TITLE PAGE

Title:

Selection of quality indicators to evaluate quality of care for patients with esophageal atresia using a Delphi method

Authors and affiliations:

Nadine M. Teunissen[1,2,3] MD, Julia Brendel[1,4] MD, Ernest L.W. van Heurn[1,5] MD PhD, Benno M. Ure[1,4] MD PhD, René Wijnen[1,3] MD PhD, Simon Eaton[1,3,6] PhD on behalf of the EPSA|ERNICA Registry Group[#] and the EA Quality of Care Initiative*

1 European Reference Network for rare Inherited and Congenital Anomalies (ERNICA)

2 Dutch Institute for Clinical Auditing, Leiden, The Netherlands

3 Department of Pediatric Surgery, Erasmus University Medical Center - Sophia Children's Hospital, Rotterdam, The Netherlands

4 Department of Pediatric Surgery, Hannover Medical School, Hannover, Germany

5 Department of Pediatric Surgery, University of Amsterdam and Vrije Universiteit, Emma Children's Hospital, Amsterdam University Medical Center, Amsterdam, The Netherlands

6 UCL Great Ormond Street Institute of Child Health, London, United Kingdom

See the online supplementary appendix for a list of the EPSA|ERNICA Registry Group members and their affiliated hospitals. Members of the EPSA|ERNICA Registry Group should be listed as authors and should be Pubmed citable

* See the online supplementary appendix for a list of the EA Quality of Care Initiative members and their affiliated hospitals or organizations. Members of the EA Quality of Care Initiative should be listed as authors and should be Pubmed citable

Corresponding author:

Nadine Teunissen, MD

Department of Pediatric Surgery

Erasmus University Medical Center – Sophia Children's Hospital

Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

Telephone: +31 (0) 6 52 55 26 15

E-mail: n.teunissen@erasmusmc.nl

Contributions

All authors contributed to the study protocol. NT and JB drafted the manuscript. SE, BU, and RW provided clinical expertise on EA care and research. All authors read, provided feedback, and approved the final manuscript.

ABSTRACT

Aim of the study

Survival of neonates with esophageal atresia (EA) is relatively high and stable, resulting in increased attention to optimizing care and longer-term morbidity. This study aimed to reach consensus on a quality indicator set for benchmarking EA care between hospitals, regions, or countries in a European clinical audit.

Methods

Using an online Delphi method, a panel of EA health care professionals and patient representatives rated potential outcome-, structure- and process indicators for EA care identified through systematic literature and guideline review on a nine-point Likert scale in three questionnaires. Items were included based on predefined criteria. In rounds two and three, participants were asked to select the five to ten most essential of the included indicators.

Main results

An international panel of 14 patient representatives and 71 multidisciplinary health care professionals representing 41 European hospitals completed all questionnaires (response rate 81%), eventually including 22 baseline characteristics and 32 indicators. After ranking, ten indicators were prioritized by both stakeholder groups. In addition, each stakeholder group highly prioritized one additional indicator. Following an additional online vote by the other group, these were both added to the final set.

Conclusions

This study established a core indicator set of twenty-two baseline characteristics, eight outcome indicators, one structure indicator, and three process indicators for evaluating (quality of) EA care in Europe. These indicators, covering various aspects of EA care, will be implemented in the European Pediatric Surgical Audit to enable recognition of practice variation and focus EA care improvement initiatives.

INTRODUCTION

Esophageal atresia (EA) is a rare congenital anomaly affecting approximately one newborn in 3000 to 4000 births[1], which requires surgical repair, generally within the first days of live. The mortality rate is low and usually determined by co-morbidity after decreasing considerably over the last decades of the 20th century[2]. Subsequently, this caused a shift of focus in EA research and care, now acknowledging the importance of longer-term morbidity and quality of care.

Clinical auditing is a method to ascertain high quality of care, using quality indicators to identify, monitor, and evaluate variation in clinical practice and outcomes[3]. The European Pediatric Surgical Audit (EPSA) is such an audit. In 2014, it was established as a Dutch, nationwide, prospective quality assurance system for six congenital malformations: esophageal atresia, congenital diaphragmatic hernia, Hirschsprung's disease, anorectal malformation, gastroschisis, and omphalocele. In 2020 it expanded internationally through funding of the European Reference Network for Inherited and Congenital Anomalies (ERNICA), a project of the European Commission to promote knowledge of and care for rare diseases. Currently, 18 hospitals across 10 European countries register their patients born with EA or one of the other conditions in this audit.

Determining a set of quality indicators is paramount in monitoring and evaluating esophageal atresia care. Many quality indicator sets have been developed for conditions with a higher prevalence, such as hip fractures or esophageal cancer[4,5]. However, as clinical auditing in rare diseases is uncommon, evidence-based quality indicators for such conditions do not exist. To fill that gap, quality indicators for all conditions in the EPSA - including EA - were initially established through an expert consensus meeting of Dutch

pediatric surgeons at the inception of the Dutch nationwide audit in 2014[6]. These quality indicators reflect Dutch care principles and care, but they may not be clinically relevant and applicable in other participating countries. In addition to this international endorsement by pediatric surgeons, it is crucial to ensure wider validity amongst other specialties caring for infants and children with EA and amongst patient representatives.

This study aimed to develop an internationally applicable and supported comprehensive set of structure-, process- and outcome indicators that will be used to monitor, evaluate, and compare the quality of lifelong care for patients with esophageal atresia. This core indicator set should also comprise patient characteristics and characteristics of treatment and care processes to enable the interpretation and correction of indicator results. In time, implementing this "EA Core Indicator Set" in a clinical audit, such as the EPSA/ERNICA registry, may lead to a better understanding of EA care and ultimately improve the quality of care throughout Europe.

METHODS

The modified Delphi study design

The EA core indicator set was developed using a modified Delphi method: a consensus method that aims to obtain the most reliable consensus within a group of experts through anonymous voting, thereby preventing the dominance of the views of a select few[7–9]. This method builds on the assumption that opinions converge in the light of peers' opinions[9,10]. In several subsequent rounds, surveys are distributed to persons considered experts. After every round, a summary of all (anonymous) votes is fed back to the experts, enabling them to adapt their answers to this new information in the following rounds.

Finally, after three rounds, if consensus has not been reached, a consensus meeting can be held, discussing and taking a final vote on the remaining indicators.

Generating item list

An extensive literature search for known quality indicators in EA-care was performed in Medline, Embase, and the Cochrane Library in collaboration with an experienced librarian, which yielded no results apart from four indicators developed by Dutch pediatric surgeons in 2015[6]. A systematic review was then performed of literature concerning the primary EA care process, published between 2015 and 2021, extracting all studied patient characteristics, treatment- and care process characteristics, and outcomes[11]. Identified parameters were included on the item longlist if studied in more than 5% of all included articles. Care process characteristics and outcomes were then translated to quality indicators by a workgroup of five pediatric surgeons and one patient representative by defining underlying concepts, specifying a numerator and denominator, and determining a fixed time of measurement. Definitions for characteristics were determined similarly. Additionally, recommendations of the ESPGHAN guidelines[12] and the ERNICA consensus statements on esophageal atresia[13,14] were translated to quality indicators by JB and NT and subsequently added to the item list as process indicators. Finally, two patient representatives reviewed the longlist for clarity of wording and comprehension and adapted it if necessary.

Panel selection

To maximize representativeness, the participating panel consisted of two stakeholder groups: health care professionals with experience in EA care (1) and patient representatives active in EA patient organizations throughout Europe (2). Per stakeholder

group, a minimum of 20 participants representing at least five European countries was pursued to increase international generalizability and support of the results.

All hospital representatives of medical centers in the ERNICA network were approached via e-mail, containing a brief study introduction and a link to preregister for participation, which was available for approximately ten weeks. It requested the recipient to forward this invitation ("snowballing") to their peer experts in EA care, including but not limited to pediatric surgeons, gastroenterologists, pediatricians, pulmonologists, nurse specialists, speech- and language therapists, and physical therapists. The invitation for preregistration specifically requested the participation of health care professionals with experience in EA management; however, no minimum number of years of experience nor proof thereof was demanded. As standard practices of their respective hospital might influence health care professionals' responses, no more than three health care professionals per medical specialty in each center were allowed to participate. Similarly, patient representatives associated with ERNICA through ERNICA's European Patient Advocacy Group (ePAG) were contacted and asked to invite their peer patient representatives active in EA care. Patient representatives were considered experts if they had experience in at least two cases or care processes for EA patients. In addition to direct correspondence, social media was used to target potential participants and disseminate the brief study information and preregistration link. An official invitation to participate in the Delphi, including a more detailed description of the process, was only sent out to those who indicated their willingness to participate through preregistering.

The three-round Delphi questionnaires

Round one

All identified quality indicators were categorized in structure (regarding the structure of the health system), process (regarding care processes), and outcome (regarding outcomes of provided care), according to the Donabedian model[15]. These three categories of quality indicators represent three interacting but substantially different aspects of quality of care. We recorded definitions, the numerator, denominator, time of measurement, and mode of retrieval (literature or guidelines or both) for every quality indicator in the corresponding "help text." Text boxes were added for participants' comments on the importance, interpretation, and wording of separate indicators. The principal question participants were asked was: "*Should this quality indicator be used to evaluate and compare EA-care between hospitals, regions, or countries*"?, rated on a Likert scale from 1 through 9 ("*Totally disagree"* to "*Totally agree*"). Previously identified patient (baseline) characteristics and treatment characteristics were rated similarly, based on the question: "*Should this characteristic be registered to enable the interpretation of quality indicators in EA-care / Do you think we need to register this variable to allow for correction for case-mix in the future?*". Finally, in round one, participants were invited to propose additional quality indicators if they considered specific topics or items missing. These proposed indicators were then added to round two if deemed distinct from those already in round one.

Rounds two and three

In the next two rounds, all participants who finished the previous round were asked to rate the items once more after receiving a summary of the previous round's results. This summary included their own response, the voting results of their own stakeholder group – including the median score, score distribution, and comments of their peers - as well as

the voting results of the other stakeholder group. Finally, a ranking section was included in rounds two and three. In this section, participants were asked to prioritize the previously selected indicators by choosing the five to ten most important ones to implement in the clinical audit. The other stakeholder groups' ranking of indicators in the previous round determined in which order the indicators would be presented in the next round. Where participants chose more than 50% of possible indicators, rather than the 5-10 requested, these responses were excluded to prevent an overestimation of the deemed importance of each item.

Selection criteria

An indicator or characteristic was selected (i.e., voted 'consensus in') if it received a median score of ≥8, with ≥75% of the ratings being in the highest tertile. Conversely, an indicator or characteristic was discarded (i.e., voted 'consensus out') if the median score was ≤2, with ≥75% of the ratings being in the lowest tertile. If voted 'consensus in' or 'consensus out,' the characteristic or indicator was not included in the following rounds.

Finalization of the core indicator set

It was decided that if consensus were not achieved after three Delphi rounds, a consensus meeting would be organized to decide on the final core indicator set and to converge the opinions of the two stakeholder groups.

Data collection, management, and confidentiality

Data was collected using Welphi, an online Delphi questionnaire management software[16]. This software program distributes the questionnaires via e-mail and automatically calculates and returns previous scores of the participants and their own stakeholder

group. Scores of the other stakeholder group were manually calculated and added to the Delphi before the following round. Response time for each round was at least four weeks to allow as many responses as possible. Participants were regularly reminded to complete the questionnaire if a response was not received after some time, with the last reminder being a personal e-mail. The aim was to achieve a response rate of >70% for the three questionnaires[17]. At the start of the first round, all participants provided electronic informed consent.

Data analysis and publication

Partially completed questionnaires were excluded, but the responses of these participants in earlier rounds were not. We examined whether attrition affected the scoring graphically and compared median scores per round and item between participants who did and did not complete the respective Delphi round(s). Data were processed using Excel and analyzed using SPSS (version 25, IBM Corp. Armonk, NY, USA). Findings were reported to comply with the practical guidance provided by Boulkedid in his systematic review[8].

RESULTS

Item list

Over 700 studied patient- and care-process characteristics and outcomes were extracted from 209 EA publications in the systematic review of recent literature on the primary EA care process[11]. The most frequently described parameters (in more than five percent of included publications) were subsequently translated into process- and outcome indicators. Similarly, recommendations from the widely recognized ESPHGAN and ERNICA consensus guidelines were translated into possible indicators[12–14]. Combining the results of both efforts generated a long list of 33 patient characteristics and 142 quality

indicators, of which 2 were structure indicators, 59 were process indicators, and 81 were outcome indicators. The list was further categorized according to clinical topics, such as primary surgery, stricture, and feeding, to increase the readability of the questionnaire. The complete item list, including the origin of each item, is included in the supplementary materials (S1 through S4).

Participants

One hundred and ten health care professionals pre-registered and were thus invited to participate. Seventy-one health care professionals completed all three questionnaires, representing 41 hospitals in 12 European countries. The median number of years of experience of these health care professionals was 14 years (IQR 9; 22). For the patient representatives panel, 23 registered to participate, of which 14 representing eight countries completed all three questionnaires. These patient representatives had a median of 7.5 years of experience in that role (IQR 3, 15). 13 out of 14 patient representatives were attached to their respective national EA patient organizations, which are all recognized, full-member organizations of the Esophageal Atresia Global Support Groups (EAT). Three of the patient representatives additionally underscored their attachment to this global organization. A panel description of both stakeholder groups is shown in Table 1 and Table 2.

The three-round Delphi questionnaires

Between January and August 2022, three rounds of questionnaires were distributed amongst preregistered health care professionals and patient representatives, which were available to complete for four, nine, and nine weeks, respectively. The duration depended on the response rate and the number of added reactions after each reminder. After round

one, twenty-seven participants proposed twenty-eight additional indicators and thirteen additional characteristics, all of which were decided to be distinct and therefore included in the subsequent rounds. Following the predetermined selection criteria, no items were voted 'consensus out' in all three rounds. However, after round two, it was decided that to increase feasibility, items that were not voted 'consensus in' or 'consensus out' in two subsequent rounds would be discarded, except for those items voted 'consensus in' by one of both stakeholder groups. All scores, per round, per stakeholder group, are outlined in the supplementary materials (S1 through S4). Thirty-two indicators and twenty-two patient- and treatment characteristics were voted 'consensus in' by both stakeholder groups. Eleven indicators and two characteristics were voted 'consensus in' by the patient representatives but not by the health care professionals. Conversely, three characteristics were voted 'consensus in' by the health care professionals but not by the patient representatives. An overview of the entire study process, including response rate, item selection and item addition, is presented in Figure 1.

Subgroup analysis of health care professionals

Just over 50% (n=38) of the participating health care professionals were surgeons, while the number of participants in other specialties ranged from 1 to 8, depending on the specific profession (Table 2). When comparing individual item scores in both subgroups, some differed significantly, speaking to the importance of having both subgroups in the professionals' panel. However, overall mean and median scores of surgeons versus scores of other health care professionals did not differ significantly. An overview of the subgroup analysis results is included in the supplementary materials (S5).

Attrition analysis

No differences were found in scoring distributions when comparing the scores of participants who only completed round one with those who completed rounds one and two. Similarly, no differences were found when comparing scoring distributions between those who only completed round one and those who completed all three rounds; or those who completed rounds one and two, compared to those who completed rounds one, two, and three. The mean overall score of round one did not differ significantly between participants who only completed round one and those who completed at least rounds one and two: 7.20 (sd 0.88) versus 7.15 (0.92) (p=0.84). For participants who completed all three rounds, the average score for round one was 7.16 (0.93) (p=0.97). Median individual item scores between complete responders and incomplete responders were comparable. An overview of attrition analysis results is included in the supplementary materials (S6).

Indicator prioritization, consensus meeting, and finalizing the core indicator set

No other indicators were voted 'consensus in' after round 3. During rounds two and three, participants were asked to select the five to ten most essential indicators out of those voted 'consensus in' in the earlier round(s). By accumulating the number of votes each indicator received, four rankings were established: one per stakeholder group per round. To prevent overestimating the value attributed to an indicator, we excluded participants' answers of those choosing "Yes" for over 50% of all indicators voted 'consensus in' instead of selecting five to ten as was the original task. Both stakeholder group rankings established in rounds two and three are included in the supplementary materials (S7 and S8). Comparing the final rankings, ten of the eleven most frequently voted-on indicators proved similar in both stakeholder groups, albeit in a somewhat different order. Two indicators were dissimilar. Health care professionals prioritized intra-operative

complications, ranking it eighth, while patient representatives ranked this indicator fourteenth. Patient representatives prioritized gastroesophageal reflux, ranking it ninth, whereas health care professionals ranked it sixteenth. It was decided that organizing a consensus meeting to resolve this minor discrepancy was unnecessary. Instead, we sent all participants an e-mail asking: "Do you agree that this quality indicator will be included in the final core indicator set, as the other stakeholder group ranked it as one of the twelve most important quality indicators in EA care?". This question was answered by 79% of participating health care professionals and 93% of the patient representatives. Respectively, 87.5% and 100% of those voted "Yes, we agree," thus supporting the inclusion of these two 'extra' indicators in the final core indicator set. The rating, preliminary proposed definition, timing of measurement and the origin of the baseline characteristics and quality indicators of the final core indicator set are listed respectively in Table 3 and Table 4. A concise, summarized overview of the core indicator set is presented in Table 5.

DISCUSSION

This study established a core indicator set for EA care through a Delphi consensus method, incorporating the opinions of experienced health care professionals of multiple disciplines and patient representatives across Europe. The final core indicator set, consisting of one structure indicator, three process indicators, eight outcome indicators, and twenty-two patient- and treatment characteristics, will be implemented in the ERNICA European Pediatric Surgical Audit (EPSA). This audit is a quality monitoring tool to measure and evaluate variation in EA care and its outcomes, enabling comparison of these findings between participating medical centers, regions, or countries. Insight into

variations in EA care and EA outcomes, measured through the quality indicators selected in this study, should allow for recognizing best practices and determining improvement actions to increase the quality of EA care.

Of 46 patient and treatment characteristics and 142 quality indicators included in the first Delphi round, 22 and 32 were voted 'consensus in' by both stakeholder groups. No characteristics or indicators were voted 'consensus out,' and overall median scores were high, with the lowest median score attributed to a characteristic or indicator being "Mildly disagree (4)". The participants thus found all items on the longlist to reflect some aspect of quality of care. Within the health care professionals' panel, scores of surgeons were relatively similar to those of professionals with other disciplines. Health care professionals were generally more critical of the proposed quality indicators than patient representatives, with fewer indicators voted 'consensus in' by this stakeholder group and generally lower median scoring for the rated items. Moreover, throughout the Delphi rounds and prioritizing segments, the scoring of both stakeholder groups displayed a convergence of opinions. This susceptibility to move towards each other resulted in a robust core indicator set, representing consensus within and between stakeholder groups. The remaining differences in the prioritization of quality indicators between and within the stakeholder panels emphasize the value of including health care professionals of various specialties, as well as patient representatives in this study. Furthermore, these findings might help to focus efforts on teaching other stakeholders involved and establish and promote mutual understanding of aspects of the EA care process deemed important by these different stakeholders.

The final quality indicator set is coherent and well-rounded. All four prioritized structure- and process indicators were identified in existing guidelines[12,13]. For the consensus statements on which these indicators are based, the underlying level of evidence was lower than CEBM level 1. The other prioritized quality indicators measure outcomes such as mortality and anastomotic leakage and are consistent with EA research's most frequently studied parameters[11]. This coherence speaks to a close relationship between the focus of EA research and (quality of) clinical care.

This European core indicator set also largely corresponds with the EA-specific quality indicators developed in 2015 by pediatric surgeons of all six Dutch pediatric surgical centers. In national consensus meetings, these ten experts agreed upon quality indicators for several neonatal surgical conditions for the Dutch clinical audit for pediatric surgery, including four for esophageal atresia care[6]. As a sound methodological foundation and international support are paramount to ensure the generalizability and acceptance of the quality indicator results in benchmarking, we decided to revise these indicators. Ultimately, three of these four previous 'Dutch' EA quality indicators were selected and prioritized in this study. One might theorize that consensus meetings with comparatively representative participants might be preferable if the applied Delphi method, which is more time-consuming, leads to similar results. However, the core indicator set identified in this study is more extensive, and digital Delphi questionnaires allow for the involvement of a more geographically diverse panel.

Although clinical auditing in conditions such as esophageal atresia is unusual, it is more common in conditions with a higher prevalence, such as esophageal carcinoma, for which several different quality indicator sets are available[18,19]. Compared to these more

established quality indicators, most of the indicators in our final set appear to be not specific for EA care but more so for esophageal surgery and surgery itself, for example, patient volume, multidisciplinary treatment, complicated postoperative course, certain complications (i.e., anastomotic leakage), and mortality or survival. More disease-specific quality indicators in the final set, for example, on the transition from pediatric to adult care, reflect the value and urgency currently attributed to specific topics in EA care, as evidenced by health care professionals and patient representatives alike.

Providing care for rare conditions such as esophageal atresia can be challenging. Due to the low patient numbers and limited funding, prospective, trial-based research is difficult, resulting in little to no CEBM level 1 evidence. Guidelines and treatment standards are thus primarily based on expert consensus instead of evidence-based outcomes. Consequently, there is great potential to improve (the quality of) care for rare conditions, which may be even more significant than in other diseases, where standardization of care and research has already progressed much further. Developing a clinical audit enables recognizing and analyzing variations in treatment and corresponding outcomes and identifying best practices[3]. It may also guide science-based standardization efforts and generate research hypotheses to focus future EA research[20]. Implementing the core indicator set established in this study in such a clinical audit will be the first step in confirming this intended utilization and the applicability of clinical auditing as a quality monitoring instrument in esophageal atresia and rare conditions in general.

Standardization is also the focus of the OCELOT ("Oesophageal atresia CorE outcomes Long Term") Delphi study currently being conducted to generate a core outcome set for EA[21]. A core outcome set aims to provide "*an agreed standardized set of outcomes that*

should be measured and reported, as a minimum, in all clinical trials in specific areas of health or health care"[22]. Hence it focuses on outcomes and variables used for interpreting the results of trials and other types of studies rather than on structure or process indicators. There may well be an overlap between the current core indicator set and the finalized OCELOT core outcome set, which will provide additional reassurance of the importance placed on these outcomes. Furthermore, the complementary measurement of these core outcomes in the clinical audit will benefit EA knowledge by increasing the collected data's interpretability, usability, and comparability. Several study team members of both projects were involved in both study steering groups to maximize these synergetic opportunities, recognizing that collaboration is paramount to establishing a good quality of care.

To our knowledge, this is the first time that quality indicators were identified, defined, and selected for EA care using a European international panel comprising experienced patient representatives and health care professionals of several disciplines. EA patients throughout Europe are well-organized in patient support organizations, nationally and globally (e.g., Esophageal Atresia Global Support Groups, EAT). They are also involved in the ERNICA network. We consider their involvement in designing the study and participation as a panelist as a major strength of this study. However, for other rare conditions, patient organizations are not always that active, which might challenge the future development of core indicator sets in these rare conditions.

Even though much effort was made to include representative panels for both stakeholder groups, there might be some limitations to the generalizability of the core indicator set. Firstly, despite the multidisciplinary character of the health care professionals' panel, most

participants were pediatric surgeons. Comparatively, the number of professionals in other disciplines was relatively low (i.e., general pediatricians, pediatric gastroenterologists). This imbalance could result in overestimating the value attributed to quality indicators regarding (peri)surgical treatment. A more equal distribution of health care professionals across disciplines might have resulted in a different outcome, although the absolute differences in aggregated subgroup analysis seem limited. Secondly, health care professionals or patient representatives of other continents were excluded from participation to maintain applicability to European care, reducing the generalizability of the final core indicator set to health care systems in other parts of the world. Moreover, most participants were of Western European origin, and 43% of the participating health care professionals practices medicine in the Netherlands or the United Kingdom, potentially overestimating country-specific values concerning the quality of care and limiting applicability to Eastern-European countries. Thirdly, the size of both stakeholder groups was different (71 health care professionals versus 14 patient representatives), causing the vote of one patient representative to weigh relatively more than the vote of one health care professional. This imbalance is partially neutralized by including patient representatives rather than patients themselves. They consider and represent multiple patient cases, legitimizing an equal collective vote. The difference in panel size was further accounted for by the separate analyses of the results of both stakeholder groups and only feeding back collective scores. Lastly, in both stakeholder groups, no attrition was found.

Ultimately, twelve quality indicators were prioritized, yet more indicators were voted 'consensus in' by one or both stakeholder groups. These indicators, such as (start of)

feeding or dysphagia, did not receive enough votes to be included in the final set; however, the vote for 'consensus in' does demonstrate the value ascribed to them. This knowledge could provide additional focus for the improvement of care and research.

The core indicator set agreed upon in this study will now be implemented in the EPSA/ERNICA registry for EA. Part of this process will be the assessment of the validity and feasibility of selected quality indicators. Because of the limited availability of comparative research or large patient numbers in esophageal atresia, it is of the utmost importance to establish that the selected indicators are indeed associated with outcome and (quality) of care. If this association can indeed be discerned, it will enable the recognition of best practices. In that case, we can move on to the next step of the quality improvement cycle: initiating concrete improvement initiatives and improving (the quality of European) EA care.

CONCLUSION

In this study, through a Delphi consensus method, we established a core indicator set comprising one structure indicator, three process indicators, eight outcome indicators, and twenty-two patient and treatment characteristics prioritized by health care professionals and patient representatives. The selected characteristics and quality indicators uniquely reflect the urgency and importance attributed to the pertaining topics by both stakeholder groups, which may help focus future research efforts. Moreover, this set will be implemented in the EPSA/ERNICA registry, a clinical audit to evaluate and compare (the quality of) EA care in Europe. After confirming the validity and feasibility of these prioritized quality indicators, their measurement will, it is expected, enable the

recognition of best practices, thereby allowing us to start concrete improvement initiatives for EA care throughout Europe.

ACKNOWLEDGEMENTS

We thank Anke Widenmann-Grolig and Melanie Wildermuth for their involvement and support in recruiting the patient representatives' panel and for their input on the comprehensibility of the Delphi questionnaires. Additionally, we thank the EA expert team for their involvement in translating outcomes to quality indicators: Prof. J. Thorup, Prof. F. Morini, Dr. J. Suominen, and Anke Widenmann-Grolig, in addition to the experts in the author group.

REFERENCES

1. Nassar N, Leoncini E, Amar E, et al. Prevalence of esophageal atresia among 18 international birth defects surveillance programs. *Birth Defects Res A Clin Mol Teratol*. 2012;94(11). doi:10.1002/bdra.23067

2. Bell JC, Baynam G, Bergman JEH, et al. Survival of infants born with esophageal atresia among 24 international birth defects surveillance programs. *Birth Defects Res*. 2021;113(12):945-957. doi:10.1002/bdr2.1891

3. Beck N, van Bommel AC, Eddes EH, van Leersum NJ, Tollenaar RA, Wouters MW. The Dutch Institute for Clinical Auditing. *Ann Surg*. 2020;271(4):627-631. doi:10.1097/SLA.0000000000003665

4. Voeten SC, Krijnen P, Voeten DM, Hegeman JH, Wouters MWJM, Schipper IB. Quality indicators for hip fracture care, a systematic review. *Osteoporosis International*. 2018;29(9):1963-1985. doi:10.1007/s00198-018-4558-x

5. Courrech Staal EFW, Wouters MWJM, Boot H, Tollenaar RAEM, van Sandick JW. Quality-of-care indicators for oesophageal cancer surgery: A review. *European Journal of Surgical Oncology*. 2010;36(11):1035-1043. doi:10.1016/j.ejso.2010.08.131

6. van Heurn E, de Blaauw I, Heij H, et al. Quality Measurement in Neonatal Surgical Disorders: Development of Clinical Indicators. *Eur J Pediatr Surg*. 2015;25(6):526-531. doi:10.1055/s-0034-1396416

7. Sinha IP, Smyth RL, Williamson PR. Using the Delphi Technique to Determine Which Outcomes to Measure in Clinical Trials: Recommendations for the Future

Based on a Systematic Review of Existing Studies. *PLoS Med*. 2011;8(1):e1000393. doi:10.1371/journal.pmed.1000393

8. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. Wright JM, ed. *PLoS One*. 2011;6(6):e20476. doi:10.1371/journal.pone.0020476

9. Dalkey N. An experimental study of group opinion. *Futures*. 1969;1(5):408-426. doi:10.1016/S0016-3287(69)80025-X

10. Hsu CC, Sanddford BA. The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research, and Evaluation*. 2007;12:1-8. doi:doi.org/10.7275/pdz9-th90

11. Teunissen N, Brendel J, Eaton S, et al. Variability in the Reporting of Baseline Characteristics, Treatment, and Outcomes in Esophageal Atresia Publications: A Systematic Review. *European Journal of Pediatric Surgery*. 2023;33(02):129-137. doi:10.1055/s-0042-1758828

12. Krishnan U, Mousa H, Dall'Oglio L, et al. ESPGHAN-NASPGHAN Guidelines for the Evaluation and Treatment of Gastrointestinal and Nutritional Complications in Children With Esophageal Atresia-Tracheoesophageal Fistula. *J Pediatr Gastroenterol Nutr*. 2016;63(5):550-570. doi:10.1097/MPG.0000000000001401

13. Dingemann C, Eaton S, Aksnes G, et al. ERNICA Consensus Conference on the Management of Patients with Esophageal Atresia and Tracheoesophageal Fistula: Diagnostics, Preoperative, Operative, and Postoperative Management. *European Journal of Pediatric Surgery*. 2020;30(04):326-336. doi:10.1055/s-0039-1693116

14. Dingemann C, Eaton S, Aksnes G, et al. ERNICA Consensus Conference on the Management of Patients with Esophageal Atresia and Tracheoesophageal Fistula: Follow-up and Framework. *European Journal of Pediatric Surgery*. 2020;30(06):475-482. doi:10.1055/s-0039-3400284

15. Donabedian A. Evaluating the Quality of Medical Care. *Milbank Quarterly*. 2005;83(4):691-729. doi:10.1111/j.1468-0009.2005.00397.x

16. Welphi. Welphi Online Survey Platform. Accessed October 25, 2022. www.welphi.com

17. Iqbal S, Pipon-Young L. The Delphi Method. *Psychologist*. 2009;22(7):598-601.

18. Busweiler LAD, Wijnhoven BPL, van Berge Henegouwen MI, et al. Early outcomes from the Dutch Upper Gastrointestinal Cancer Audit. *British Journal of Surgery*. 2016;103(13):1855-1863. doi:10.1002/bjs.10303

19. Stordeur S, Vlayen J, Vrijens F, et al. Quality indicators for oesophageal and gastric cancer: a population-based study in Belgium, 2004-2008. *Eur J Cancer Care (Engl)*. 2015;24(3):376-386. doi:10.1111/ecc.12279

20. Posada de la Paz M, Groft SC. *Rare Diseases Epidemiology*. Vol 686. (Posada de la Paz M, Groft SC, eds.). Springer Netherlands; 2010. doi:10.1007/978-90-481-9485-8

21. Thursfield RM, Gorst S, Adams A, Losty P, Hopwood L, Slater G. Establishing Core Outcomes to Improve Quality Health Care For Patients Born With Oesophageal Atresia (OA) And Tracheoesophageal Fistula (TOFS) [Submitted].

22. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials*. 2017;18(S3):280. doi:10.1186/s13063-017-1978-4

LEGENDS: FIGURES AND TABLES

Figure 1: Study flow chart

Table 1: Panel description: patient representatives who completed all three rounds

Table 2: Panel description: health care professionals who completed all three rounds

Table 3: Characteristics included in the final core indicator set

Table 4: Quality indicators included in the final core indicator set

Table 5: Summarized overview of the established core indicator set for esophageal atresia care

Figure 1. Study flow chart

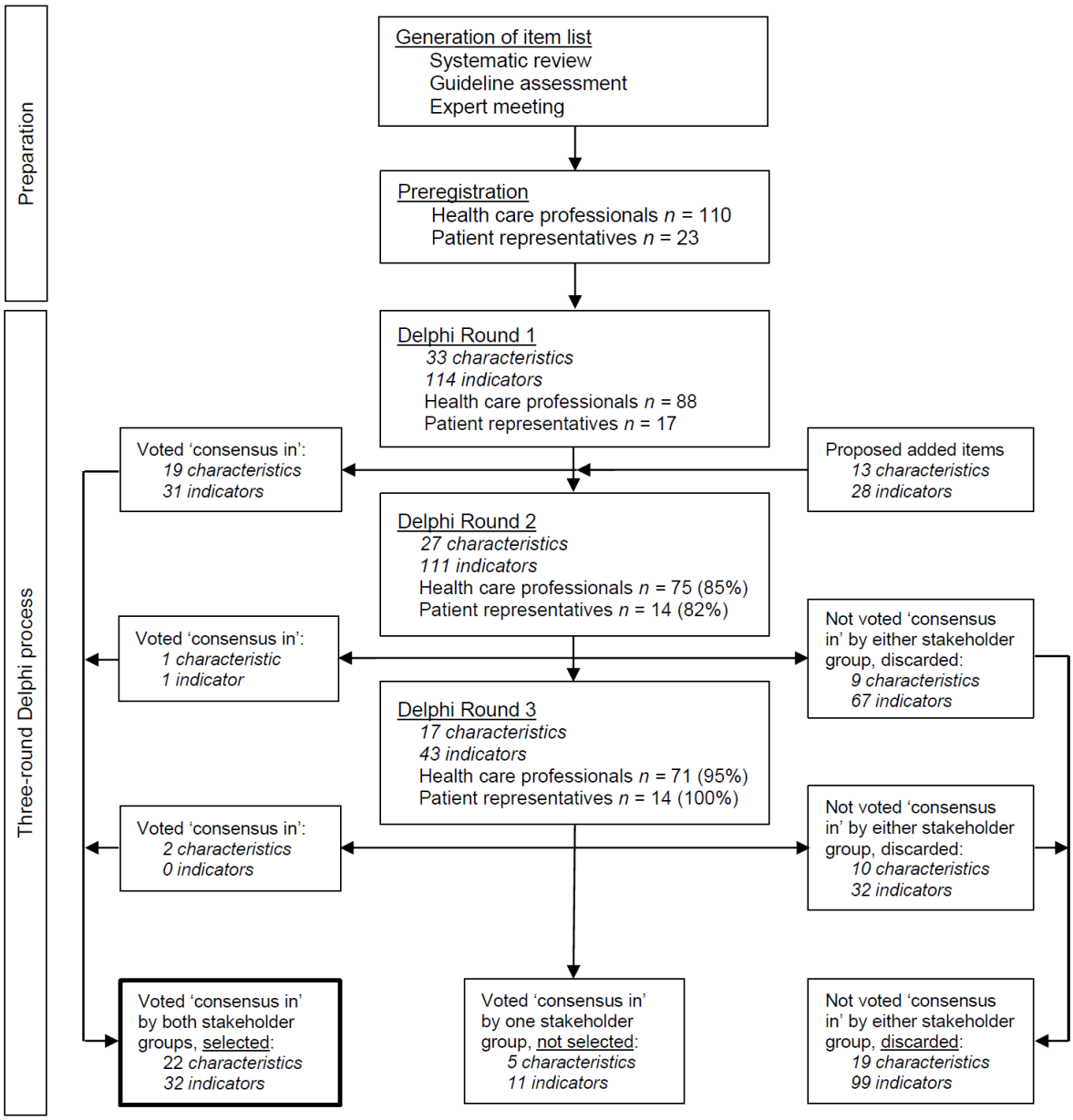

Table 1. Panel description: patient representatives who completed all three rounds

		Number of patient representatives (n=14)
Years of experience (median, IQR)		7.5 (3, 15)
Attached to a patient organization		13 (93%)
Country	Austria	1 (7%)
	Belgium	2 (14%)
	Croatia	1 (7%)
	France	1 (7%)
	Germany	2 (14%)
	Spain	1 (7%)
	The Netherlands	1 (7%)
	UK	5 (36%)

Due to rounding errors, some categories may not add up to 100%

Table 2. Panel description: health care professionals who completed all three rounds

		Health care professionals (n=71)	Origin of health care professionals
Years of experience (median, IQR)		14 (9, 22)	
Profession	Pediatric surgeon	38 (54%)	AT/BE/DK/FI/FR/DE/ IT/PL/ES/NL/UK
	Speech- and language therapist	8 (11%)	UK/NL/DE
	Pediatric gastroenterologist	4 (6%)	BE/FR/IT
	Pediatric pulmonologist	4 (6%)	FR/ES/NL/UK
	Dietician	2 (3%)	NL/UK
	Gastroenterologist (adult care)	2 (3%)	NL
	Pediatrician	2 (3%)	HR/FR
	Neonatologist	2 (3%)	PL/HR
	Nurse practitioner/nurse specialist	2 (3%)	NL/UK
	Care- and quality manager	1 (1%)	DE
	Medical rehabilitation specialist	1 (1%)	ES
	Pediatric anesthesiologist	1 (1%)	NL
	Pediatric intensive care doctor	1 (1%)	NL
	Pediatric nutritionist	1 (1%)	ES
	Physical therapist	1 (1%)	IT
	Pulmonologist (adult care)	1 (1%)	NL
Country	Austria	1 (1%)	
	Belgium	2 (3%)	
	Croatia	2 (3%)	
	Denmark	2 (3%)	
	Finland	3 (4%)	
	France	6 (8%)	
	Germany	7(10%)	
	Italy	8 (11%)	
	Poland	2 (3%)	
	Spain	9 (13%)	
	The Netherlands	15 (21%)	
	UK	14 (20%)	

Due to rounding errors, some categories may not add up to 100%

Table 3. Characteristics included in the final core indicator set

Characteristic	Definition and origin
Gender	1. *Definition:* Male / Female / Undetermined / Fetus (unknown) 2. *Identified in* 85% of published papers between 2015 and 2021
Type of EA	1. *Definition*: Gross Classification 2. *Identified in* 74% of published papers between 2015 and 2021
Long gap/gap length	1. *Definition*: all type A and type B according to the Gross classification 2. *Identified in* 46% of published papers between 2015 and 2021
Gestational age	1. *Definition:* in weeks and days 2. *Identified in* 74% of published papers between 2015 and 2021
Weight at birth	1. *Definition*: in grams 2. *Identified in* 69% of published papers between 2015 and 2021
Age at surgery	1. *Definition*: in days 2. *Identified in* 47% of published papers between 2015 and 2021
Any other (associated) congenital malformation	1. *Definition*: is there (at least one) another associated anomaly: yes or no 2. *Identified in* 40% of published papers between 2015 and 2021
Chromosomal/genetic abnormalities	1. *Definition*: is there a chromosomal or genetic anomaly: yes or no (e.g., trisomy 18 or 21) 2. *Identified in* 39% of published papers between 2015 and 2021
VACTERL	1. *Definition*: at least three or more characteristic abnormalities 2. *Identified in* 42% of published papers between 2015 and 2021
Anorectal malformation	1. *Definition*: is there an anorectal malformation: yes or no (e.g., anal atresia, rectourethral fistula) 2. *Identified in* 32% of published papers between 2015 and 2021
Cardiac malformation/congenital heart disease	1. *Definition:* is there a cardiac malformation: yes or no (e.g., hemodynamically significant ASD, VSD, PDA, Fallot) 2. *Identified in* 72% of published papers between 2015 and 2021
Renal/genitourinary anomalies	1. *Definition*: is there a renal or genitourinary anomaly: yes or no (e.g., renal agenesis, polycystic kidneys or duplicated collected system, hypospadias) 2. *Identified in* 36% of published papers between 2015 and 2021
Musculoskeletal/limb anomalies	1. *Definition*: is there a musculoskeletal malformation or malformation of the extremities: yes or no (e.g., hemivertebrae, costal fusion, limb reduction defect, clubfoot) 2. *Identified in* 35% of published papers between 2015 and 2021
Intestinal malformation	1. *Definition*: is there an intestinal malformation: yes or no (e.g., duodenal atresia, malrotation) 2. *Identified in* 27% of published papers between 2015 and 2021
Pulmonary/respiratory anomalies or conditions	1. *Definition*: is there a pulmonary or respiratory condition or malformation: yes or no (e.g., NRDS, pulmonary agenesis) 2. *Identified in* 17% of published papers between 2015 and 2021
Neurologic/central nervous system anomalies	1. *Definition*: is there a neurological malformation or condition, or a condition of the central nervous system: yes or no (e.g., microcephaly, hydrocephaly) 2. *Identified in* 15% of published papers between 2015 and 2021
Laryngeal anomalies	1. *Definition*: is there a laryngeal or laryngotracheal malformation: yes or no (e.g., laryngeal cleft, subglottic stenosis, laryngotracheoesophageal cleft) 2. *Identified in* 6% of published papers between 2015 and 2021
Previous esophageal surgery	1. *Definition*: did the patient previously undergo esophageal surgery, for example, a primary anastomosis: yes or no 2. *Identified in* 13% of published papers between 2015 and 2021
Elongation procedure: yes or no	1. *Definition*: No elongation or traction Elongation (myotomy) Traction (Foker, Kimura) Other (magnetic) 2. *Identified in* 9% of published papers between 2015 and 2021
Type of repair	1. *Definition*: Primary anastomosis Delayed primary anastomosis

	Secondary anastomosis and type Gastric pull-up Gastric sleeve Jejunal interposition Colonic interposition Other 2. *Identified in* 60% of published papers between 2015 and 2021
Thoracotomy or thoracoscopy	1. *Definition*: thoracotomy (open surgery) or thoracoscopy (minimally invasive surgery) 2. *Identified in* 38% of published papers between 2015 and 2021
Existence Multidisciplinary Team Clinic	1. *Definition*: Does the hospital have a multidisciplinary clinic specialized in EA care? 2. *Identified in* round one (proposed by health care professional)

Table 4. Quality indicators included in the final core indicator set

Health care professionals prioritizing indicator	Patient representatives prioritizing indicator	Indicator	Proposed definition and timing of measurement
49 (72%)	12 (92%)	Number of patients that underwent surgical treatment for esophageal atresia in that hospital	1. *Definition*: 2. *Timing of measurement*: 1 year 3. *Identified in* ERNICA guidelines
31 (46%)	7 (54%)	Percentage of patients that underwent surgical treatment for esophageal atresia that underwent tracheoscopy or bronchoscopy during or before surgery	1. *Definition*: 2. *Identified as a studied parameter in* 11% of published papers between 2015 and 2021; ERNICA consensus conference
52 (76%)	11 (85%)	Percentage of patients that underwent surgical treatment for esophageal atresia that was treated by a multidisciplinary team	1. *Definition*: a multidisciplinary team should at least include the following specialties: surgery, gastroenterology, pulmonology, otolaryngology, and nutrition counseling 2. *Timing of measurement*: 2 years 3. *Identified in* ESPGHAN guidelines
28 (41%)	10 (77%)	Percentage of patients that underwent surgical treatment for esophageal atresia and as an adult was transitioned from pediatric care to an adult physician with expertise in EA	1. Definition: adult physicians such as general practitioners, surgeons, gastroenterologists, pulmonologists, or any informed specialist aware of the specificities of the care of adults operated for EA 2. Timing of measurement: 20 years 3. *Identified in* ERNICA consensus conference and ESPGHAN guidelines
30 (44%)	3 (23%)	Percentage of patients that underwent surgical treatment for esophageal atresia, in which there were intraoperative complications*	1. *Definition*: intraoperative complications include a.o. accidental extubation, tube dislodgement; iatrogenic injury to the phrenic nerve, trachea, bronchus, or lung; hemodynamic complications such as arrhythmia, tachycardia, bradycardia, bronchospasm, desaturations; 2. *Timing of measurement*: 30 days 3. *Identified as a studied parameter in* 5% of published papers between 2015 and 2021
37 (54%)	11 (85%)	Percentage of patients that underwent surgical treatment for esophageal atresia and developed an anastomotic leakage	1. *Definition*: clinically significant if symptomatic with objective leakage of contrast or a pneumothorax on imaging 2. *Timing of measurement*: 30 days 3. *Identified as a studied parameter in* 69% of published papers between 2015 and 2021
35 (51%)	6 (46%)	Percentage of patients that underwent surgical treatment for esophageal atresia in which reoperation was required because of an esophageal complication	1. *Definition*: reoperation on the esophagus (e.g., resection of stricture, persistent leakage not responding to conservative treatment, and recurrent fistula) 2. *Timing of measurement*: 1 year 3. *Identified as a studied parameter in* 38% of published papers between 2015 and 2021
38 (56%)	7 (54%)	Percentage of patients that underwent surgical treatment for esophageal atresia that subsequently developed a recurrent fistula	1. *Definition*: clinically significant if symptomatic and diagnosed by bronchoscopy/endoscopy 2. *Timing of measurement*: 1 year 3. *Identified as a studied parameter in* 41% of published papers between 2015 and 2021

17 (25%)	6 (46%)	Percentage of patients that developed clinically significant gastroesophageal reflux*	1. *Definition*: clinically significant if the presence of typical symptoms (frequent postprandial vomiting, retrosternal pain/heartburn, poor weight gain), with pH monitoring showing pathologic reflux OR upper endoscopy showing typical reflux-induced mucosal lesions OR for which fundoplication surgery was needed 2. *Timing of measurement*: 1 year 3. *Identified as a studied parameter in* 53% of published papers between 2015 and 2021
38 (56%)	9 (69%)	The number of dilatations the patient underwent since undergoing surgical treatment for esophageal atresia	1. *Definition*: median 2. *Timing of measurement*: 1 year 3. *Identified as a studied parameter in* 36% of published papers between 2015 and 2021
28 (41%)	6 (46%)	Percentage of patients that underwent surgical treatment for esophageal atresia, in which there were postoperative complications within 1 year	1. *Definition*: postoperative complications such as wound infections, anastomotic complications, vocal cord issues, cosmetic complications, etcetera. 2. *Timing of measurement*: 1 year 3. *Identified as a studied parameter in* 7% of published papers between 2015 and 2021
29 (43%)	8 (62%)	Percentage of patients that underwent surgical treatment for esophageal atresia that died within one year from surgery	1. *Definition*: 2. *Timing of measurement*: 1 year 3. *Identified as a studied parameter in* 66% of published papers between 2015 and 2021

* This indicator was only prioritized (top 10) by one of both stakeholder groups but was accepted by the other stakeholder group in a final vote.

Table 5. Summarized overview of the established core indicator set for esophageal atresia care

Baseline characteristics		*Quality indicators*
Sex	One or more associated anomaly	*Number of patients* that underwent surgical treatment for esophageal atresia in that hospital
Type of EA	Chromosomal/genetic anomalies	Percentage of patients that underwent *tracheoscopy or bronchoscopy during or before surgery*
Long gap	VACTERL	Percentage of patients in which there were *intraoperative complications**
Gestational age	Cardiac malformation/CHD	Percentage of patients that developed an *anastomotic leakage*
Weight at birth	ARM	The *number of dilatations* the patient underwent since undergoing surgical treatment for esophageal atresia
Age at operation	Intestinal malformation	Percentage of patients that subsequently developed a *recurrent fistula*
Type of repair	Pulmonary/respiratory conditions	Percentage of patients that developed clinically significant *gastroesophageal reflux**
Open or scopic procedure	Laryngeal anomalies	Percentage of patients in which *reoperation* was required because of an esophageal complication
Lengthening procedure	Musculoskeletal/limb anomalies	Percentage of patients in which there were *postoperative complications* within one year
Previous esophageal surgery	Neurologic or CNS anomalies	Percentage of patients that *died* within one year from surgery
Existence multidisciplinary team clinic	Renal/genitourinary anomalies	Percentage of patients that was treated by a *multidisciplinary team*
		Percentage of patients and as an adult was *transitioned* from pediatric care to an adult physician with expertise in EA

*This indicator was only prioritized (top 10) by one of both stakeholder groups but was accepted by the other stakeholder group in a final vote.