ORIGINAL ARTICLE

A benchmarking project on the quality of previous guidelines about the management of malignant pleural effusion from the European Society of Thoracic Surgeons (ESTS) Pleural Diseases Working Group

Luca Bertolaccini[a,*], Benedetta Bedetti[b], Alessandro Brunelli[c], Katerina Marinova[d], Federico Raveglia[e],
Gaetano Rocco[f], Yaron Shargall[g], Piergiorgio Solli[a,h], Gonzalo Varela[i], Kostas Papagiannopoulos[c],
Jaroslaw Kuzdzal[j], Gilbert Massard[k], Enrico Ruffini[l], Pierre-Emmanuel Falcoz[k], Carlos Martinez-Barenys[m],
Isabelle Opitz[n], Hasan F. Batirel[o], Alper Toker[p] and Marco Scarci[b]; for the Members
of the ESTS Pleural Disease Working Group

[a] Department of Thoracic Surgery, AUSL Romagna Teaching Hospital, Ravenna, Italy
[b] Thoracic Surgery Unit, University College London Hospital, London, UK
[c] Department of Thoracic Surgery, St. James's University Hospital, Leeds, UK
[d] Department of Surgery, ESS Thoracic Surgery, Thoracic Surgery Clinic, University Hospital St. Marina, Varna, Bulgaria
[e] Thoracic Surgery Unit, Azienda Ospedaliera San Paolo, Milan, Italy
[f] Division of Thoracic Surgery, Department of Thoracic Surgical and Medical Oncology, National Cancer Institute—Pascale Foundation, Naples, Italy
[g] Division of Thoracic Surgery, McMaster University, St. Joseph's Healthcare, Hamilton, Ontario, Canada
[h] Department of Thoracic Surgery, AUSL Romagna Teaching Hospital, Forlì, Italy
[i] Thoracic Surgery, Hospital Universitario de Salamanca, Salamanca, Spain
[j] Department of Thoracic Surgery, Jagiellonian University Collegium Medicum, John Paul II Hospital, Krakow, Poland
[k] Department of Thoracic Surgery, Nouvel Hopital Civil, Strasbourg, France
[l] Department of Thoracic Surgery, University of Torino, Turin, Italy
[m] Department of Thoracic Surgery, Hospital Germans Trias i Pujol, Badalona, Spain
[n] Department of Thoracic Surgery, University Hospital Zurich, Zurich, Switzerland
[o] Department of Thoracic Surgery, Marmara University Faculty of Medicine, Istanbul, Turkey
[p] Department of General Thoracic Surgery, Istanbul Medical Faculty, Istanbul, Turkey

* Corresponding author. Department of Thoracic Surgery, AUSL Romagna Teaching Hospital, Viale Vincenzo Randi 5, 48121 Ravenna, Italy.
  Tel: +39-0544-285111; fax: +39-0544-1761304; e-mail: luca.bertolaccini@gmail.com (L. Bertolaccini).

Abstract

OBJECTIVES: In the European Society of Thoracic Surgeons (ESTS) survey about management of malignant pleural effusions (MPE), 56% of respondents are not informed of any relevant clinical guidelines and 52%, who are aware of the existence of guidelines, declared that they are in need of updating or revision. The ESTS Pleural Diseases Working Group developed a benchmarking project on quality of previous guidelines on the management of MPE.

METHODS: The Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument was used to assess each guideline. Each item was scored on a 7-point scale. Scores for each domain were calculated. Economic data for the nations which have issued the guidelines were collected from the Organisation for Economic Cooperation and Development health statistics database.

RESULTS: Six guidelines fitted the inclusion criteria and were assessed. Five out of 6 guidelines were produced by a multinational collaboration. Observers would recommend only 2 guidelines with minimal modification. Two areas that received the best score were clarity of presentation and scope and purpose (objectives and health questions target population). The applicability of guideline domain had the lowest score. Multivariate analysis demonstrated that clarity of presentation, international guidelines and publication through medical journal were related to improved scores. A strong correlation was observed between the measures of economic status.

CONCLUSIONS: The quality of guidelines assessed by the AGREE II criteria was found to be extremely variable. Guidelines achieving higher AGREE II scores were more likely to come from the European Union with the direct involvement of scientific societies in their development. It was also recognized that some fundamental unanswered questions remain about the management of MPE.

Keywords: Lung cancer · Benchmarking project · Quality · Malignant pleural effusion · AGREE II

INTRODUCTION

Modern medical practice is best evidence-based and largely influenced by guidelines. Therefore the quality of guidelines is reflected in their real-clinical application. In this setting, availability, content and regular implementation are key factors and tend to be determined not only by clinician's willingness but also by organisational, political and socio-economic elements. Malignant pleural effusion (MPE) commonly complicates advanced malignancy, posing a high burden of symptomatic disease. The European Society of Thoracic Surgeons (ESTS) survey about the management of MPE revealed that 56% of respondents are not aware of the existence of any clinical guidelines and 52% declared that the current guidance needed updating or revision [1]. Last available clinical guidance was published in 2010 by the British Thoracic Society [2]. Since then, it has been an expansion in our understanding of this subject and our pharmacological and interventional armamentarium. A search of the largest clinical trials registry (clinicaltrials.gov) in January 2017 revealed 136 ongoing trials dealing with the topic of MPE at the time of writing this article [3]. Clinical practice is thought to be resultantly highly variable, although no published data previously existed to support this claim. The ESTS Pleural Diseases Working Group was then created with the aim of assessing the current literature and benchmarking the quality of previous guidelines on the management of MPE. The objectives of the group were to create a database of the current guidelines and to assess the methodology adopted in producing such guidelines.

MATERIAL AND METHODS

A professionally qualified health librarian carried out the literature search using recognized health and social care databases (detail in Supplementary Material). Guidelines databases and Internet search engines were selected as appropriate. A mixture of keywords (free text) and subject headings, mapped to the thesaurus, were used to ensure a thorough search of the selected databases and websites. Moreover, as guidelines are not necessarily published in scientific journals, other sources were explored. The results were deduplicated using the Healthcare Database Search (HDAS) tool and Endnote Web and scanned by the librarian to abstract level to ensure relevancy. The librarian then removed irrelevant results and any remaining duplicates manually (details in Supplementary Material). Working Group members selected guidelines for review. The last update of the selected guidelines was considered. National guidelines were included only if published in a peer-reviewed journal and the Google Translate tool (http://translate.google.it/) was used for conversion into the English language when necessary. The translations were checked within the working group. The Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument was used to assess each guideline [4]. The purpose of the AGREE II is to provide: (i) a framework to evaluate the quality of guidelines, (ii) a methodological strategy for the development of guidelines and (iii) guidance on the content and best reporting strategy. Six observers (A.B., K.M., F.R., G.R., Y.S. and P.S.) independently scored each guideline. The AGREE II consists of 23 key items organised within 6 domains followed by 2 global rating items ('Overall Assessment'). Each domain captures a unique dimension of guideline quality. The AGREE Collaboration defined the quality of guidelines as the confidence that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible for practice [5]. The 6 evaluating domains in the AGREE II instrument are:

1. Scope and purpose (3 items). Concerned with the overall aim of the guideline, the specific health questions and the target population.
2. Stakeholder involvement (3 items). Focuses on the extent to which the guideline was developed by the appropriate interested parties and represents the views of its intended users.
3. Rigour of development (8 items). Relates to the process used to gather and synthesize the evidence, the methods to formulate the recommendations and to update them.
4. Clarity of presentation (3 items). Deals with the language, structure and format of the guideline.
5. Applicability (4 items). Deals with the possible barriers and facilitators to implementation, strategies to improve uptake and resource implications of applying the guideline.
6. Editorial independence (2 items). Concerned with the formulation of recommendations not being unduly biased with competing interests.

Each item was rated on a 7-point scale. Overall assessment includes the rating of the overall quality of the guideline and whether the guideline would be recommended for use in practice. For each domain, a score (expressed as a percentage) was calculated based on the individual domain-specific item scores of the 6 observers, as recommended in the AGREE II manual [6]: the higher the score, the better the methodological quality of the guideline for the corresponding AGREE II domain. Each item was scored on a 7-point scale (1 = strongly disagree and 7 = strongly agree). Scores for each domain were calculated by the sum of all items within a domain and scaling the score as a percentage of the maximum possible score using the following formula:

$$\text{Scaled domain score} = \frac{\text{Obtained score} - \text{Minimum possible score}}{\text{Maximum possible score} - \text{Minimum possible score}}$$

The results from each guideline were summarized with values for each domain. All 23 items of the AGREE II instrument were assessed with results reported in percentage form for each of the 6 domains. Each reviewer received a user's manual of the AGREE II instrument, containing instructions. Furthermore, the following information was recorded: the country, the language, the year of publication and the scientific society behind the guideline. Economic data for the nations, which have issued the guidelines were collected from the Organisation for Economic Cooperation and Development health statistics database [7]. Organisation for Economic Cooperation and Development data included the percentage of gross domestic product allocated to health expenditure and the absolute amount of health expenditure (per capita). Data were collected for the year of the guideline publication. Costs were converted from US$ to € as per the conversion rate on 15 September 2016.

Statistical analysis

The characteristics of the guidelines and the AGREE II scores were descriptively analysed. Correlation between the AGREE II

THORACIC

Table 1:   Guidelines included in the AGREE II evaluation with information on language and sources of retrieval

Issuing society	Title of guideline	Year	Country	Language	Availability in PubMed	Ref.
American Thoracic Society	Management of malignant pleural effusions	2000	USA	English	Available	[9]
European Respiratory Society	Management of malignant pleural effusions	2001	Europe	English	Available	[10]
Dutch Association of Physicians for Pulmonary Diseases and Tuberculosis	Diagnosis and treatment of malignant pleural effusion	2006	Netherland	Danish		[11]
British Thoracic Society	Management of a malignant pleural effusion	2010	UK	English	Available	[2]
American College of Chest Physicians	Symptom management in patients with lung cancer—Diagnosis and management of lung cancer, 3rd ed	2013	USA	English	Available	[12]
National Comprehensive Cancer Network	Clinical practice guidelines in oncology—Non-small cell lung cancer version 4.2016	2016	USA	English		[13]

domains was measured by the Bravais–Pearson correlation coefficients. The analyses of variance (for categorical factors) or regression models (for continuous covariates) were used to assess the role of guideline characteristics on the AGREE II scores. Univariate analyses selected factors with a P-value < 0.30 for inclusion in the multivariate analyses. The inter-rater reliability between the 6 observers was determined using the Krippendorff's alpha, an extension of Cohen's kappa to evaluate concordance or agreements between multiple raters, objects and categories. A P-value < 0.05 was defined as significant. Since all analyses were exploratory, there was no adjustment for multiplicity. R (version 3.2.3, Wooden Christmas-Tree with standard, *rcmdr*, and *irr* packages) was used for statistical analyses [8].

RESULTS

In total, 188 guidelines were found; 6 guidelines fitted the inclusion criteria and were assessed by the 6 observers (Table 1). Publication year ranged from 2000 to 2016 and 3 (50%) guidelines were updated between 2010 and 2016. Five out of 6 guidelines (83%) were produced by a multinational collaboration (29 countries). A governmental organisation was involved in 2 (33%) guidelines and a scientific society in 5 (83%). Four guidelines (67%) were published in a scientific journal and 2 (33%) in other media (e.g. website). Four guidelines were identified through Medline, one through guideline databases and another through homepages of national and international societies. As expected, some guidelines focused on MPE, while others had a wider content. The 6 observers would only recommend 2 guidelines with minimal modification. Table 2 shows the analysis of each of the 6 domains score related to each guideline. The lowest score regarding overall assessment was received from the American Thoracic Society guidelines [9]; the highest score was received from the Dutch Association of Physicians guidelines due to a very high score of the clarity of presentations [11]. Nevertheless, Table 3 shows the descriptive analysis of the 6 domain scores and the overall assessment. The 2 areas that received the best score were D4—clarity of presentation and D1—scope and purpose (objectives and health questions target population). The domains

D5—applicability of the guideline had the lowest score (Table 2). All 6 observers rated all the domains of the AGREE II without data missing. Krippendorff's alpha showed that D1—scope and purpose and D6—editorial independence coefficients yield high inter-rater reliability values; the overall assessment of the guidelines showed an optimal inter-rater correlation. The D5—applicability of guidelines and the D3—rigour of development showed the lowest inter-rater reliability (Table 3).

Involvement of stakeholders was strongly correlated with the rigour of development (Pearson correlation coefficient = 0.92) and clarity of presentation (Pearson correlation coefficient = 0.96). Rigour of development was also strongly correlated with the clarity of presentation (Pearson correlation coefficient = 0.93). These correlations were significantly different from the null hypothesis (all P < 0.05). The lowest correlation was between the rigour of development and editorial independence (Pearson correlation coefficient = 0.08). Table 4 shows the results of the univariate analyses of variance for the categorical variables. The scores for 2 domains, D3—rigour of development and D4—clarity of presentation, were influenced by 2 factors: international guidelines for the rigour of development and international guidelines for clarity of presentation. The guidelines published in the scientific journals had better scores than those published in other media. The international guidelines also had a positive influence on scope and purpose. Finally, very recent guidelines were associated with involvement of the stakeholder. Multivariate analysis showed that for the clarity of presentation, the international guidelines and the publication through scientific journal were related to improve scores (Table 5).

A strong correlation was observed between the 2 covariates that were measures of economic status (Pearson correlation coefficient = 0.98, P < 0.0001). The percentage of health spending of the national gross domestic product was analysed as a continuous variable. The absolute amount of health expenses (per capita) were dichotomized according to the median value (€3573), and it was evident that the countries with higher expenditure on healthcare produced significantly better guidelines (Table 6). The involvement of the scientific societies and the publication year were also associated with more expenditure on healthcare. Consequently, multivariate modelling was limited to health

Table 2: AGREE II scores (%) by different domains of the guidelines analysed

Issuing society	Title of guideline	D1—scope and purpose	D2—stakeholder involvement	D3—rigour of development	D4—clarity of presentation	D5—applicability	D6—editorial independence	Overall assessment
American Thoracic Society	Management of malignant pleural effusions	57	37	44	69	33	57	53
European Respiratory Society	Management of malignant pleural effusions	64	48	51	72	47	63	69
Dutch Association of Physicians for Pulmonary Diseases and Tuberculosis	Diagnosis and treatment of malignant pleural effusion	88	69	85	94	53	61	83
British Thoracic Society	Management of a malignant pleural effusion	72	62	61	88	56	86	75
American College of Chest Physicians	Symptom management in patients with lung cancer—Diagnosis and Management of lung cancer, 3rd ed.	80	68	75	87	53	67	78
National Comprehensive Cancer Network	Clinical practice guidelines in oncology—Non-small cell lung cancer version 4.2016	58	65	65	85	48	57	75

Table 3: AGREE II scores by domain with the inter-rater reliability between observers

Domain	Mean ± SD %	95% CI for the mean %	Median %	Minimum %	Maximum %	Krippendorff's alpha
D1—scope and purpose	70 ± 12	57–83	68	57	88	0.86
D2—stakeholder involvement	58 ± 13	45–72	63	37	69	0.61
D3—rigour of development	64 ± 15	48–79	63	44	85	0.43
D4—clarity of presentation	83 ± 10	73–93	86	69	94	0.54
D5—applicability	48 ± 8	40–57	51	33	86	0.36
D6—editorial independence	65 ± 11	54–77	62	57	86	0.96
Overall assessment	72 ± 11	61–83	83	53	75	0.86

SD: standard deviation; CI: confidence interval.

expenditure (per capita), and publication in a scientific journal. Only 1 domain, rigour of development, influenced the score with adjustment for the health spending. Quality seems to be mostly driven by guidelines writing organisations promoting audits.

DISCUSSION

While we acknowledge that there are several accepted methods for quality improvement research in healthcare, the ESTS Pleural Diseases Working Group chose the AGREE II instrument based on a previously published paper about quality management [14]. The AGREE II is intended to be used by the following stakeholder groups: by health care providers who wish to undertake their assessment of a guideline before adopting its recommendations into their practice; by guideline developers to follow a structured and rigorous development methodology, to conduct an internal

evaluation to ensure that their guidelines are sound, or to evaluate guidelines from other groups for potential adaptation to their own context; by policy makers to help them decide which guidelines could be recommended for use in practice or to inform policy decisions; and by educators to help enhance critical appraisal skills amongst health professionals and to teach core competencies in guideline development and reporting [5]. This review of the guiding principles of the management of MPE is, to the authors' knowledge, the most comprehensive systematic analysis. The 6 guidelines that were assessed differed widely in content and goals but had some similarities in the subject and structure. Guideline development methodology was highly variable across all domains of AGREE II. The management of MPE varies widely depending on the operating surgeon, as almost half of the participants who responded to the ESTS survey for MPE did not follow any guideline and felt the need to have a revision of the current guidance. The current guidance was published over 5

THORACIC

Table 4: Analysis of variance for the 6 AGREE II domains. Data are presented as *n* (%) or mean

Factor	Evaluated guidelines	AGREE II domains					
		D1	D2	D3	D4	D5	D6
Level of the guidelines							
International	5 (83)	66	56	59	80	48	66
National	1 (17)	87	69	85	94	53	61
P-value		0.106	0.394	0.127	0.245	0.572	0.739
Involvement of a scientific society							
Yes	5 (83)	73	65	63	82	49	67
No	1 (17)	58	57	65	85	48	57
P-value		0.356	0.628	0.918	0.799	0.949	0.480
Publication in a scientific journal							
Yes	4 (33)	68	54	58	78	47	68
No	2 (67)	73	67	75	89	51	59
P-value		0.697	0.268	0.217	0.257	0.696	0.400
Publication year							
Before 2012	4 (33)	70	54	60	81	47	67
2012–2016	2 (67)	69	66	70	86	51	62
P-value		0.912	0.300	0.511	0.580	0.696	0.662

D1: scope and purpose; D2: stakeholder involvement; D3: rigour of development; D4: clarity of presentation; D5: applicability; D6: editorial independence.

Table 5: Results of multivariate analysis

	AGREE II domains					
	D1	D2	D3	D4	D5	D6
Level of the guidelines	0.081		0.115	0.043		
Involvement of a scientific society		0.895	0.148	0.041		
Publication in a scientific journal		0.331				
Publication year						

Data are presented as *P*-values.
D1: scope and purpose; D2: stakeholder involvement; D3: rigour of development; D4: clarity of presentation; D5: applicability; D6: editorial independence.

Table 6: The impact of the economic situation on the AGREE II domains

Factor	AGREE II domains					
	D1	D2	D3	D4	D5	D6
Percentage of gross domestic product dedicated to health expenditure						
Regression coefficient	0.64	0.70	0.60	0.66	0.59	0.58
Health expenditure per capita (€)						
<3573€	71	55	67	83	47	60
>3573€	68	60	56	80	51	74
P-value	0.048	0.144	0.445	0.167	0.315	0.158

D1: scope and purpose; D2: stakeholder involvement; D3: rigour of development; D4: clarity of presentation; D5: applicability; D6: editorial independence.

years ago, and it is widely composed of recommendations based on moderate (Level B) or low (Level C) quality evidence [15].

The ESTS recognized this need and, through the Pleural Disease Working Group, set the ambitious project to elaborate new recommendations. The first step in this task is to assess the current guidelines to understand their clinical effectiveness and the criteria that the new ones should follow. From our analysis, there is a limited applicability of the current guidelines, as the domain D5-applicability had the lowest scores. So much so that the reviewers recommended only 2 out of the 6 selected guidelines. The rigour of development (D3) was strongly correlated with the clarity of presentation (D4) and these 2 domains substantially associated with the involvement

of stakeholders (D2). That means that a new set of guidelines should be clear, concise and involve surgeons as stakeholders so that, hopefully, a larger number of them will feel encouraged to adopt them in their clinical practice. Furthermore, the international collaboration, the involvement of the scientific societies and the publication in a scientific journal had a positive influence on the quality of the guidelines (overall higher scores). Countries with larger health care financial resources also had better guidelines. Multivariate analyses showed that AGREE II scores were higher when guidelines were developed at an international level with the involvement of scientific societies. Therefore organisations supported guideline development groups probably could produce a better outcome.

Our analysis supports the fact that new guidelines should then be international, supported by at least 1 scientific society with the aim to be published in a scientific journal, to use the reviews as a further improvement tool.

The main limitations of our project were the subjective nature of the AGREE II tool and the potential bias of the reviewers performing the assessment. The AGREE II tool is a 23-question instrument established to evaluate guidelines quality. AGREE II is sensitive to differences in important aspects of clinical practice guidelines, and it can be used consistently by a wide range of professionals from different cultural backgrounds. Health professionals, policy makers and consumers were all able to appraise guidelines with the AGREE questions and user guide. The appraisers found the instrument easy to apply and perceived it to be useful for judging the quality of guidelines. The AGREE II is built for the guideline developers to follow a structured and rigorous development methodology. Although it is a subjective tool, it is the current gold standard; the AGREE II guidelines suggest using at least 2 and preferably 4 appraisers with content-specific knowledge [6]. We employed 6 reviewers who all had content-specific knowledge and experience in research and evidence-based methods. The AGREE II is an instrument to assess the methodological rigour and transparency of the guideline's development. It informs about what is written in the guideline and how this should be reported in the guideline. However, there is no threshold for discrimination from high-quality to low-quality criteria. And moreover, some guidelines did not report detailed methodology, so a low-domain score may not always reflect low quality. Thus the AGREE II overall scores should be interpreted with caution and in specific contexts. Given that the AGREE II assessment requires an evaluation of the guidelines based on the descriptions available in the published manuscripts, there was a small chance that inaccurate estimates would be due to poor descriptions in the document. A multidisciplinary management of MPEs is necessary. Likewise, we agree the surgeons may significantly alter the diagnostic and therapeutic pathway of these patients yielding possible over treatment. However, we feel that the above represent perfectly sound reasons to have guidelines written by surgeons to influence the role of specialists in the decision-making process. Indeed, in other multidisciplinary contexts (i.e., lung cancer screening), focused work on the role of surgeons has been recognized as a significant contribution to the development of an interdisciplinary approach to the condition [16].

Despite these limits, this benchmarking project represents the first attempt to record the current practice of MPE. It is also important to note that some guidelines did not report detailed methodology, so a low-domain score may not always reflect low quality. Besides, use of assessment tools such as AGREE II during guideline development phase allows the guideline-developing team to follow the process step-by-step.

CONCLUSIONS

Although there are many guidelines published in the literature, the quality of these, assessed by the AGREE II criteria, was found to be extremely variable with relatively low average scores. Those guidelines achieving higher AGREE II scores were more likely to come from the European Union with a direct involvement of scientific societies in development. It was also recognized that some fundamental unanswered questions remain about the management of MPE. The ESTS Pleural Diseases Working Group is an ambitious initiative with the aim to improve the quality of care for patients with MPE across Europe. The next steps of the ESTS Pleural Diseases Working Group will be an extensive review of the literature on management of MPE and the development of updated recommendations or Clinical Practice Guideline.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *EJCTS* online.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES

[1] Scarci M, Caruana E, Bertolaccini L, Bedetti B, Brunelli A, Varela G et al. Current practices in the management of malignant pleural effusions: a survey among members of the European Society of Thoracic Surgeons (ESTS). Interact CardioVasc Thorac Surg 2016; doi:10.1093/icvts/ivw373.

[2] Roberts ME, Neville E, Berrisford RG, Antunes G, Ali NJ; BTS Pleural Disease Guideline Group. Management of a malignant pleural effusion: British Thoracic Society Pleural Disease Guideline 2010. Thorax 2010;65:ii32–40.

[3] ClinicalTrials.gov [Internet]. Bethesda, MD: National Library of Medicine (US). https://clinicaltrials.gov/. (27 January 2017, date last accessed).

[4] AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003;12:18–23.

[5] AGREE Next Steps Consortium (2009). The AGREE II Instrument [Electronic version]. http://www.agreetrust.org (03 October 2017, date last accessed).

[6] Brouwers MC, Kho ME, Browman GP, Burgers J, Cluzeau F, Feder G et al. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Can Med Assoc J 2010;182:E839–42.

THORACIC

[7]   Organisation for Economic Cooperation and Development. Data Visulisation for Key OECD Data. https://www.oecd.org/ (15 September 2016, date last accessed).

[8]   R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2016. https://www.R-project.org/.

[9]   Management of malignant pleural effusions. Am J Respir Crit Care Med 2000;162:1987–2001.

[10] Antony VB, Loddenkemper R, Astoul P, Boutin C, Goldstraw P, Hott J *et al.* Management of malignant pleural effusions. Eur Respir J 2001;18:402–19.

[11] Groen H, Welling A, Postmus PE, van Meerbeeck JP, van Kasteren JHLM. Diagnostiek en behandeling van maligne pleuravocht. Richtlijn van de Nederlandse Vereniging van Artsen voor Longziekten en Tuberculose. [Diagnosis and Treatment of Malignant Pleural Effusions. Guideline of the Dutch Society of Pulmonologists.] Alphen aan den Rijn: van Zuiden Communications, 2006.

[12] Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e142S–65S.

[13] National Comprehensive Cancer Network. Non-Small Cell Lung Cancer (Version 4.2016). www.nccn.org/professionals/physician_gls/pdf/nscl.pdf (15 September 2016, date last accessed).

[14] Blum TG, Rich A, Baldwin D, Beckett P, De Ruysscher D, Faivre-Finn C *et al.* The European initiative for quality management in lung cancer care. Eur Respir J 2014;43:1254–77.

[15] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.

[16] Rocco G, Allen MS, Altorki NK, Asamura H, Blum MG, Detterbeck FC *et al.* Clinical statement on the role of the surgeon and surgical issues relating to computed tomography screening programs for lung cancer. Ann Thorac Surg 2013;96:357–60.