemed 'Grade 4', as are common duct stricture dilatation. In these cases, the difficulty is in the therapy; hence, the grading is not attributable to cannulation difficulty, as supported by our results, where an indication of stricture was not significantly associated with cannulation failure (OR=0.9, CI 0.5 to 1.6).

The principal reason for failed cannulation was papillary access and included anatomical abnormalities such as duodenal stricturing, tumour impingement of the papilla or duodenum and anatomy-changing surgery such as Billroth II (OR=12.5, CI 5.5 to 28.5; OR=15.7, CI 5.1 to 47.8 and OR=12.2, CI 3.3 to 45.4, respectively) (table 1). Equipment failure (OR=7.9, CI 1.4 to 43.5) and patient agitation (OR=27.1, CI 7.9 to 92.7) were both implicated in failure; however, it is likely that these are surrogates for duration and/or difficulty of procedure, but without performing a prospective study this remains

Figure 1 Breakdown of endoscopic retrograde cholangiopancreatography (ERCP) cases by procedure and patient.

uncertain. We suggest that patient agitation, and therefore inadequate sedation, should mandate Propofol sedation to be more readily accessible. Duodenal diverticula were present in 74 of the failed cases, although may not have directly contributed to failure depending on their size and location.

We recognise certain limitations to the analysis of the indications associated with failure and reasons for failure—primarily the utilisation of retrospective data from the Endosoft reporting software. While Endosoft as a program allows the institution to determine how many of the data fields are mandatory, ours did not choose to include such fields as ‘second endoscopist’ or ‘quality of patient sedation’ as mandatory fields, hence did not always have the complete picture retrospectively (although this information is captured in a separate system). In addition, it would be preferable to have had just one indication per procedure, as this would more clearly reflect the case mix.

Success rates

Considering all of the 1862 ERCP procedures analysed, cannulation of the intended duct was successful in 1609 cases, giving an 86% success rate, as summarised in [figure 1](#). For virgin papillae, this falls to 79.5%. The overall cannulation success rate for patients who presented to the department was 83%, falling to 79.3% just considering patients with virgin papillae. This may include repeat attempts, providing no permanent change such as pre-cut sphincterotomy during previous procedures.

[Figure 2](#) demonstrates the basic success rates in each category, when all cases are included, and by eliminating cases deemed relatively inaccessible, one step at a time, shows how the success rate improves. By discounting the extremely challenging cases, such as chronic pancreatitis and previous anatomy-changing surgery, according to Schutz and Abbott,⁷ the remaining cases would correspond to a diagnostic ERCP (Grade 1) or basic therapy (Grade 2). It is then

Table 1 Statistical analysis for failed cannulation of virgin papillae

	n (%)	OR (1)	CI lwr	CI upr	p Value	χ^2	Probability	0.05
Indication								
Chronic pancreatitis	44 (5)	3.9	2.1	7.1	0.00	21.0	0.0	Significant
Biliary stones	448 (47)	0.3	0.2	0.4	0.00	52.1	0.0	Significant
Accessory duct	8 (1)	1.3	0.3	6.5	0.38	0.1	0.8	
Jaundice	194 (20)	1.3	0.9	1.9	0.11	1.6	0.2	
Normal	25 (3)	1.0	0.4	2.6	0.48	0.0	1.0	
Stricture in duct	70 (7)	0.9	0.5	1.6	0.34	0.2	0.7	
Acute pancreatitis	55 (6)	0.9	0.4	1.7	0.33	0.2	0.7	
Dilated duct	179 (19)	0.8	0.5	1.2	0.12	1.4	0.2	
Cancer	155 (16)	0.7	0.5	1.2	0.11	1.6	0.2	
Reasons for failure								
Agitation	22 (2)	27.1	7.9	92.7	0.00	60.0	0.0	Significant
Tumour impingement	19 (2)	15.7	5.1	47.8	0.00	40.7	0.0	Significant
Duodenal stricture	31 (3)	12.5	5.5	28.5	0.00	56.8	0.0	Significant
Previous surgery	12 (1)	12.2	3.3	45.4	0.00	22.2	0.0	Significant
Equipment failure	6 (1)	7.9	1.4	43.5	0.00	7.9	0.0	Significant
Other associations with failed cannulation								
Previous failure	66 (7)	1.7	0.6	4.4	0.14	1.1	0.3	
Use of Xylocaine	431 (46)	0.8	0.6	1.2	0.15	1.0	0.3	
Duodenal diverticula	74 (8)	0.7	0.3	1.3	0.11	1.6	0.2	

possible to push the success rate from below the recommended 80% to well above it. As can be seen, the way the duct is defined can greatly impact the ability to meet the JAG guidelines, and excepting the special case of chronic pancreatitis, all comes down to papillary access.

The consultants concerned have performed a minimum of 1000 procedures each, averaging >300 procedures between them per year. The apparent variation in performance shown in [table 2](#) might be attributed to the ERCP subspecialty of each consultant, with differing degrees of difficulty and types of technical skills required. For example, one of the regular lists is associated with extracorporeal shock-wave lithotripsy cases, with subsequent stone removal, another contains most of the day cases and the third covers most of the urgent inpatient procedures. The success rates could also be influenced depending on whether procedures were emergency or day case, but unfortunately this information was not available on Endosoft. By applying the same criteria of accessibility, and removal of the special case of chronic pancreatitis, the success rates of virgin duct cannulation by the three consultants (A, B and C) were 79%, 92% and 82%, respectively. If we interpret our data per patient, including instances of multiple procedures for particular patients, then the overall success rates were 84%, 93% and 86%, respectively. Additionally, we have reason to believe that the number of supervised procedures carried out by trainees may have been under-reported, as noted in our limitations.

Cannulation is the first essential step for any ERCP procedure; hence, it can be analysed separately to overall or therapeutic success. It was not the intention of this study to review therapeutic success, and therefore we must emphasise that these success rates do not give a complete picture of our service, since there is no analysis of therapeutic intent, or complication rates, and each hospital's case mix varies compared with its peers. However, the principal aim of this report is to emphasise the reproducibility of the type of duct being cannulated when reporting success.

ERCP success rates in the literature

The UK's largest audit of ERCP data by Williams *et al*,³ which included both immediate and late outcome measures, concluded that completion of intended treatment occurred in 70.4% of procedures, with 84% cannulation success at first ever attempt, which we infer as 'virgin' per our definition. Although Williams' study was intended as an 'all-comers' audit, there was quite a high exclusion rate, so the figures are not directly comparable with the present audit.

Published success rates, where the nature of the duct being cannulated is not clearly defined include: Chatterjee *et al*,⁸ who report a completion rate of 80.2%; Ragunath 87%⁴; Tanner,⁹ who in 1996 reported 88% procedural success; Penston *et al*¹⁰ who state an overall cannulation success rate of ducts which are virgin or otherwise of 95% and Cotton *et al*⁵ declaring 96% success for biliary cannulation at an academic centre. Similarly, the excellent figures from the US-centred 'ERCP quality network'⁶ are self-

Figure 2 Cumulative gap assessment.

selected and not inclusive of all cases, so the success rates must be interpreted with caution.

Considerations for future revision of success benchmarking

Our report serves to highlight the need not only for defining a revised benchmark for ERCP success, but also for reliability and reproducibility in the way we analyse cannulation/therapeutic success. We would also recommend the precise definition of the papillae included in any success rate, and that the exclusion of 'difficult' cases (including previous failures and intended pancreatic therapy) would make the criteria fairer and more realistic (figure 2).

In addition, this raises the question as to what should be the requirement for trainee and revalidation success rates, and how they should be measured.

Bearing in mind the JAG target of $\geq 80\%$ success for therapeutic intent, by defining the duct cannulated as 'virgin', two out of the three highly experienced consultants at St Thomas' did not meet the JAG requirements for cannulation alone (table 2).

We believe that our data show that the 80% JAG benchmark for therapeutic success at initial attempt for trainees, and even for established ERCP-ists, is currently very ambitious, since therapy requires cannulation to be achieved, and we know from many publications that therapeutic success is not universal after successful diagnostic ERCP.^{3 4 11} It is also worth noting that standards are only likely to get tougher given the current ERCP working party discussions within the British Society for Gastroenterology.¹² As a corollary of this work, we hope to encourage other units to publish clearer definitions when defining success.

Table 2 Success rates summary and cumulative gap assessment per consultant

	Summative factors					
	All-comers (%)	Eliminate previous failure (%)	Eliminate anatomy-changing surgery (%)	Eliminate duodenal stricture (%)	Eliminate CBD* tumour (%)	Eliminate chronic pancreatitis (%)
Consultant A (total ERCP procedures on Endosoft=445)						
Successful Virgin Cannulations	71	71	72	74	76	79
Overall Success per Procedure	80	79	80	83	86	89
Overall Success per Patient	78	78	79	80	82	84
Consultant B (total ERCP procedures on Endosoft=664)						
Successful Virgin Cannulations	84	85	85	88	89	92
Overall Success per Procedure	88	87	88	90	92	95
Overall Success per Patient	86	86	87	89	91	93
Consultant C (total ERCP procedures on Endosoft=603)						
Successful Virgin Cannulations	78	78	78	80	81	82
Overall Success per Procedure	87	85	86	87	87	87
Overall Success per Patient	82	82	83	84	85	86

ERCP, endoscopic retrograde cholangiopancreatography; CBD, Common Bile Duct.

Significance of this study

What is already known about this subject?

- ▶ This paper references the relevant body of work on the subject, and provides a platform for discussion, pertinent to the recent Joint Advisory Group Working Party on ERCP standards.

What are the new findings?

- ▶ A basis for precisely defining cannulation success, presented together with actual rates per consultant within a tertiary ERCP referral centre.

How might it impact on clinical practice in the foreseeable future?

- ▶ By agreeing a more precise definition of cannulation success, it is hoped to improve our understanding of present gaps, and a path for closure.

Acknowledgements Our thanks to James Dean STH, Dr R Ede and Dr T Wong.

Contributors DPS and SJC: data and drafting of the article. BDW: review of article. MLW: concept of the study and reviewing of article.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data have been published and included in our study.

REFERENCES

- 1 Joint Advisory Group on GI Endoscopy. A guide to auditing quality and safety items of the Endoscopy Global Rating Scale [Online]. 2009. <http://www.thejag.org.uk> (accessed Apr 2013).
- 2 Joint Advisory Group on GI Endoscopy. BSG Quality and Safety Indicators for Endoscopy [Online]. 2007. <http://www.thejag.org.uk> (accessed Apr 2013).
- 3 Williams EJ, Ogollah R, Thomas P, *et al.* What predicts failed cannulation and therapy at ERCP? Results of a large-scale multicenter analysis. *Endoscopy* 2012;44:674–83.
- 4 Ragnauth K, Thomas LA, Cheung WY, *et al.* Objective evaluation of ERCP procedures: a simple grading scale for evaluating technical difficulty. *Postgrad Med J* 2003;79:467–70.
- 5 Cotton PB, Eisen G, Romagnuolo J, *et al.* Grading the complexity of endoscopic procedures: results of an ASGE working party. *Gastrointest Endosc* 2011;73:868–74.
- 6 Cotton PB, Romagnuolo J, Faigel DO, *et al.* The ERCP quality network: a pilot study of benchmarking practice and performance. *Am J Med Qual* 2013;28:256–60.
- 7 Schutz SM, Abbott RM. Grading ERCPs by degree of difficulty: a new concept to produce more meaningful outcome data. *Gastrointest Endosc* 2000;51:535–9.
- 8 Chatterjee S, Rees C, Dwarakanath D. ERCP practice in district general hospitals in north east England—a NREG Study. *J R Coll Physicians Edin* 2011;41:109–13.
- 9 Tanner AR. ARERCP: present practice in a single region. Suggested standards for monitoring performance. *Eur J Gastroenterol Hepatol* 1996;8:145–8.
- 10 Penston J, Southern P, Daws J. ERCP in a district general hospital in England: a review of 1550 procedures over nine years. *Internet J Gastroenterol* 2009;8:1528–8323.
- 11 Zinsser E, Hoffmann A, Will U, *et al.* Rates of success and complication in diagnostic and therapeutic ERCP—a prospective study. *Z Gastroenterol* 1999;37:707–13.
- 12 Joint Advisory Group Working Party. ERCP—The way forward, a standards framework. BSG Quality and Safety Indicators for Endoscopy [Online]. 2014. <http://www.thejag.org.uk> (accessed Aug 2014).