Defining Core Patient Descriptors for Perforated Peptic Ulcer Research: An International Delphi Consensus Exercise

The PPU CDS collaborators (see Supplement 1)

Address correspondence to:

Matthew Lee

Address: FU32, Department of Oncology and Metabolism, The Medical School, University of Sheffield, Sheffield, UK, S10 2RX

Email: m.j.lee@sheffield.ac.uk

Twitter: @wannabehawkeye @segun_lamidi @petecoe1 @Kate_M_Williams

Word count: 3497

Tables:2

Figures: 3

Short running head: Core patient descriptors in PPU research

ABSTRACT

Background: Perforated peptic ulcer (PPU) remains a common condition globally with significant morbidity and mortality. Research in the field may be impaired by poor characterisation of patients, making comparison of studies and outcomes difficult. Previous work has demonstrated variation in reporting of patient characteristics in PPU studies.The aim of this study was to standardise the reporting of patient characteristics, by creating a Core Descriptor Set (CDS) of important descriptors that should be consistently reported in PPU research.

Methods: Candidate descriptors were identified through systematic review and stakeholder proposals. An International Delphi exercise involving three survey rounds was undertaken to obtain consensus on key patient characteristics for future research. Participants rated items on a scale of 1-9 on their importance. Items meeting a predetermined threshold (rated 7-9 by >70% of stakeholders) were included in the final set and ratified at a consensus meeting. Feedback was provided between rounds to allow refinement of ratings.

Results: 116 clinicians were recruited from **29** countries. 63 descriptors were longlisted from the literature, and 27 were proposed by stakeholders. After three survey rounds and consensus meeting, 27 descriptors were included in the CDS. These covered demographic and comorbidities, risk factors for PPU, presentation and pathway factors, need for organ support, biochemical parameters, prognostic tools, perforation details, and surgical history.

Conclusions: This study defines the core descriptive items for PPU research. Delphi methodology can be used to define this type of dataset. The CDS could support more future research and allow more robust synthesis.

Introduction

Perforated peptic ulcer (PPU) is an emergency surgical condition which occurs due to a perforation in either the stomach or duodenum[1]. Despite the advent of proton-pump inhibitors, and decreasing rates of Helicobacter Pylori, this remains an important condition[2], with particularly high rates in low and middle income countries[1]. Patients with perforation often have peritonitis, and therefore significant morbidity and mortality. The evidence base for the treatment of PPU is limited by the quality of studies[3]. Understanding patient characteristics and their roles in heterogeneity of treatment effects, direct or indirect, might help to stratify patients and treatment strategies. Previous work has reviewed 23 studies of PPU treatment, which identified 76 unique descriptors and a range of 4-22 descriptors reported in each study[4].

The lack of clarity when reporting patient characteristics in randomised trials has been identified as a challenge[5]. Guidelines for trial reporting (CONSORT) recommend that the first table in a study reports baseline demographic and clinical characteristics for each group. This table should "allow readers, especially clinicians, to judge how relevant the results of a trial might be to an individual patient"[6]. However, its implied function is to reassure the reader that baseline imbalances in prognostic characteristics are not so great that they – rather than treatment allocation - might account for different results[7]. Implicit in both claims are condition-specific theories about heterogeneity of treatment effects[8]. For any trial, the items in Table one embody claims about why treatments might work differently for different people. Science advances by testing hypotheses so, for medical research to progress, it is worth making such claims explicit and subject to consensus within specialties[9]. To do so enables prognostic research that can lead to evidence-based stratified care[10]. It also aids assessment of external validity across various healthcare settings such as low and middle income countries.

The aim of this study was to generate a consensus on the core patient characteristics (descriptors) for PPU studies.

Methods

This research project was approved by the University of Sheffield Research Ethics Committee (Application 034049). This study was developed with reference to the Core Outcome Set-STAndards for Development (COS-STAD) recommendations[11] and reported in line with Core Outcome Set-STAndards for Reporting (COS-STAR) guidelines[12].

Scope

This study aimed to develop a global consensus on clinically important patient characteristics to be reported in adult studies of PPU disease. This includes patients undergoing surgical or conservative treatment of a perforated ulcer. A core descriptor set (CDS) describes what descriptors should be measured but does not recommend how they should be measured. A CDS can be used for all research types. A summary of the study design is presented in figure 1.

Steering group

A steering group of 15 clinicians was established. 13 (80%) were clinicians from the United Kingdom and two (20%) from Italy. Members were all consultant surgeons, with publications in emergency surgery, or leadership roles in gastrointestinal or emergency surgery organisations. The steering group provided direction and oversight for the study.

Stakeholders

Patients and healthcare professionals were consulted separately during the development of the core descriptor set. Healthcare professionals contributing were general surgeons or anaesthetists with more than one year of postgraduate experience in the management of perforated peptic ulcer disease. Participants were recruited internationally. A public and patient involvement group in emergency care contributed to the study, providing feedback on burden of descriptor set, and perceived relevance.

Study design

Delphi methodology was used to gain consensus on the important descriptors in PPU. Delphi is a widely used methodology for establishing reliable consensus from a group of experts. It uses a series of questionnaires/surveys interspersed with controlled feedback[13].

The survey was disseminated via mailing lists and social media. World Society of Emergency Surgery (WSES) and Association of Surgeons of Great Britain and Ireland (ASGBI) promoted the projects, advertising through weekly emails and/or website pages. A second and third survey were sent directly to emails of clinicians that responded in the previous rounds. On completion of all three rounds an online consensus meeting was held to vote on the final list of eligible descriptors.

Longlisting of descriptors

The initial longlist of descriptors were sourced from a systematic review of patient characteristics in randomised clinical trials of peptic ulcer repairs[4].

Inclusion of new Descriptors

Inclusion and exclusion processes for new descriptors were determined *a priori*. There were two opportunities for descriptors to be added to the initial list sourced from the systematic review. Members of the steering group provided suggestions prior to finalisation of the list of descriptors to be used in the first round. During the first survey round, responders were asked to suggest descriptors they thought should be included for consideration. Any descriptors suggested at both times were immediately included into the list unless determined by the steering group to be vague, already described, irrelevant to PPU, related to the treatment process or not a patient descriptor.

Consensus process and consensus definition

A three-round online Delphi process was used to gain consensus on important descriptors. **This followed practice in previously completed Delphi studies**[14]. In each round clinicians were asked to rate descriptors on their importance in PPU, on a 9-point Likert scale. 1-3 signified a 'Not important' descriptor, 4-6 an 'Important but not critical' descriptor, and 7-9 signified an 'Important and critical' descriptor. Descriptors rated 7-9 by ≥70% of clinicians were removed from subsequent rounds and formed a list of 'consensus descriptors' to be discussed in the consensus meeting. Descriptors rated 1-3 by ≥70% of clinicians were removed from the longlist of descriptors and not included in subsequent rounds.

Round One

Round one was open to responses for three months, this was to allow a sufficient number of clinicians to complete the survey. During round one clinicians were asked to rate descriptors on their importance in the management of PPU. Participants were able to suggest new descriptors for inclusion in subsequent rounds. Descriptors were presented in a randomised order on a single page in order to minimise bias[15].

Principal component analysis (PCA) was performed on round one results to organically cluster descriptors based on patterns in voting by responders. **PCA is a dimension reduction technique which can be used to reduce a list of many items into a shorter list. It creates new unifying items or groupings whilst retaining the dimensions or coverage of the longlist. It does this by comparing the variation in measurements (i.e ratings) of items across raters and between the items themselves. Essentially, items where ratings show similar behaviours or trends can group together.**

Rounds Two and Three

Round two and three were open to responses for one month each. Clinicians needed to have completed the previous round to be eligible to complete the next rounds. As well as access to the survey, clinicians were provided with results from the previous round and how this differed to their own ratings. As a vital part of the Delphi process, feedback between rounds allows consolidation of personal ratings as well as an awareness of how it fits with the generic opinion of the cohort. This reflection leads to consensus with each subsequent round. Feedback was presented as a Likert scale and median rating for each descriptor alongside their rating for the descriptor (Supplementary file 2). Descriptors suggested in round one were included in the second and third rounds (if not meeting the consensus threshold).

Participants were made aware that completion of all survey rounds were required for collaborative co-contributorship.

Patient and Public Involvement

Descriptors meeting the consensus threshold during the survey rounds were presented to a patient and public involvement group. They assessed the acceptability and burden of measuring these 'consensus descriptors'. This discussion was held online during their forum meeting after the completion of round three and prior to the consensus meeting. PPI was delayed up to this point, due to the nature of descriptors having a role in prognosis and decision making. It was felt only clinicians with their experience and insight could provide valid ratings on the importance of these descriptors. The PPI group was consulted on the appropriateness of delaying their involvement to this point in the study.

Consensus meeting

A number of survey participants were invited to participate in a consensus meeting to finalise the CDS. **There was no set criteria for participation in the consensus meeting apart from completion of all 3 rounds. Participants completing all three survey rounds were asked to indicate if they would be willing and able to participate in a consensus meeting. Of those voting yes, selection was based on role and location with the intention of getting good geographical representation. Participant names and previous round ratings were not available to the team members reviewing the long list. Members were selected to be reflective of the real word population to whom the results will be applied to. A sample of 10 participants was selected as a pragmatic group size to facilitate discussion in an online setting.**

A number of suggestions were presented to the consensus group to vote on. A threshold of 80% agreement was required for any suggestion to be accepted. If unsuccessful a period of discussion was allowed to collect comments on the proposed vote. New suggestions could be voted on based on the results of the discussion. The purpose of the suggestions was to refine the phrasing and clustering of descriptors. Descriptors having consensus in the survey rounds could not be removed, but could be combined as long as both descriptors were equally represented in the refined descriptor.

Principal component analysis

PCA is a statistical method that identifies patterns in the correlation between variables. It is used to reduce the dimensionality of highly dimensional data sets [16]. This method can be used to identify and group descriptors representing the same concepts[17]. Using PCA meant groupings of descriptors were determined by survey participant ratings. Having groups developed in this manner prevents the preconceived ideas of the research team influencing structure or

arrangement of the surveys. The position of items has been shown to influence participant ratings [15], therefore it was vital that this potential source of bias was removed.

Ratings from round one were entered into SPSS (Version 23) [18] and PCA was conducted using varimax rotation. Reliability tests measure the suitability of using PCA on the data set. Kaiser-Meyer-Olkin Measure (KMO) of Sampling Adequacy indicates the proportion of variance that might be caused by underlying factors. High values (close to 1.0) indicate that a factor analysis will be reliable in the data set. Bartlett's test of sphericity tests indicates if variables are unrelated and therefore unsuitable for structure detection. Small values (<0.05) indicate a factor analysis will be useful. Communalities, which are a check for correlation analyses indicate specific variables that do not fit well with the others. A communalities value of <0.6 was determined for removal of descriptors with bad fit [19]. Components were extracted using the criteria eigenvalue ≥1. Descriptors with rotated component matrix value >0.6 to the same component were grouped together. Components were reviewed by the research team to rationalise where cross loading of items was noted.

During round two and three descriptors were presented in a random order within the components identified in PCA. One component was presented per page, **labelled alphabetically by descending Eigenvalue**. During the consensus meeting, descriptors could be reorganised and groupings renamed if the consensus threshold of 80% was met.

Attrition and Sample Size

The sample size for a Delphi panel cannot be statistically determined but good results can be developed from a relatively small homogenous group of experts[20] . Using numbers recruited in other similar Delphi studies, it was determined that 100 was a realistic target [14].

Results

Stakeholders

A total of 116 clinicians responded to the first survey. The survey link was accessed on 772 occasions, a conversion rate of 15%. It is unclear what proportion of clicks were unique or what proportion of those that accessed the link were eligible to submit a response. Responses came from six geographical areas (Asia, Europe, Africa, Australia, North America and South America) spanning **29** countries. The vast majority of responders were from the UK (47%) and Italy (16%) and South Africa (3%). Of the 116 clinicians that responded 89 (77%) were consultant surgeons, 26(22.4%) surgical trainees, and one (0.9%) consultant anaesthetist. In round two, 86/116 (74.1%) clinicians completed a response. In round three, 80/86 (93%) clinicians responded. Overall 69% of all participants completed all three survey rounds (Table 1).

Descriptor list

The systematic review identified 76 unique descriptors used in PPU trials. These 76 descriptors were collated and paraphrased to create an initial list of 63 descriptors for consideration. These 63 descriptors were included into the first survey and reviewed by members of the steering group. Steering group suggested an additional descriptor to be included in the longlist, leaving a total of 64 descriptors in the first round.

Figure 3 illustrates the flow of descriptors through the study. Prior to the second round, 17 (round one and newly suggested) descriptors were combined by the steering group to form an amalgamated list of seven descriptors. 27 suggested descriptors were added. Seven descriptors reaching the consensus threshold in round one were removed. A total of 74 descriptors were rated in round two. Prior to round three, 16 descriptors meeting the consensus threshold in round two were removed. A total of 57 descriptors were rated in round 3. No descriptors met the threshold to be dropped in any survey round.

Clustering of descriptors

At 0.887, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy was satisfactory and Bartlett's test was significant ($p<0.001$). Therefore the sample size at n=116 was sufficient for PCA. Communalities for all descriptors were >0.6. Using the Eigenvalue method, 10 components were identified. Descriptors with a rotated component matrix value >0.4 were assigned to the corresponding component. This was possible for 36 (56.3%) descriptors. 20 (31.3%) descriptors demonstrated cross-loading between components. These were preferentially placed in the component with the highest correlation value. Seven (10.9%) descriptors despite correlating better with a component were clustered elsewhere. As their component matrix values were low and similarly correlated to multiple components, the decision on their placement was decided by the research team. Body mass index was the sole descriptor not to enter a component.

All 10 components were labelled after analysing the descriptors that were clustered together. Acute organ failure (16 descriptors), risk factors for PPU (10 descriptors), perforation details (six descriptors), chronic disease (nine descriptors), laboratory tests (seven descriptors), prognostic tools (four descriptors), presentation & pathway factors (seven descriptors), immunocompromised (two descriptors), Demographics (two descriptors). BMI remained outside of these groups.

Protocol modification

It was envisioned that only descriptors meeting the consensus thresholds would be discussed in the consensus event. However, six descriptors came within 5% of the threshold in round three. In order to be robust and allow for a margin of error caused by a reduced sample size, these descriptors were voted on for inclusion in the CDS. The research team determined an agreement threshold of 80% was needed for inclusion. A higher threshold than during the survey rounds was

required as fewer (n=10) participants contributed to the vote. Along with a period of discussion, this allowed for greater scrutiny over descriptors that had failed to meet the consensus threshold over multiple rounds. (Figure 2)

PPI feedback

No disagreements were raised with the descriptors to be included in the CDS. PPI feedback indicated that the burden of measuring all the descriptors was not perceived as great and should not pose a problem for trialists. The PPI group agreed it was not appropriate for the public in general to vote on the importance of prognostic descriptors.

Consensus meeting

26 descriptors met the consensus threshold during the three survey rounds. Presence of diffuse peritoneal contamination (met threshold in round one) and global score or peritoneal contamination were combined as presence of generalised peritonitis which was included in the CDS. Six borderline descriptors were voted on in the consensus meeting, only 'presence of immunosuppression' met the required threshold, and was included in the CDS. Several descriptors were rearranged from their original clusters and two groups were renamed. **This was undertaken by the consensus group.**

Twenty-seven descriptors were included in the core descriptor set. 26 descriptors could be clustered into seven groups. Need for organ support (three descriptors), risk factors for PPU (five descriptors), perforation details (four details), demographics and comorbidities (four descriptors), biochemical parameters (two descriptors), Prognostic tools (four descriptors), and presentation and pathway factors (four descriptors). Previous gastric surgery did not cluster well with other descriptors. See summary in table 2.

Discussion

This study has used a novel, multimodal approach, with international stakeholders to agree on a list of common descriptors to be used in research in PPU. It has used PCA to explore the structure of ideas behind this data. Twenty seven descriptors were included in the final CDS.

A previous prognostic review undertook meta-analysis of prognostic factors in PPU[21]. Many of the factors associated with mortality in the review are identified as core descriptors here, lending credibility to this approach. Many of these factors relate to chronic health and physiological reserve, such as ASA status, age, comorbidities. Acute physiology is represented with aspects of cardiovascular dysfunction, organ failure, and the need for critical care support, covered in several clusters. Participants felt that the risk factors for PPU should be clearly documented, including malignant and drug related causes. There was explicit feedback that the differing drug aetiologies should be kept as separate items as these might be associated with different patient phenotypes. For example, a patient with PPU from NSAIDs might expect a different trajectory to a patient with PPU resulting from steroid use, due to the effects of steroids on tissue quality and immune function. **It is worthwhile remembering that as well as varied functions, drugs may also be proxies for underlying diseases with impact on outcomes. In the Delphi, presence of a perforated cancer was considered an important descriptor, possibly as malignancy is not always apparent and may lead to treatment as a 'simple' perforated ulcer, but may be associated with failure of repair.** In addition, collaborators have identified factors related to technical aspects of treatment including site and size of perforation, and whether previous gastric surgery had taken place. Presumably these items inform operative planning related to laparoscopic approach, or the need for distal gastrectomy; both of which might be considered in appropriate patient groups [22,23].

Although frequently reported[4], patient sex is not associated with outcomes, and is not included in the CDS. Respondents felt that this factor was not material to treatment decisions or risk stratification for this population. **Whilst it may not have a role as a core prognostic factor, this does not preclude the importance of recording sex and gender in studies to ensure equity of representation[24].** It is notable that smoking status, which was included in the Delphi and is associated with mortality[21], is not considered important here. This may be a reflection of generally declining use of tobacco products[25], leading to the perception that the characteristic is relatively infrequent. **This decline may be more common in higher income countries which accounted for significant voting numbers in early rounds, potentially removing some LMIC influence.**. It is also notable that the presence of *H. pylori* is not considered relevant for inclusion here, despite its well documented role in the development of peptic ulceration[26]. Participants felt that this was not a factor that would change their management, given that testing and results were only available after an initial treatment strategy had been selected.

The value of international engagement is clearly demonstrated in the consensus meeting where the borderline descriptor 'immunosuppression' was discussed. This was considered a potential 'treatment effect mediator'. Treatment effect mediators are patient characteristics measured at baseline or during treatment that impact on outcome. These characteristics may or may not interact with intervention choice to predict response to treatment[27]. Representatives from LMIC health systems reported that they would consider immunosuppression an important prognostic tool, given the high prevalence of Human Immunodeficiency Virus in their practice. Representatives from HICs noted that they would consider this important too, albeit for different reasons. Patients with underlying malignancy might develop a peptic ulcer during chemotherapy, and experience similarly poor outcomes.

This study is not without limitations. As with all consensus exercises, there may be some form of responder bias, meaning that people with a special interest or focus on the field may have self selected. This could lead to a consensus at odds with the general clinical population. This study predominantly recruited from a single clinical group (surgeons) and did not have much in the way of representation from the wider clinical or multi-disciplinary team. The input of anaesthetists, intensivists, nurses, and allied health groups might again have altered this. **Whilst efforts were made to engage with societies, these were unsuccessful. Future perioperative work should ensure input of these key stakeholders.** However, in much of the world, research and initial prognostication for this disease is typically performed by surgeons. **There is also the possibility that inclusion of some items may have been impacted by dominance of respondents from high income countries or LMICs. Future work should consider strategies to address this, perhaps by treating them as individual panels.**

The multi-method approach used here contributes to the strength of the work. The international consensus approach aggregates the tacit knowledge of the clinical community, and aids external validity of the CDS. The use of PCA helps us to understand the structure of data and begin to understand ideas related to disease and pathology underlying the voting. This approach could be adopted for other clinical problems where consistent description of study populations is needed to understand treatment effect heterogeneity.

Researchers might also see additional benefits from use of this CDS. Consistent reporting of descriptors across studies will lead to more reliable meta-analysis by allowing exploration of treatment effect heterogeneity caused by patient factors. Conduct of a large cohort of patients using this descriptor set might allow the identification of key subgroups of patient phenotypes with worse outcomes, which might explain treatment heterogeneity effects. This might be achieved using approaches such as latent class analysis[28].

For clinicians, this CDS might be used to provide a standard reporting framework for patients receiving treatment for PPU. This could aid robust comparison of characteristics across different institutions, regions, and nations. Comparisons such as these might enable us to identify areas with comparatively higher risk populations, and target resources appropriately. Finally, consistent reporting of patient descriptors which are prognostic in themselves, as well as informing treatment decisions, will allow robust comparison and benchmarking between clinical sites.

Conclusion

This study has identified 27 items which should be consistently reported in future research on PPU. This method might be adapted to other disease areas to improve consistency of reporting.

Acknowledgements

The authors want to thank the World Society of Emergency Surgery (WSES), Association of Surgeons of Great Britain and Ireland (ASGBI) and Sheffield Emergency Care Forum (SECF) for their support and contribution to the study.

Declarations

Availability of data and materials - No data are available for sharing. All findings presented in this manuscript.

Competing interests - No competing interests declared by any authors.

Funding - This study was unfunded

References

1. Søreide, K. *et al.* Perforated peptic ulcer. *Lancet* **386**, 1288–1298 (2015).
2. Hooi, J. K. Y. *et al.* Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. *Gastroenterology* **153**, 420–429 (2017).
3. Collaborators, E. *et al.* Methodological overview of systematic reviews to establish the evidence base for emergency general surgery. *Br. J. Surg.* **104**, 513–524 (2017).
4. Lee, M. J., Coe, P. O., O'Donoghue, R., Peirson, M. & Saha, A. Variation in descriptors of patient characteristics in randomized clinical trials of peptic ulcer repair: a systematic review. *Br. J. Surg.* **107**, 1570–1579 (2020).
5. Nguyen, T.-L. & Xie, L. Incomparability of treatment groups is often blindly ignored in randomised controlled trials – a post hoc analysis of baseline characteristic tables. *Journal of Clinical Epidemiology* vol. 130 161–168 (2021).
6. Schulz, K. F., Altman, D. G., Moher, D., Group, C. & Others. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c332 (2010).
7. Roberts, C. & Torgerson, D. J. Understanding controlled trials: Baseline imbalance in randomised controlled trials. *BMJ* vol. 319 185–185 (1999).
8. Ransohoff, D. F. Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer* **5**, 142–149 (2005).
9. Gillies, D. *Causality, Probability, and Medicine*. (Routledge, 2018).
10. Hemingway, H. *et al.* Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* **346**, e5595 (2013).
11. Kirkham, J. J. *et al.* Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med.* **14**, e1002447 (2017).
12. Kirkham, J. J. *et al.* Core Outcome Set-STAndards for Reporting: The COS-STAR Statement. *PLoS Med.* **13**, e1002148 (2016).
13. Okoli, C. & Pawlowski, S. D. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management* **42**, 15–29 (2004).
14. Sahnan, K. *et al.* Developing a core outcome set for fistulising perianal Crohn's disease. *Gut* **68**, 226–238 (2019).
15. Brookes, S. T., Chalmers, K. A., Avery, K. N. L., Coulman, K. & Blazeby, J. M. Impact of question order on prioritisation of outcomes in the development of a core outcome set: a randomised controlled trial. *Trials* **19**, 1–11 (2018).

16. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
17. Todhunter, F. Using principal components analysis to explore competence and confidence in student nurses as users of information and communication technologies. *Nursing Open* **2**, 72–84 (2015).
18. IBM Corp. *IBM SPSS Statistics*. (2020).
19. IBM Docs. https://www.ibm.com/docs/en/spss-statistics/23.0.0?topic=detection-kmo-bartletts-test.
20. Akins, R. B., Tolson, H. & Cole, B. R. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Med. Res. Methodol.* **5**, 37 (2005).
21. Møller, M. H., Adamsen, S., Thomsen, R. W. & Møller, A. M. Preoperative prognostic factors for mortality in peptic ulcer perforation: a systematic review. *Scand. J. Gastroenterol.* **45**, 785–805 (2010).
22. Chung, K. T. & Shelat, V. G. Perforated peptic ulcer - an update. *World J. Gastrointest. Surg.* **9**, 1–12 (2017).
23. Coe, P. O., Lee, M. J., Boyd-Carson, H., Lockwood, S. & Saha, A. Open Versus Laparoscopic Repair of Perforated Peptic Ulcer Disease: A Propensity-matched Study of the National Emergency Laparotomy Audit. *Ann. Surg.* (2020) doi:10.1097/SLA.0000000000004332.
24. Tadiri, C. P. *et al.* Methods for prospectively incorporating gender into health sciences research. *J. Clin. Epidemiol.* **129**, 191–197 (2021).
25. Organization, W. H. & Others. *WHO global report on trends in prevalence of tobacco use 2000-2025*. (World Health Organization, 2019).
26. Marshall, B. & Robin Warren, J. UNIDENTIFIED CURVED BACILLI IN THE STOMACH OF PATIENTS WITH GASTRITIS AND PEPTIC ULCERATION. *The Lancet* vol. 323 1311–1315 (1984).
27. Kraemer, H. C., Wilson, G. T., Fairburn, C. G. & Agras, W. S. Mediators and moderators of treatment effects in randomized clinical trials. *Arch. Gen. Psychiatry* **59**, 877–883 (2002).
28. Nielsen, A. M., Kent, P., Hestbaek, L., Vach, W. & Kongsted, A. Identifying subgroups of patients using latent class analysis: should we use a single-stage or a two-stage approach? A methodological study using a cohort of patients with low back pain. *BMC Musculoskelet. Disord.* **18**, 57 (2017).

Tables

Table 1: Demographic of survey participants

Demographic of survey participants				
Characteristics	Round 1 N=116	Round 2 N=86	Round 3 N=80	Consensus N=10
Gender				
Male	91 (78.4%)	69 (80.2%)	64 (80%)	9 (90%)
Female	25 (21.6%)	17 (19.8%)	16 (20%)	1 (10%)
Country				
United Kingdom	54 (46.6%)	35 (40.7%)	30 (37.5%)	2 (20%)
Italy	18 (15.5%)	15 (17.4%)	15 (18.8%)	1 (10%)
South Africa	4 (3.4%)	3 (3.5%)	3 (3.8%)	
Ireland	3 (2.6%)	1 (1.2%)	1 (1.3%)	
Romania	3 (2.6%)	2 (2.3%)	2 (2.5%)	1(10%)
Turkey	3 (2.6%)	3 (3.5%)	3 (3.8%)	
Greece	3 (2.6%)	3 (3.5%)	3 (3.8%)	2 (20%)
Other	28 (24.1%)	24 (27.9%)	23 (28.8%)	4 (40%)
Position				
Consultant Surgeon	89 (76.7%)	68 (79.1%)	63 (78.8%)	9 (90%)
Surgical trainee (>PGY1)	26 (22.4%)	17 (19.8%)	16 (20%)	1 (10%)
Consultant anaesthetist	1 (0.86%)	1 (1.2%)	1 (1.3%)	

Table 2: Descriptors included in the Core descriptor set

	Descriptors	Cluster name
1 2 3 4	Presence of chronic comorbid disease Age Use of Anticoagulants Presence of immunosuppression	A Demographics and comorbidities
5	Previous gastric surgery	
6 7 8 9 10	Diagnosis of upper gastro-intestinal cancer Use of ulcerogenic drugs NSAID use Steroid use Previous peptic ulcer disease	B Risk factors for PPU
11 12 13 14	Time from symptoms to admission Time from admission to theatres (if operated on) On admission systolic blood pressure On admission heart rate	C Presentation and pathway factors
15 16 17	Presence of organ failure ITU/HDU support requirement on admission Invasive ventilation requirement on admission	D Need for organ support
18 19	Score on an intensive care mortality predictor tool On admission Lactate	E Biochemical parameters
20 21 22 23	Score on a post-operative mortality predictor tool Score on a measure of frailty tool Summary score of comorbidities ASA score	F Prognostic tools
24 25 26 27	Presence of generalised peritonitis Type of peritoneal contamination at operation (if operated on) Site of perforation Size of defect	G Perforation details

Figures