Standard Outcome Measures for Thymic Malignancies

James Huang, MD,† Frank C. Detterbeck, MD, Zuoheng Wang, PhD,* and Patrick J. Loehrer, Sr, MD‡*

Abstract: Thymic malignancies present particular issues due to the pace of disease progression, patterns of recurrence, and causes of death that make nuances of how outcomes are reported particularly important. The relatively limited number of patients also creates a challenge to glean as much as possible from the available experience, but risks over-interpretation and potentially misleading conclusions. Therefore the International Thymic Malignancy Interest Group has developed a set of standards for reporting of outcome measures of clinical studies, which have been adopted for collaborative projects undertaken by the organization. Widespread adoption of this baseline will enhance the ability to compare results from different series.

Key Words: Thymoma, Mediastinal disease, Thymic carcinoma, Statistics, Outcomes, Recurrence.

(*J Thorac Oncol.* 2010;5: 2017–2023)

Thymic malignancies are relatively uncommon, with an incidence of approximately 2.5 to 3.2 per 10^6 people,[1,2] and care is widely dispersed across many institutions. The literature consists almost exclusively of retrospective single institution series, which often extend over multiple decades of experience to have a reasonable number of patients. Comparing the results from one center to another is often difficult because of differences in the outcomes that are reported and the definitions used. Significant progress cannot be made unless a standard and uniform set of definitions and outcomes measures are adopted.

The International Thymic Malignancy Interest Group (ITMIG) is a collaborative effort of interested individuals around the world to develop an infrastructure that facilitates progress in this disease. One of the first steps in this process is the development of standard outcome measures and definitions. This article describes the measures adopted by the ITMIG membership that form the basis for ITMIG collaborative projects.

*Division of Thoracic Surgery, Department of Surgery, Yale University School of Medicine and Department of Biostatistics, School of Epidemiology and Public Health, Yale University, New Haven, Connecticut; †Thoracic Service, Memorial Sloan-Kettering Cancer Center, New York City, New York; and ‡HH Gregg Professor of Oncology, Division of Medical Oncology, Department of Internal Medicine, Indiana University School of Medicine, Indianapolis, Indiana.

Disclosure: The authors declare no conflicts of interest.

Address for correspondence: Frank C. Detterbeck, MD, Yale University School of Medicine, 330 Cedar St. BB 205, New Haven, CT 06520-8062. E-mail: Frank.detterbeck@yale.edu

ISSN: 1556-0864/10/0512-2017

METHODS

The process used in development of this document was designed to represent both underlying evidence and a broad consensus of ITMIG members. An initial workgroup consisting of surgeons, a medical oncologist, and a statistician (J.H., F.C.D., P.J.L., and Z.W.) was assembled to review measures that have been used in the existing literature This workgroup formulated preliminary recommendations, which were refined by an extended workgroup (Giuseppe Giaccone, Gregory Riely, Nicolas Girard, Meinoshin Okumura, Charles Thomas, Edith Marom, Andrea Bezjak, and Alexander Brunelli) and distributed to all ITMIG members for further discussion and input. The final recommendations, which are presented in this article, were approved and adopted by ITMIG members at the annual meeting in New York on May 6, 2010.

PROPOSED MEASURES

Stage Classification

No official stage classification for thymic malignancies has been defined by the Union Internationale Contre le Cancer and the American Joint Commission on Cancer. Various staging systems have been proposed,[3] including the Masaoka system,[4] the Koga modification of the Masaoka system,[5] the French Groupe d'Étude des Tumeurs Thymiques system,[6] and a T, N, and M system.[7] Most of centers and published reports use the Masaoka Stage Classification System, with studies since 1995 generally using the Koga modification (Table 1, Masaoka-Koga). The Masaoka-Koga stage classification system is recommended by ITMIG for current use.

The Koga modification differs from the original Masaoka system in that microscopic invasion into (but not through) the capsule is classified as a stage IIb by Masaoka but as stage I by Masaoka-Koga. This modification is supported by the fact that most pathologists do not consider partial invasion into the capsule to be significant, and survival data appear to bear this out.[5,8,9] Furthermore, this definition of the staging system is consistent with the definition of encapsulated and invasive thymoma adopted by ITMIG (Which way is up? A collaborative position paper on standards of handling and processing of thymic tissue by surgeons and pathologists, submitted). Another difference is that adherence to adjacent structures or microscopic invasion into but not through the mediastinal pleura or pericardium is classified as stage IIb by Masaoka-Koga but is not clearly defined in the original Masaoka definition.

TABLE 1. Masaoka-Koga Staging System

Stages	Definitions
I	Grossly and microscopically completely encapsulated tumor
IIa	Microscopic transcapsular invasion
b	Macroscopic invasion into thymic or surrounding fatty tissue or grossly adherent to but not breaking through mediastinal pleura or pericardium
III	Macroscopic invasion into neighboring organ (i.e., pericardium, great vessel, or lung)
IVa	Pleural or pericardial metastases
b	Lymphogenous or hematogenous metastasis

Adapted from *Pathol Int*.[5]

Despite widespread use, the Masaoka-Koga system has many ambiguities that have not been clearly defined. These issues are beyond the scope of this article and will be addressed in a separate publication by ITMIG. Furthermore, evaluation and statistical validation of the stage classification of thymic malignancies are being undertaken by ITMIG and the International Staging Committee of the International Association for the Study of Lung Cancer. This requires collection of data and details beyond the Masaoka-Koga stage and will include evaluation of alternative stage classification schemas. This will be addressed separately in a manual associated with the ITMIG database.

The focus of staging has been on pathologic stage (i.e., as defined after resection). However, the clinical stage (the assessment before treatment is initiated) is of much more clinical importance, especially because surgery is not always the first step in the treatment. Unfortunately, the correlation of tumor characteristics and the reliability of staging tests in defining clinical stage have not been well defined. This subject must be addressed prospectively in a more detailed manner in a future publication. Until such definitions are available, we suggest that authors estimate the stage according to the Masaoka-Koga system based on their best judgment. We strongly encourage authors to report not only the pathologic but also the clinical stage.

Survival

A standard outcome measure is overall survival. It is concrete and generally easy to verify and certainly should be reported in any clinical outcomes study of thymoma. For many cancer types, this is an adequate general measure of outcomes related to the cancer, because survival after a recurrence is generally short, and the majority of deaths are due to the original cancer. However, thymomas have a number of characteristics that make overall survival problematic for assessment of thymoma-related outcomes. Many patients die of other causes, especially in stage I and stage II tumors (Figures 1*A, B*). Patients may also live for many years despite a recurrence. Therefore, more specific measures are needed in addition to overall survival.

More specific measures generally involve either considering a specific cause of death or type of recurrence, or both. The issue of whether the cause of death is determined accurately or unduly attributed to cancer has been studied in general and found to be quite reliably assigned,[10–12] and it is highly likely to apply to thymoma as well. However, there are issues with what is considered a relevant cause of death in the existing literature on thymoma. Disease-free survival, recurrence-free survival, cancer-specific survival, progression-free survival, etc., each involve a different definition of a relevant outcome event as shown in Table 2. These differences are important in the case of thymoma, because the incidence of these events is high (Figure 1*A*). Furthermore, approximately 15 to 20% of patients with thymoma have or develop another type of cancer as well.[3,13,14] As an example, the estimated outcomes for a stage III thymoma using these different definitions is provided in Figure 2.

Actuarial outcome curves that depict a specific type of event (e.g., only local recurrence or a particular cause of death) are often misleading, usually yielding overly optimistic results.[15] This is because the actuarial method requires statistical independence of the specific event from others, which is generally not the case (e.g., the time to local recurrence and time to distant recurrence are likely to be correlated).[15] Depending on the degree to which outcome events are linked, an actuarial estimate of only one type of outcome (a flawed method) can easily underestimate the actual rate of this outcome by 30 to 50%.[15] Therefore, it is better to analyze death or failure in general, and then to compare proportional causes of death or failure to avoid this problem of "competing risks."[15] We propose, therefore, that whenever possible, the proportion of recurrence types and the proportion of causes of death be reported.

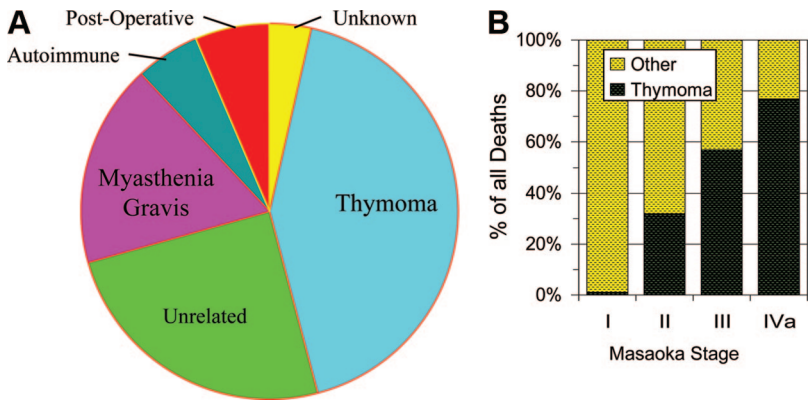

FIGURE 1. Overall cause (*A*) and stage-specific (*B*) cause of death after resection of patients with thymoma. Results are an average of studies from 1980 to 2009 of ≥100 patients reporting this data.[3]

TABLE 2. Survival Measures

Measures	Events (End Point of Interest)	Censored Observations[a]	Included Patients
Overall survival	Death, any cause		All[b]
Disease-related survival	Death from TM, MG,[c] Treatment	Unrelated death,[d] unknown cause of death	All[b]
Disease-specific survival	Death from TM, MG[c]	Unrelated death,[d] unknown cause of death	All[b]
Cause-specific survival	Death from TM	Unrelated death,[d] unknown cause of death	All[b]
Cancer-specific survival	Death from any Cancer	Unrelated death,[d] unknown cause of death	All[b]
Disease-free survival[e]	Death, recurrence	Unknown recurrence status	R0/CR
Freedom from recurrence	Recurrence	Dead without recurrence; Unknown recurrence status	R0/CR
Progression-free survival	Death, progression of TM	Unknown status of TM	R1,2/PR, SD
Time-to-progression	Progression of TM	Dead without progression, Unknown status of TM	R1,2/PR, SD

[a] In all categories includes patients lost to follow-up or without an event at termination of study period.
[b] Either all patients or may be restricted to a specific group (e.g., R0 resected patients only).
[c] Or other thymoma-related disease (e.g., red cell dyscrasia and hypogammaglobunlinemia).
[d] Definition of related determined by the definition of an event.
[e] Disease-free survival usually means the same as recurrence-free survival but sometimes may include MG as an event.
CR, complete radiographic response (no residual tumor identified); MG, myasthenia gravis; PR, partial response; SD, stable disease (after treatment with chemoradiotherapy or radiotherapy); TM, thymic malignancy.

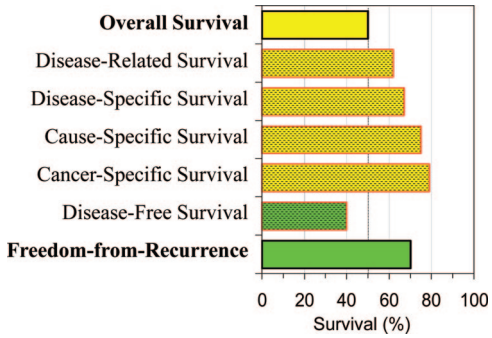

FIGURE 2. Specific outcomes for a stage III resected thymoma at 10 years, estimated from data regarding overall survival, cause of death, incidence of recurrence, and incidence of other cancers.[3]

TABLE 3. Recommended Outcome Measures

Measures	Patient Cohort	Starting Point
Overall survival	For all patients	Date of diagnosis
Freedom-from-recurrence	For patients after successful curative treatment (R0 resection or radiographic CR after chemotherapy or radiotherapy)	Completion of treatment
Time-to-progression	For treated patients in whom all disease was not eradicated (R1,2 resection, radiographic SD, or PR)	Completion of treatment

CR, complete response, PR, partial response, SD, stable disease.

The approach proposed by ITMIG is that assessment of the efficacy of treatment of a thymic malignancy is best measured by the rate of recurrence of the thymic malignancy (at any site). The ability to control related diseases such a Myasthenia Gravis should be viewed separately. Furthermore, the cause of death is also a suboptimal measure that is affected by other factors, and death is best not mixed together with a recurrence end point. Therefore, we recommend that freedom-from-recurrence is the best measure for patients who have successfully undergone curative-intent treatment (patients with no evidence of disease, i.e., either an R0 resection or a radiographic complete response). For patients in whom the disease was never eradicated, time-to-progression is the best measure (Table 3). Time-to-progression should also be used after an R1 resection, because residual disease is thought to exist. (The end point of these measures—recurrent disease—is the same; but the appropriate term depends semantically on whether disease is still thought to be still present after treatment or not. Time-to-progression is chosen for palliative treatments because it is a well-accepted measure among medical oncologists. Freedom-from-recurrence is chosen for curatively treated patients rather than time-to-recurrence because the former emphasizes the likelihood of a positive outcome, whereas the latter creates the impression that recurrence is only a matter of time.)

Because of the indolent behavior of thymoma and the fact that recurrence is associated with a mean survival of many years, we suggest that for overall survival, 10-year outcomes should be reported in addition to 5-year outcomes. For thymic carcinoma, on the other hand, we suggest that 5-year survival is an appropriate measure.

The average time to recurrence of a completely resected thymoma has been found to be approximately 5 years (range of reported average, 3–7 years).[13,16–21] One study suggested there was a difference according to stage, observing a mean time to recurrence of 10 years in patients with a stage I thymoma, compared with 3 years in patients with a stages II to IV thymoma.[22] (This suggests that more indolent tumors, with a longer time to recurrence, are more likely to be detected while they are still stage I.) Given these results, we suggest that 5-year outcomes are appropriate for freedom-from-recurrence studies in general, although it is probably better to have 5- and 10-year results for stage I tumors.

A concept that is often forgotten is that actuarial survival curves (e.g., Kaplan-Meier) provide an estimate of survival (because some patients have been lost or have not been followed up for the entire duration of the study). The variance of this

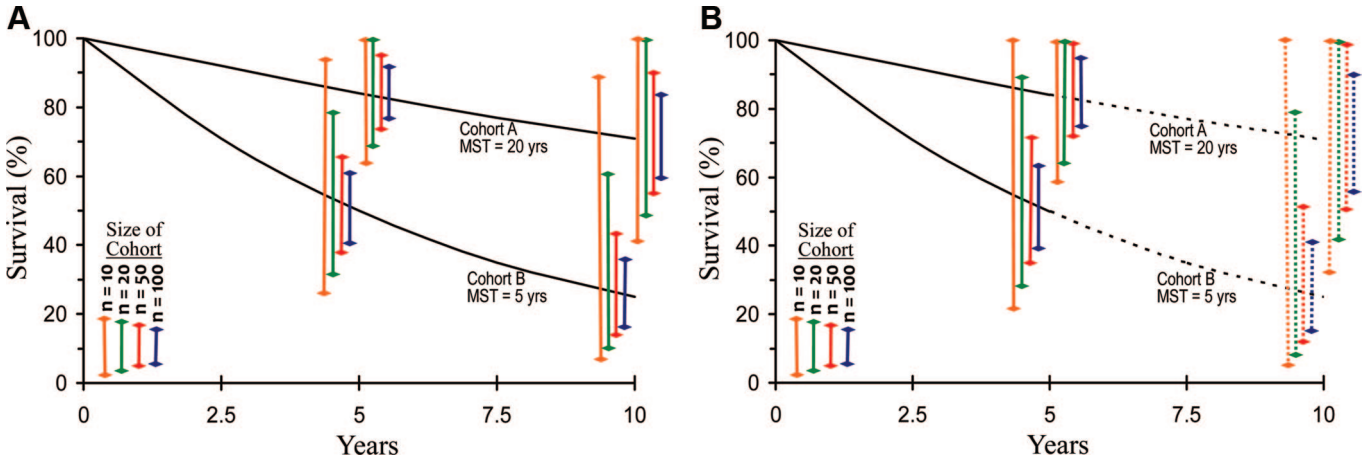

FIGURE 3. Variance in actuarial survival estimates by size of cohort for (*A*) a 10-year study duration and (*B*) a 5-year study duration. The vertical bars are 95% confidence intervals for the survival estimate at 5 and 10 years, based on a standard model of exponentially decreasing survival, a constant rate of accrual of patients during the course of the study until study termination, and no loss to follow-up. MST, median survival time.

estimate is dependent on the sample size, the duration of follow-up, the study duration, and the median survival. The variance is fairly wide for cohorts of less than approximately 50 patients, and this is accentuated in studies of shorter duration (Figure 3*A, B*). In a rare disease, large patient cohorts are often simply not available; however, it is important to have an assessment of the limitations of the available data. Therefore, we recommend that the confidence interval for a survival estimate should be provided. We also propose that the median follow-up be reported in all studies (duration from time of patient entry into the study until the event of interest, termination of the study, or loss to further follow-up, as a median for the entire study cohort).

Completeness of Surgical Resection

The completeness of resection has frequently been identified by multivariate analysis as a major prognostic factor.[3,23] However, details of how specimens are handled and how margins are reported vary among studies. Because of the nature of the anatomy and the often loose areolar tissue surrounding the thymus, there is potential for significant misclassification (ITMIG. Which way is up? A collaborative position paper on standards of handling and processing of thymic tissue by surgeons and pathologists, submitted). ITMIG has proposed standards for handling, processing, and reporting of pathologic results of resected thymic tumors to minimize this (ITMIG. Which way is up? A collaborative position paper on standards of handling and processing of thymic tissue by surgeons and pathologists, submitted). In brief, areas of concern should be marked immediately during dissection in a way that minimizes disruption of the overlying tissues and provides clear orientation and communication between the surgeon and pathologist. The pathology staff should take care in the gross preparation of the sample to section in a way that allows definition of these areas. A positive margin denotes tumor extending to an inked surface of the specimen that involves tissues that have been cut or dissected. Exposed tumor on a mediastinal pleural or pericardial surface, which is bounded by the space of a normal body cavity (i.e., no adhesions present), does not constitute a positive margin. One of the difficulties in thymoma resection is that the loose areolar tissues surrounding the thymus can easily be disrupted during handling, resulting in exposed tumor in the specimen that was in fact not due to dissection on the surface of the tumor (ITMIG. Which way is up? A collaborative position paper on standards of handling and processing of thymic tissue by surgeons and pathologists, submitted). Such areas should not be counted as a positive margin, provided the area has been clearly identified as not having been grossly concerning in situ and having been disrupted during handling, as outlined in a separate publication (ITMIG. Which way is up? A collaborative position paper on standards of handling and processing of thymic tissue by surgeons and pathologists, submitted).

Definition of Recurrence

The term recurrence is appropriate if it is thought that all disease has been potentially eradicated (an R0 resection or a complete radiographic response has been achieved). We propose that a recurrence be defined to have occurred when there is a strong clinical suspicion (or proof), without a specific requirement of what tests (i.e., imaging and biopsy) this is based on. Given that the survival with disease can sometimes justify observation alone, requiring tissue confirmation or a decision to treat will underestimate recurrence. The time of the recurrence should be recorded as the time when a strong suspicion first existed, even if subsequent events (i.e., radiographic progression or biopsy) make the clinical diagnosis even more secure at a later time. When significant doubt exists, a biopsy is encouraged whenever feasible.

Rarely, there may be a strong suspicion that is later conclusively shown to have been in error (i.e., biopsy or spontaneous regression). In such cases, the diagnosis of recurrence should be (retroactively) rescinded. However, the onset of a recurrence should not be retroactively assigned, meaning that one should not reassign the onset of recurrence to a time before there was any clinical suspicion, even if a retrospective review of imaging demonstrates an abnormality that was missed but was later found to be a recurrence.

The frequency of follow-up imaging may affect when a recurrence is first suspected. We suggest that at a minimum, yearly computed tomography (CT) scans of the thorax be performed for 5 years after surgical resection and then alternating annually with a chest radiograph until year 11, followed by annual chest radiographs alone. Resected stage III or IVa thymoma, thymic carcinoma, incomplete resection, or other high-risk tumors are suggested to undergo additional CT imaging every 6 months for 3 years. Obtaining a new "baseline" examination after resection when acute inflammatory effects have resolved (i.e., 4–12 weeks postoperatively) may be very useful for comparison. Magnetic resonance imaging may be useful instead of CT either for better visualization or to minimize cumulative radiation dose (especially in young patients). Positron emission tomography (PET) imaging is not recommended for routine surveillance but, if available, may selectively be of benefit (e.g., to investigate a clinical or radiographic suspicion of recurrence). A more frequent follow-up schedule has been proposed[21] (annual CT for life), but this may be less broadly applicable in the disparate health care systems around the world.

A local recurrence should be defined as disease appearing in the bed of the thymus (i.e., the anterior mediastinum) or tissues immediately contiguous with the normal thymus or with the thymoma (Table 4). This should include involvement of lymph nodes or pleural or pericardial tumor that is immediately adjacent to the previous primary tumor. Similarly, a recurrence in the exact area of a previously resected noncontiguous pleural metastasis (stage IVa) should be classified as a local recurrence, with the additional notation of "recurrence of a noncontiguous metastasis." Finally, disease in the lower neck contiguous with the location of the upper poles of the thymus should be classified as local recurrence.

A regional recurrence should be defined as intrathoracic tumor that is not immediately contiguous with the thymus gland or the previous thymic neoplasm. This would include pleural (parietal or visceral) and pericardial nodules as noted in Table 4, unless these are in the bed of the previously treated tumor. A nodule that is contiguous with the pleura (including the fissures) is designated as a visceral pleural nodule. A regional recurrence also includes lymph nodes that are not adjacent to the thymus or the thymic tumor (e.g., periesophageal or more distant neck nodes).

A distant recurrence includes disease outside the thorax or the lower neck. Disease in the peritoneal cavity or retroperitoneum should be classified as distant unless it is arising from local extension through the diaphragm of an intrathoracic nodule(s).

We propose that the term distant recurrence also include nodules that are clearly intrapulmonary, with a clinically visible rim of lung tissue between it and the visceral pleura (either radiographically or on gross examination). This is based on speculation that the mechanism of spread of a nodule under the visceral pleura is through the pleural space and that of an intraparenchymal pulmonary nodule is through the bloodstream. However, there are no data to substantiate this speculation and no data defining the clinical impact of making this distinction. To prospectively study this, a distant recurrence consisting only of an intraparenchymal nodule should be recorded separately from other extrathoracic recurrences.

TABLE 4. Definitions of Recurrence (After R0 Resection or Radiographic Complete Response)

Local recurrence—anterior mediastinum
Tumor occurring in bed of thymus or previously resected thymoma
Includes pericardial, pleural, or pulmonary tumor that is immediately adjacent to the thymus or previously resected thymoma
Lymph nodes immediately adjacent to the thymus or previously resected thymoma (including nodes in the neck immediately adjacent to the upper poles of the thymus)
Recurrence at the site of a previous noncontiguous metastasis (stage IVa)—should be specifically noted as such
Regional recurrence—intrathoracic recurrence not contiguous with thymus or previous thymoma
Parietal pleural nodules
Pericardial nodules
Visceral pleural nodules
Mediastinal lymph nodes not adjacent to the normal thymus or the previous thymic malignancy
Distant recurrence
Extrathoracic recurrence
Intraparenchymal pulmonary nodules (with rim of normal lung between the nodule and the visceral pleura)

Response to Chemotherapy and Radiation

Chemotherapy and radiation are often used as induction therapy before resection, as definitive therapy with curative intent for inoperable disease, or for palliation. The standard way to assess a patient's response to therapy is unidimensional tumor measurement by CT according to the RECIST criteria (version 1.1)[24] Although these criteria may be suitable for most solid tumors in general, the anatomic aspects of thymic malignancies, with their large size, location, irregular shape, and intimate relationship with neighboring structures make it particularly difficult to attain consistent measurements. Measurements using either hand-held or electronic calipers have high intra- and interobserver variability, particularly for tumors with irregular or vague borders.[25,26] Therefore, we recommend that tumor response assessment be performed by one person, ideally a radiologist experienced with tumor measurements.[25]

Thymic malignancies may also respond by undergoing cystic change, central necrosis, and density changes that may not be captured by conventional measurements of the largest lesion diameter. Another issue is that in lymphocyte-rich thymomas, a significant response can be seen simply by the effect of chemotherapy (or even prednisone) on the normal lymphocytes and not the tumor cells. Other complicating factors include the predilection of thymic malignancies for pleural involvement, which poses a significant challenge for administering standard RECIST methods. Although the standard RECIST criteria exclude pleural nodules, because of their importance in thymoma, these should be measured and included, similar to mesothelioma.[27] Tumor volume measurements may enable more objective, accurate, and consistent assessment of therapy response than the traditional unidimensional technique. These approaches require further investigation and validation and are currently being studied.

In the interim, conscientious adherence to standard revised RECIST criteria (version 1.1) for response assessment is recommended.[24] All measurable lesions up to a maximum of two lesions per organ and five lesions in total, representative of all involved organs, should be identified as target lesions and recorded and measured at baseline. Target lesions should be selected on the basis of their size (i.e., the longest diameter [LD]) and their suitability for accurate repeated measurements (either by imaging techniques or clinically). A sum of the longest (or most accurately determined diameters) for all target lesions is calculated and serves as the reference by which objective tumor response is assessed.[28]

An exception to this rule is pleural lesions, as the tumor thickness is more clearly and consistently measured than the long axis spread along the lining of the chest wall or the mediastinum.[27] For extensive pleural involvement, the tumor thickness perpendicular to the chest wall or mediastinum is measured in two positions at three separate levels on thoracic CT scans.[27] The sum of the six measurements defines a pleural unidimensional measure. The sum of the pleural unidimensional measure with the sum of the LD (or the most accurately determined diameter) for all non-pleural target lesions up to a total of five defines the baseline sum LD, against which subsequent tumor response or progression is compared.[24,27]

Disease progression should be defined according to the revised RECIST criteria (version 1.1).[24] The timing of recognition of progression is, of course, potentially also influenced by the frequency of assessment. We propose that for resected stage III or IVa thymoma, thymic carcinoma, incomplete resection, or other high risk tumors undergo imaging at least every 6 months for 3 years.[21] Earlier stage tumors should follow annual imaging as outlined earlier in the Definition of Recurrence section.

Evidence of treatment effect can also be assessed by histologic evaluation of tumor tissue after chemotherapy or radiotherapy. Inflammation, necrosis, and fibrosis are often seen admixed with viable tumor, and the degree of treatment effect can be visually quantified as a percentage of viable tumor in increments of 10%. A pathologic complete response exists when no viable tumor is identified throughout the entire specimen. The number of sections examined to determine this is important. At least one section for every centimeter of tumor diameter is recommended (ITMIG. Which way is up? A collaborative position paper on standards of handling and processing of thymic tissue by surgeons and pathologists, submitted).

Multivariate Analysis

The availability of statistical software has made it much easier to perform multivariate analyses to identify potential prognostic factors. However, the statistical science is often not well understood by the authors, and conclusions are often overstated. Essentially, all prognostic factor studies in thymic malignancies to date have been exploratory, so-called phase I prognostic factor studies. There is nothing wrong with such exploration, but external validation is essential.[29] There are many potential sources of bias. For example, when using stepwise regression analysis, different predictive factors may emerge, depending on whether a forward model or backward model is used and depending on the order in which variables are introduced. The regression coefficients in the final model are generally overestimated, thus the p value in these analyses is not actually valid (because the same data used to define the variable is used to estimate the prognostic effect).[29] The same applies to survival curves with and without the prognostic factor that have not adjusted for other factors. Furthermore, when a cutoff value is chosen to dichotomize a continuous variable based on the best discriminatory power, the chance of a false-positive prognostic factor is about 40%, despite a p value of <0.05.[30] The technique of bootstrapping can overcome some of these difficulties.[31]

There is also a high chance of false-negative results when investigating potential prognostic factors, mainly because of insufficient sample size. It is easy to see how a conclusion can be reached that a factor has no independent prognostic significance, when in reality the sample size was too small to detect a difference (for a given effect size). A general estimate of the minimum sample size needed can be calculated using an online tool for multiple regression analysis, with results shown in Table 5[32] This must be viewed as only a rough estimate, because other characteristics of the cohort (e.g., uneven distribution of a potential prognostic factor) also play a role. A more detailed analysis requires the involvement of a statistician.

Because sample size is a major issue in a rare disease, such as thymoma, performing and interpreting the results of a multivariate analysis must be done with caution. We recommend that multivariate studies of prognostic factors involve the help of a statistician. The study should include definition of the parameters chosen for the analysis (i.e., level of significance and power). Furthermore, the results should be stated in a way that represents an accurate perspective (e.g., "this study suggests that several factors are not likely to have a large prognostic effect, although the sample size is insufficient to evaluate a small or medium-sized effect" or perhaps "the factors did not appear to have prognostic significance, but the power of detection of even a large effect in this limited study was only 0.40"). Ideally, the hazard ratio of

TABLE 5. Sample Size Needed for Multiple Regression Analysis

No. of Predictors to be Tested	Power = 0.80 Effect Size[a]			Power = 0.60 Effect Size[a]		
	Small	Medium	Large	Small	Medium	Large
2	478	67	31	308	44	21
3	543	76	36	356	52	25
4	597	84	39	395	58	28
5	643	91	43	429	63	31
6	684	97	46	460	68	33
7	721	103	49	488	72	36
8	755	108	52	513	76	38
9	788	113	54	538	80	40
10	818	118	57	561	84	42

Estimate of minimum sample size needed to assess a given number of potential predictors, assuming a level of significance of $p = 0.05$.

Values calculated using an online tool available at: http://www.danielsoper.com/statkb/topic01.aspx.[32]

[a] Effect size using conventional levels of (f^2) of 0.02, 0.15, and 0.35 for small, medium, and large effects, respectively.

each variable should be reported together with 95% confidence intervals. However, in most cases, the findings must be externally validated before drawing any serious conclusions.

DISCUSSION

A common language is essential to sharing experiences across different centers, and because thymic malignancies are uncommon, such collaboration is crucial. ITMIG is an organization of individuals interested in this disease and is devoted to providing infrastructure and processes that foster collective research. This article establishes a baseline for ITMIG regarding how outcomes should be reported. Details of how endpoints are defined are clarified, to avoid the ambiguity and inability to combine data that have persisted in the past because of a lack of uniformity. These definitions will be used in ITMIG collaborative projects. We also hope that this will be a useful guide in general for studies of thymic malignancies.

We have also included basic facts regarding the statistics of clinical outcomes as they apply to a relatively uncommon and often indolent disease such as thymoma. There are limitations to the strength of conclusions that can be drawn from studies involving a modest number of patients. Pointing this out is not meant to discourage analysis of the experience in individual centers, which will often suffer from a limited number of patients. Instead, it is meant to help the general medical community remain realistic about the strength of the data that we have. There is nothing wrong with trying to glean as much knowledge as possible from the experience we have, but there are potential issues if we overstep what can be concluded without realizing it.

Any "standard" must be viewed as merely a baseline in a fluid process. We hope this article will encourage all who are reporting on thymic malignancies to use the proposed definitions and hope it will stimulate a careful appraisal of the weaknesses and of unproven measures. Furthermore, we hope this article will inspire an exploration of new measures. We fully anticipate subsequent formal changes to occur to at least some of the proposed measures once a good scientific basis or at least a solid consensus has been established.

In summary, this article establishes outcome measures and definitions appropriate for thymic malignancies. This represents a consistent baseline agreed on by ITMIG that will be used in ITMIG collaborative projects, and we hope it will be a useful guide for the medical community in general.

REFERENCES

1. de Jong WK, Blaauwgeers JLG, Schaapveld M, et al. Thymic epithelial tumours: a population-based study of the incidence, diagnostic procedures and therapy. *Eur J Cancer* 2008;44:123–130.
2. Engel P, Marx A, Muller-Hermelink HK. Thymic tumours in Denmark. A retrospective study of 213 cases from 1970–1993. *Pathol Res Pract* 1999;195:565–570.
3. Detterbeck F, Parsons AM. Thymic tumors: a review of current diagnosis, classification, and treatment. In, Patterson GA, Cooper JD, Deslauriers J, Lerut A, Luketich JD, Rice TW, Pearson FG (eds). Thoracic and Esophageal Surgery, 3rd Ed. Philadelphia: Elsevier, 2008. Pp. 1589–1614.
4. Masaoka A, Monden Y, Nakahara K, et al. Follow-up study of thymomas with special reference to their clinical stages. *Cancer* 1981;48: 2485–2492.
5. Koga K, Matsuno Y, Noguchi M, et al. A review of 79 thymomas: modification of staging system and reappraisal of conventional division into invasive and non-invasive thymoma. *Pathol Int* 1994;44:359–367.
6. Gamondès JP, Balawi A, Greenland T, et al. Seventeen years of surgical treatment of thymoma: factors influencing survival. *Eur J Cardiothorac Surg* 1991;5:124–131.
7. Yamakawa Y, Masaoka A, Hashimoto T, et al. A tentative tumor-node-metastasis classification of thymoma. *Cancer* 1991;68:1984–1987.
8. Kornstein MJ, Curran WJ, Turrisi AT III, et al. Cortical versus medullary thymomas: a useful morphologic distinction? *Hum Pathol* 1988;19: 1335–1339.
9. Kornstein MJ. Controversies regarding the pathology of thymomas. *Pathol Annu* 1992;27:1–15.
10. Satariano WA, Ragland KE, Stephen VDE. Cause of death in men diagnosed with prostate carcinoma. *Cancer* 1998;83:1180–1188.
11. Penson DF, Albertsen PC, Nelson PS, et al. Determining cause of death in prostate cancer: are death certificates valid? *J Natl Cancer Inst* 2001;93:1822–1823.
12. Albertsen PC, Walters S, Hanley JA. A comparison of cause of death determination in men previously diagnosed with prostate cancer who died in 1985 or 1995. *J Urol* 2000;163:519–523.
13. Wilkins KB, Sheikh E, Green R, et al. Clinical and pathologic predictors of survival in patients with thymoma. *Ann Surg* 1999;230:562–574.
14. Welsh JS, Wilkins KB, Green R, et al. Association between thymoma and second neoplasms. *JAMA* 2000;283:1142–1143.
15. Gelman R, Gelber R, Henderson IC, et al. Improved methodology for analyzing local and distant recurrence. *J Clin Oncol* 1990;8:548–555.
16. Regnard J-F, Magdeleinat P, Dromer C, et al. Prognostic factors and long-term results after thymoma resection: a series of 307 patients. *J Thorac Cardiovasc Surg* 1996;112:376–384.
17. Lewis JE, Wick MR, Scheithauer BW, et al. Thymoma: a clinicopathologic review. *Cancer* 1987;60:2727–2743.
18. Maggi G, Casadio C, Cavallo A, et al. Thymoma: results of 241 operated cases. *Ann Thorac Surg* 1991;51:152–156.
19. Blumberg D, Port JL, Weksler B, et al. Thymoma: a multivariate analysis of factors predicting survival. *Ann Thorac Surg* 1995;60:908–914.
20. Ruffini E, Mancuso M, Oliaro A, et al. Recurrence of thymoma: analysis of clinicopathologic features, treatment, and outcome. *J Thorac Cardiovasc Surg* 1997;113:55–63.
21. Huang J, Rizk NP, Travis WD, et al. Comparison of patterns of relapse in thymic carcinoma and thymoma. *J Thorac Cardiovasc Surg* 2009; 138:26–31.
22. Verley JM, Hollmann KH. Thymoma. A comparative study of clinical stages, histologic features, and survival in 200 cases. *Cancer* 1985;55: 1074–1086.
23. Detterbeck FC, Parsons AM. Thymic tumors. *Ann Thorac Surg* 2004; 77:1860–1869.
24. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–247.
25. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol* 2003;21:2574–2582.
26. Schwartz LH, Ginsberg MS, DeCorato D, et al. Evaluation of tumor measurements in oncology: use of film-based and electronic techniques. *J Clin Oncol* 2000;18:2179–2184.
27. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann Oncol* 2004;15:257–260.
28. Therasse P, Arbuck SG, Eisenhauer EA, et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. *J Natl Cancer Inst* 2000;92:205–216.
29. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncolong. *Br J Cancer* 1994;69:979–985.
30. Altman DG, Lausen B, Sauerbrei W, et al. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994;86:829–835.
31. Grunkemeier GL, Wu Y. Bootstrap resampling methods: something for nothing? *Ann Thorac Surg* 2004;77:1142–1144.
32. Soper DS. Statistics Knowledge Base—A-Priori Sample Size for Multiple Regression. 2010. Available at: http:www.danielsoper.com/statkb/topic01.aspx.

