

***American Thoracic Society  
Clinical Practice Guideline  
Development Manual***

Audience: Guideline committee members

Last updated: May 2020

## **TABLE OF CONTENTS**

1. Guideline development committee
2. Conflict of interest
3. Work plan
4. Clinical Questions
5. Outcomes
6. Evidence synthesis
  - a. Search strategy
  - b. Study selection
  - c. Data extraction and meta-analysis
  - d. Summary of findings and evidence tables
7. Formulating recommendations
8. Determining the strength of recommendations
9. Writing recommendations
10. Writing the guideline manuscript

# **Chapter 1:**

## ***Guideline Development Committee***

The first and perhaps the most important step in developing a clinical practice guideline is enlisting the appropriate personnel. All clinical practice guidelines require a chair and a methodologist:

- The chair should be an individual who has experience leading groups. The ideal chair is pragmatic, diplomatic (i.e., can mediate disagreements and facilitate compromise), efficient, and persuasive (i.e., can coax individuals to get work done in a timely fashion). There is a natural tendency to recruit a chair that is prominent in the field, but this is not the best approach unless the individual has these qualities.
- The methodologist should have experience leading both systematic reviews and projects that have applied the Grading, Recommendations, Assessment, Development, and Evaluation (GRADE) approach to formulating, writing, and grading recommendations. It is helpful if the methodologist has good communication skills, as he or she must work closely with the chair and is frequently called upon to explain the rationale for various steps in the guideline development process. For projects without their own methodologist, the ATS has a methodology training program, which will provide one or more methodology trainees (i.e., methodology Scholars) to perform the methodological work with the mentorship of the

ATS' methodologist. In exchange, it is expected that the Scholar(s) will be listed as a middle co-author on the final guideline and the first author on any systematic reviews that are published as independent derivatives.

The chair selects and invites potential guideline development committee members. An ideal guideline development committee consists of 10 to 20 individuals. Too many individuals can prove unwieldy and unmanageable. Guideline development committees should include clinicians, researchers, patients or patient advocates, and other stakeholders, such as payers and representatives from healthcare organizations. Among the clinicians, multiple disciplines should be represented, including nurses, therapists, primary care physicians, and relevant specialists. Inclusion of a specialist who cares for patients with common co-morbidities is strongly encouraged; as an example, a guideline addressing chronic obstructive pulmonary disease (COPD) might include a cardiologist, since many patients with COPD have cardiac co-morbidities.

For more information on formation of a guideline development committee, please contact:

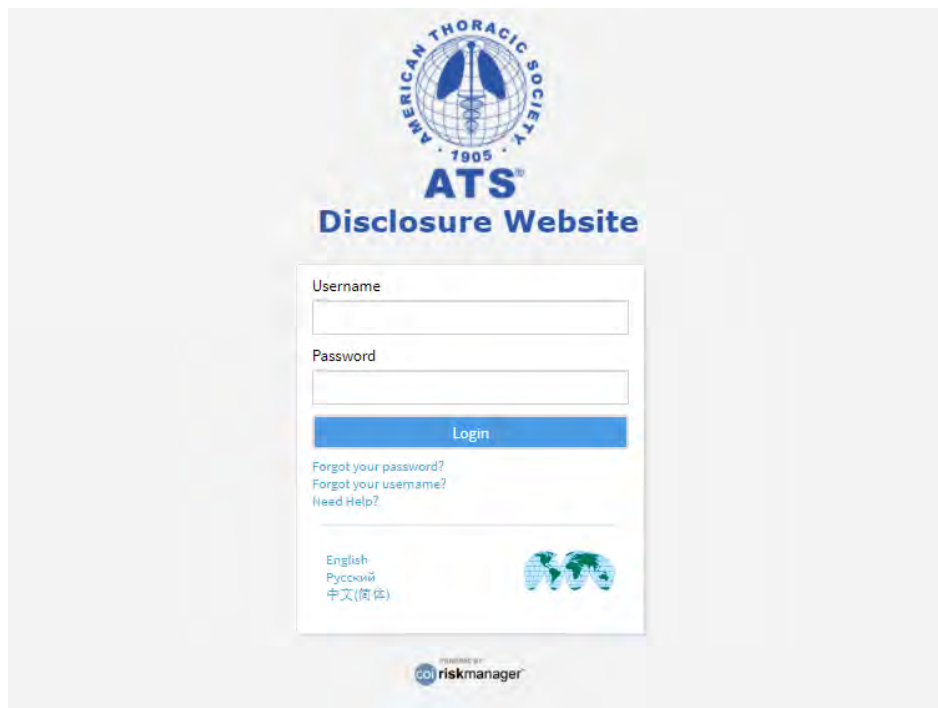
- Charlie B. Strange III, MD (Chair, DDIC), [strangec@muscc.edu](mailto:strangec@muscc.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Guidelines and Documents), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)
- Judy Corn (Sr. Director, Documents & Patient Education), [jcorn@thoracic.org](mailto:jcorn@thoracic.org)
- John Harmon (Sr. Manager, Conflict of interest [COI] Management), [jharmon@thoracic.org](mailto:jharmon@thoracic.org)

# **Chapter 2:**

## ***Conflict of Interest***

Disclosure and management of potential conflicts of interest (COI) are important aspects of guideline development. Many clinical practice guidelines have come under scrutiny for inadequate COI management, which can discredit or cast doubt on years of work performed by a guideline development committee.

Upon approval of a clinical practice guideline project, the chair is asked to disclose his or her potential conflicts of interest using the database, COI-Smart (<https://thoracic.coiriskmanager.com/>).



The chair's disclosures are vetted by the ATS' Conflict of Interest Office, with assistance from the Document Development and Implementation Committee (DDIC) and the Ethics and Conflict of Interest Committee (ECOI), if necessary. The ATS' Conflict of Interest Office then conveys any concerns and necessary actions to the chair. Occasionally, a chair's conflicts disqualify him or her from participation. In such rare instances, the relevant assembly will work with ATS leadership and the DDIC to identify a suitable replacement.

Once a chair is approved to begin, he or she invites individuals to participate on the guideline development committee. Each potential participant must disclose his or her potential conflicts of interest using COI-Smart (<https://thoracic.coiriskmanager.com>). The ATS' Conflict of Interest Office will vet the disclosures for the entire proposed committee with assistance from the DDIC and ECOI, if necessary, and then compile a list that categorizes the conflicts as acceptable or disqualifying. The implications of the findings and next steps are then discussed with the chair.

- Acceptable conflicts of interest include industry-funded research, data safety monitoring board participation, and unpaid, non-promotional, educational speaking. They are the types of activities that the ATS encourages, but they need to be acknowledged and managed. An example of managing an acceptable conflict is to disclose the conflict and allow the individual to participate in all related discussions about the evidence, but to excuse the individual from discussing, writing, and grading related recommendations.

- Disqualifying conflicts of interest include participation on speakers' bureaus and scientific advisory committees, consultation, employment, stock ownership, etc. related to the content of the guidelines. Individuals with disqualifying conflicts have three options: 1) they may decide to not participate in the guideline, 2) they may terminate the pertinent industry relationships and participate as someone who has manageable conflicts of interest as described above, or 3) they may participate as "non-voting expert contributors", meaning that they can discuss evidence but cannot formulate, vote on, write, or edit any recommendations, regardless of whether or not the recommendations are related to their conflicts. Non-voting expert contributors cannot serve as chairs or co-chairs. Employees are not eligible to serve as non-voting expert contributors.

Guideline development committees are required to have at least one co-chair and a majority of participants who are completely free of conflicts related to the content of the guidelines. The remainder of participants may have acceptable conflicts if appropriately managed. For multi-society guidelines, the approach to COI may differ slightly from that described above because compromises are frequently necessary in light of differing approaches across societies.

For more information on conflict of interest policy and procedures, please contact:

- Shane McDermott (Sr. Director, Ethics and Professionalism, [smcdermott@thoracic.org](mailto:smcdermott@thoracic.org))
- Kevin Wilson, MD (Chief, Guidelines and Documents [kwilson@thoracic.org](mailto:kwilson@thoracic.org))
- John Harmon (Sr. Manager, Document Development), [jharmon@thoracic.org](mailto:jharmon@thoracic.org)

## **Chapter 3:**

### ***Work Plan***

There are many ways to develop a rigorous, evidence-based clinical practice guideline. One approach that has consistently proven successful is described here. The premise underlying the approach is that guideline development can be conceptualized as having three parts: The first part is determining the scope of the guideline and specifying a manageable number of clinical questions that will eventually be translated into recommendations. The second part is evidence synthesis, which involves searching the literature, selecting relevant studies, extracting and pooling data, summarizing the body of evidence, and rating the quality of evidence; and the third part is developing and grading recommendations based upon the evidence. Most guideline development committee members are content experts whose interests and skills are best suited to generating and prioritizing clinical questions and developing recommendations based upon evidence; they tend to be less interested and skilled in evidence synthesis. Therefore, it is generally best to defer the evidence synthesis portion of guideline development to methodologists who are experienced in this aspect of guideline development and to allow the content experts to focus on that which interests them most and for which their skills are best suited. . . developing recommendations from the evidence.



**Step 1, Leadership teleconference:** Most guideline projects begin with a teleconference among the co-chairs and methodologist. Topics typically include introductions, disclosure of potential conflicts of interest (COI) to one another, scope of the guideline, work plan, roles, and timeline.

**Step 2, Full committee teleconference:** Next, a teleconference is held with the full committee. The chair, methodologist, and committee members are introduced to one another, potential COI are disclosed, and the scope, timeline, and work plan are described as agreed upon in step one. Feedback on the plan may be solicited. It is important to emphasize to the full committee that: 1) a clinical practice guideline is different from a narrative review in that it consists of answering diagnostic and/or treatment questions via an evidence review and evidence-based recommendations; there is no need to review epidemiology, clinical manifestations, pathogenesis, etc. except in a very limited fashion as background information, and 2) the scope of a guideline is generally five to eight clinical questions answered by recommendations; trying to do more leads to delays and burnout among the committee.

**Step 3, Selecting clinical questions:** The methodologist works with the co-chairs to design an initial survey to brainstorm for important clinical questions and a second survey to prioritize the questions. The questions are then ranked and the most important are selected for inclusion in the guideline. See chapter 4.

**Step 4, Selecting outcomes:** The methodologist works with the co-chairs to design an initial survey to brainstorm for important outcomes and a second survey to prioritize the outcomes.

The outcomes are then prioritized. Outcomes with a low priority score are discarded, while those with a moderate priority score are categorized as important and those with a high priority score are categorized as critical. See chapter 5.

**Step 5, Evidence synthesis:** Once the guideline development committee has selected the guideline's clinical questions and outcomes, the methodologist leads the evidence synthesis. This consists of looking for published high-quality systematic reviews and, if none exist, then performing a new systematic review or pragmatic evidence synthesis for each clinical question. Both systematic and pragmatic reviews require designing a search strategy, determining study selection criteria, selecting pertinent studies, extracting data from the selected studies, pooling the data via meta-analysis, and constructing both summary of findings and evidence tables. Evidence synthesis is the most labor intensive and time consuming part of guideline development.

While the evidence synthesis is largely performed by the methodologist, he or she will reach out to the co-chairs or full committee for help periodically. As examples, the methodologist may need the co-chairs' input regarding synonyms, abbreviations, etc. when designing the search strategy, and at the completion of study selection will ask the committee whether any pertinent evidence may have been missed so that oversights and errors can be corrected. See chapter 6.

**Step 6, Reviewing evidence and formulating recommendations:** Moving from evidence to recommendations often begins during a face-to-face meeting of the full guideline development committee. During the meeting, participants' potential COI are disclosed and then the evidence for each clinical question is summarized by the methodologist. Evidence-based recommendations are then formulated using the following criteria: the balance of benefits versus harms and burdens, the quality of evidence, patient values and preferences, and costs and resource use. If consensus on the appropriate recommendations cannot be achieved by discussion, voting may be required.

**Step 7, Grading the recommendations:** Once the recommendations have been determined, each must be rated according to the strength of the recommendation (strong or weak/conditional) and the quality of evidence (high, moderate, low, or very low). The strength of the recommendation indicates the degree of certainty that the committee has that the desirable consequences of the recommended intervention outweigh the undesirable consequences, while the quality of evidence indicates the confidence that the committee has in the estimated effects that it used to inform its judgments.

For more information on developing an appropriate work plan, please contact:

- Kevin Wilson, MD (Chief, Guidelines and Documents), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)

# Chapter 4:

## *Clinical Questions*

The methodologist works with the co-chairs to design a series of surveys whose purpose is to determine the clinical questions that will be answered by the guideline. The first survey consists of a single open-ended question. For example, “*What important clinical questions should be answered in this guideline?*”

The image shows a screenshot of a SurveyMonkey survey. The title bar at the top reads "Guideline on the Management of Asthma Exacerbations". Below that, the survey title is "Identifying important clinical questions". The main question is "1. What important clinical questions should be answered in this guideline?". There are ten numbered input fields, labeled "Question 1" through "Question 10", each with a corresponding text box. At the bottom of the survey, there is a "Done" button and a footer that says "Powered by SurveyMonkey" and "Check out our sample surveys and create your own now!"

Once the full committee has responded to the initial survey, the methodologist rephrases the questions in a PICO (Population, Intervention, Comparator, Outcomes) format and then sends a second survey whose purpose is to prioritize the clinical questions. For example, the question may be, “*These are the clinical questions that you and your colleagues determined are*

*important. Please prioritize each on a scale of 1 to 9.”*

**Guideline on the Management of Asthma Exacerbations**

These are the clinical questions that you and your colleagues decided are important.

**1. Should patients having an asthma exacerbation (P) be treated with short-acting beta agonists via nebulizer (I) or meter dose inhaler (C)?**  
Assign score of 1 (lowest priority) to 9 (highest priority)

**2. Should patients having an asthma exacerbation (P) be treated with short-acting anticholinergics (I) or not treated with short-acting anticholinergics (C)?**  
Assign score of 1 (lowest priority) to 9 (highest priority)

**3. Should patients admitted to the hospital due to an asthma exacerbation (P) be treated with intravenous glucocorticoids (I) or oral glucocorticoids (C)?**  
Assign a score of 1 (lowest priority) to 9 (highest priority)

**4. Should patients with an asthma exacerbation who are being treated with oral glucocorticoids (P) be prescribed 1.0-1.5 mg/kg (I) or 0.5-0.99 mg/kg (C)?**  
Assign a score of 1 (lowest priority) to 9 (highest priority)

**5. Should patients with an asthma exacerbation who are being treated with oral glucocorticoids (P) be prescribed a single dose for five days (I) or a tapered dose over two weeks (C)?**

Finally, the committee members' responses are recorded (usually in a Microsoft Excel spreadsheet) and the median priority score for each question is determined. The questions with the highest median priority scores are selected for the guideline, with the exact number of questions determined by the previously agreed upon scope of the guideline.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	RS	LPB	CB	EB	MC	PS	RD	NJ	MG	FC	SW	PG	PC	Median			
2	8	8	3	8	8	3	5	8	7	6	6	7	2	7			
3	8	6	3	8	7	3	6	8	7	4	6	7	2	6			
4	8	5	3	5	6	2	3	7	3	5	5	4	2	5			
5	1	9	3	7	3	8	2	5	4	4	9	4	4	4			
6	5	8	3	7	3	6	2	5	3	4	8	4	2	4			

For more information on selecting clinical questions, please contact:

- Charlie B. Strange III, MD (Chair, DDIC), [strange@muscd.edu](mailto:strange@muscd.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Documents & Patient Education), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)

## Chapter 5:

# Outcomes

The methodologist works with the co-chairs to design a series of surveys whose purpose is to determine which outcomes should be considered for each clinical question during guideline development. The first survey consists of a single open-ended question. For example, “*Which outcomes do you think are important for each clinical question?*” The committee should be reminded that adverse outcomes should be included; they are often overlooked, but they are essential to formulating recommendations later.

**Guideline on Management of Asthma Exacerbations: Outcomes 1**

List outcomes that matter for each clinical question

**1. Should patients having an asthma exacerbation be treated with short-acting beta agonists via nebulizer or meter dose inhaler?**

Outcome 1

Outcome 2

Outcome 3

Outcome 4

Outcome 5

**2. Should patients having an asthma exacerbation be treated with short-acting anticholinergics or not treated with short-acting anticholinergics?**

Outcome 1

Outcome 2

Outcome 3

Outcome 4

Outcome 5

Once the full committee has responded to the initial survey, the methodologist rephrases the outcomes in a way that distinguishes patient-important outcomes from surrogates. For example,

if the committee listed “FEV1” as an outcome, the methodologist may rephrase the outcome as, “dyspnea (measured by FEV1)”. The methodologist then sends a second survey whose purpose is to prioritize the outcomes. The survey says something like, *“These are the patient-important outcomes that you and your colleagues listed as important. Please prioritize each on a scale of 1 to 9 (a priority score of 7 to 9 indicates that the outcome is of sufficient importance to change your clinical practice, a priority score of 4 to 6 indicates that the outcome is important but may not change your clinical practice, and a priority score of 1 to 3 indicates that the outcome is not important).”*

**Guideline on Management of Asthma Exacerbations: Outcomes 2**

These are the patient-important outcomes that you and your colleagues decided are important.

Prioritize each outcome on a scale of 1 to 9. A score of 7 to 9 indicates that the outcome is of sufficient importance to change your clinical practice, 4 to 6 indicates that the outcome is important but may not be enough to change your clinical practice, and 1 to 3 indicates that the outcome is not important.

**1. Should patients having an asthma exacerbation be treated with short-acting beta agonists via nebulizer or meter dose inhaler?**

Mortality	<input type="text"/>
Intubation Rate	<input type="text"/>
Hospitalization Rate	<input type="text"/>
Duration of Exacerbation	<input type="text"/>
Palpitations	<input type="text"/>
Tremulousness	<input type="text"/>

**2. Should patients having an asthma exacerbation be treated with short-acting anticholinergics or not treated with short-acting anticholinergics?**

Mortality	<input type="text"/>
Intubation Rate	<input type="text"/>
Hospitalization Rate	<input type="text"/>
Duration of Exacerbation	<input type="text"/>
Frequency of Dry Mouth	<input type="text"/>

Done

Finally, the committee members’ responses are recorded (usually in a Microsoft Excel spreadsheet) and the median priority score for each outcome is determined. Outcomes with a median priority score of 1 to 3 are discarded, while those with a median priority score 4 to 6 are categorized as



“important” and those with a median priority score 7 to 9 are categorized as “critical”. It is especially important to get input about outcomes from patients or patient representatives who serve on the guideline development team.

		RS	LPB	CB	EB	MC	PS	RD	NJ	rMG	FC	SW	PG	PC	QH	KW	Median
2	Should patients having an asthma exacerbation be treated with short-acting beta agonists via nebulizer or meter dose inhaler?	9	7	9	9	8	7	9	9	9	8	8	9	8	9	9	9
3	mortality	8	6	4	9	6	3	8	7	7	7	8	8	2	5	9	7
4	intubation and ventilation	7	7	7	5	5	2	7	8	6	6	8	5	6	7	9	7
5	hospitalization	9	9	5	7	5	9	5	5	6	2	5	4	9	7	4	5
6	duration of exacerbation	2	4	5	7	5	8	4	4	5	3	7	4	4	4	1	4
7	palpitations	2	2	5	3	5	3	2	1	2	4	1	4	2	1	1	2
8	tremulousness																
9	Should patients having an asthma exacerbation be treated with short-acting anticholinergics or not treated with short-acting anticholinergics?	9	7	9	9	8	7	9	9	9	8	8	9	8	9	9	9
10	mortality	8	6	4	9	6	3	8	7	7	7	8	8	2	5	9	7
11	intubation and ventilation	7	7	7	5	5	2	7	8	6	6	8	5	6	7	9	7
12	hospitalization	9	9	5	7	5	9	5	5	6	2	5	4	9	7	4	5
13	duration of exacerbation	2	2	5	3	5	3	2	1	2	4	1	4	2	1	1	2
14	dry mouth																

For more information on prioritizing outcomes, please contact:

- Charlie B. Strange III, MD (Chair, DDIC), [strangecc@musc.edu](mailto:strangecc@musc.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Documents & Patient Education), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)

## **Chapter 6:**

# ***Evidence synthesis***

Once the guideline development committee has selected the guideline's clinical questions and outcomes, the methodologist leads the evidence synthesis. This consists of designing a search strategy, determining study selection criteria, selecting pertinent studies, extracting data from the selected studies, pooling the data via meta-analysis, and constructing both summary of findings and evidence tables. The evidence tables include the quality of evidence for each outcome of interest. Generally speaking, evidence synthesis is the most labor intensive and time consuming part of guideline development and may take nine months or more to complete.

While the evidence synthesis is largely performed by the methodologist, he or she will need to reach out to the co-chairs or full committee for help periodically. For example, when designing a search strategy, the methodologist may need the co-chairs' input regarding synonyms, abbreviations, etc. After the completion of study selection, the committee will be asked whether any pertinent evidence has been missed. If potential oversights are confirmed, the search and selection strategy are revised and the search and selection process repeated to see if other important studies may have been similarly missed.

## Search Strategy

The search strategy generally consists of the population, intervention, comparator, and synonyms and abbreviations for each. As a means of saving time and resources, it is suggested that an initial search be conducted to look for relevant, recently published, high-quality systematic reviews. Such systematic reviews can either be a) used as is if it is current or b) updated and then used to inform the guideline. As an example, for the question, “*Should patients with sickle cell disease be treated with hydroxyurea,*” a reasonable initial search strategy is:

Step	Search term
1	Anemia, sickle cell [mh]
2	Sickle cell anemia [tw]
3	Sickle cell disease [tw]
4	#1 OR #2 OR #3
5	hydroxyurea [mh]
6	hydroxyurea [tw]
7	#5 OR #6
8	systematic review [tiab]
9	meta-analysis [pt]
10	meta-analysis [tiab]
11	systematic literature review [tiab]
12	review [pt]
13	Cochrane Database Syst Rev [ta]
14	ACP journal club [ta]
15	#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14
16	#4 AND #7 AND #15

If no relevant, recently published, high-quality systematic reviews are identified, the search is repeated looking for randomized trials:

Step	Search term
1	Anemia, sickle cell [mh]
2	Sickle cell anemia [tw]
3	Sickle cell disease [tw]
4	#1 OR #2 OR #3
5	hydroxyurea [mh]
6	hydroxyurea [tw]
7	#5 OR #6
8	randomized controlled trial [pt]
9	controlled clinical trial [pt]
10	clinical trial [pt]
11	randomized [tiab]
12	randomly [tiab]
13	controlled [tiab]
14	trial [tiab]
15	Placebo [tiab]
16	#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15
17	#4 AND #7 AND #16

If no randomized trials are identified, the search is repeated looking for observational studies (i.e., prospective and retrospective cohort studies with control groups, case-control studies).

Frequently, no relevant systematic reviews, randomized trials, or observational studies are identified. Several options exist at this point, which should be discussed among the methodologist, co-chairs, and possibly the full committee. The options include: a) no further searching with the intention of making a recommendation for future research, b) searching for lower quality evidence such as case series, case reports, clinical observations, etc., with the intention of making a recommendation that acknowledges the poor evidence upon which the recommendation is based, or c) broadening the population or intervention in the search, with the intention of using indirect evidence as the basis of judgments.

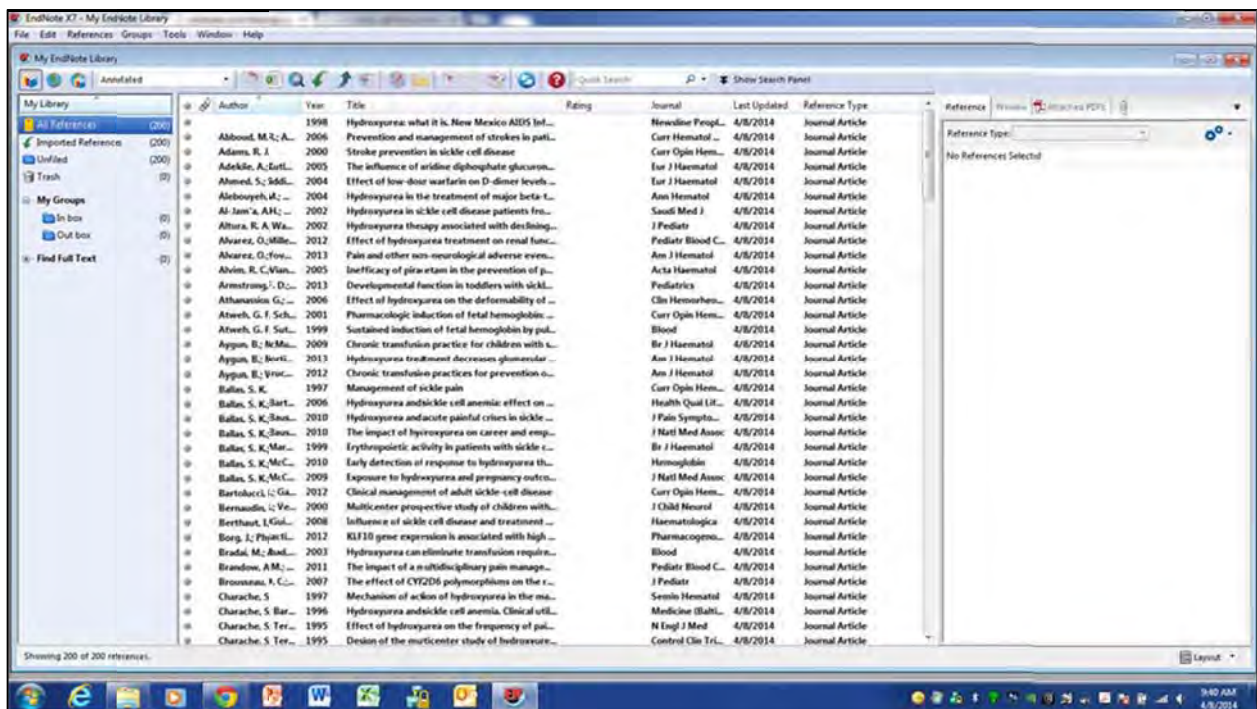
Both the search strategies and search results should be carefully documented and included in the final guideline, generally within an online supplement:

Step	Search term	Results
1	Anemia, sickle cell [mh]	17,507
2	Sickle cell anemia [tw]	17,444
3	Sickle cell disease [tw]	9,414
4	#1 OR #2 OR #3	20,404
5	hydroxyurea [mh]	7,175
6	hydroxyurea [tw]	10,101
7	#5 OR #6	10,101
8	randomized controlled trial [pt]	362,701
9	controlled clinical trial [pt]	87,154
10	clinical trial [pt]	743,778
11	randomized [tiab]	312,261
12	randomly [tiab]	212,145
13	controlled [tiab]	481,531
14	trial [tiab]	356,673
15	Placebo [tiab]	156,375
16	#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15	1,429,629
17	#4 AND #7 AND #16	225

A full systematic review requires that searches be conducted in multiple databases. The most common databases searched are Medline (searched using either the PubMed or Ovid search engine), EMBASE, CINAHL (Cumulative Index to Nursing and Allied Health Literature), CENTRAL (Cochrane Central Register of Controlled Trials), and the Cochrane Database of Systematic Reviews. In contrast, a pragmatic evidence synthesis only requires that at least one database be systematically searched. While pragmatic evidence syntheses are generally sufficient for guideline development, full systematic reviews are necessary if the evidence synthesis is going to be published as an independent manuscript.

# Study Selection

The results of the search should be downloaded into reference management software. EndNote seems to be most common, but other reasonable options include Reference Manager (RefMan), RefWorks, Mendeley, Google Drive, Dropbox, Box.net, and Distiller.



Study selection criteria should be defined prior to reviewing the search results. Criteria generally include a patient population, intervention, comparator(s), outcomes, and study design. As an example, for the searches described above, reasonable study selection criteria are: Randomized trials that included sickle cell patients who received hydroxyurea, compared such patients with those who received either placebo or no treatment, and measured mortality, frequency of

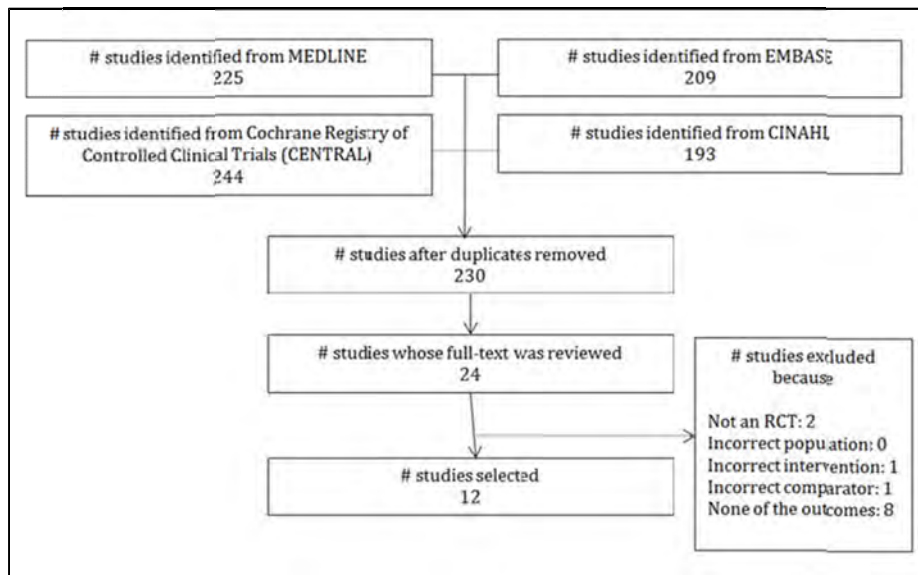
vasoocclusive crises, frequency of acute chest syndrome, and/or bone marrow suppression.

Once the study selection criteria have been determined, the search results are screened on the basis of the titles and abstracts alone. Screening may be performed by the methodologist or divided among members of the guideline development committee. If a pragmatic evidence synthesis is being performed only one screener is necessary; however, screening should be performed in pairs if a full systematic review is being performed. During screening, references that clearly do not meet the study selection criteria are discarded to an “Outbox” within the reference management software. The pace of screening is typically rapid, covering several hundred articles per hour.

Full text articles are then obtained and reviewed in detail for the titles that remain after the initial screening. Those that meet the selection criteria are moved to an “Inbox” within the reference management software, while those that do not meet the selection criteria are moved to the “Outbox”. Information should be recorded about why the articles whose full text was reviewed were either included or excluded:

	Study 1	Study 2	Study 3
1			
2	Smith	Jones	Johnson
3	2005	2010	2009
4	187987	213498	118787
5	Eur J Haematol	Blood	Ann Intern Med
6			
7	Randomized trial	Randomized trial	Randomized trial
8	Sickle cell disease	Sickle cell disease	Sickle cell disease
9	Novel agent B 288	Hydroxyurea	Hydroxyurea
10	Placebo	Placebo	Chronic transfusions
11			
12		X	
13	X	X	X
14	X	X	X
15		X	X
16			
17	Excluded	Included	Excluded
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			

The flow of information during the evidence synthesis should be carefully documented and included in the final guideline, generally within an online supplement:





## ***Pooling the evidence***

Judgments about interventions are based upon the body of evidence, not individual studies. Therefore, it is ideal if the data from multiple studies are pooled into a single estimate of the effect of the intervention on each outcome. This facilitates decision making. As an example, if you pool ten randomized trials and determine that hydroxyurea reduces mortality 3%, reduces acute chest syndrome 10%, and causes bone marrow suppression in 20% of patients, it is easier to balance these competing outcomes to derive a final recommendation than if you have ten separate estimates of the effect of hydroxyurea on each outcome. This is the rationale for pooling data whenever possible.

Pooling data from multiple studies into a single estimate of the effect begins by extracting crude data from the individual studies. A sample data extraction form can be downloaded from the ATS' methodological tools webpage (<http://www.thoracic.org/statements/document-development/methodological-tools.php>). The type of information recorded on the data extraction form can be roughly categorized as the numerical data necessary to pool the data into a single estimate of effect (e.g., number of patients and events in the intervention and control groups) and characteristics necessary to rate the quality of evidence (e.g., details about the patient population, whether the study was blinded, how many patients were lost to follow-up, etc.).

Item	Study 1	Study 2	Study 3	Study 4	Study ... n
<b>BASELINE CHARACTERISTICS</b>					
51 Number randomised					
52 Intervention	300	150		599	79
53 Comparison	300	150		595	78
54 Age					
55 Intervention (mean or median)	65	60		62	68
56 Comparison (mean or median)	67	59		62	66
57 Unit (e.g. mean and SD)					
58 Age range (e.g. 22-73)	mean (SD) 18-85	mean (SD) 18-90	mean (SD) 18-80	mean (SD) 18-79	
59 Age inclusion criterion (e.g. older than 18)	older than 18	older than 18	older than 18	older than 18	
60 Male gender					
61 Intervention	55.00%	50.00%		61.00%	92.00%
62 Comparison	50.00%	50.00%		63.00%	90.00%
63 Severity of illness					
64 Name of score (e.g. APACHE, SOFA, ...)	Apache II	Apache II	Apache II	Apache II	
65 Intervention group mean score	25	25		30	29
66 Comparison group mean score	23	27		30	31
67 Study population					
68 Please choose type of patients from the list (e.g. medical, surgical, ...)	Mixed Medical-surgical	Medical	Medical	Medical	
69 OUTCOMES					
70 Dichotomous outcome 1 - Mortality (all cause)					
71 Are the data available?	Data available	Data available	Data available	Data available	
72 Location or duration of follow-up (choose from the list)	Hospital	30 day	ICU	Hospital	
73 Intervention group: # with event	15		5	24	7
74 Intervention group: Total	300	150		599	79
75 Comparison group: # with event	20		8	32	10
76 Comparison group: Total	300	150		595	78
77 Blinding (patients) (only relevant for RCTs)	probably no	probably no	probably no	probably no	
78 Blinding (personnel) (only relevant for RCTs)	yes	yes	yes	yes	
79 Blinding (outcome assessors) (only relevant for RCTs)	yes	yes	yes	yes	
80 Blinding (data collectors) (only relevant for RCTs)	probably no	probably no	probably yes	no	
81 Blinding (analysts) (only relevant for RCTs)					
raw data					

Meta-analysis is the method used to combine the results of multiple studies. It is performed by inserting the data that was recorded on the data extraction sheet into meta-analysis software.

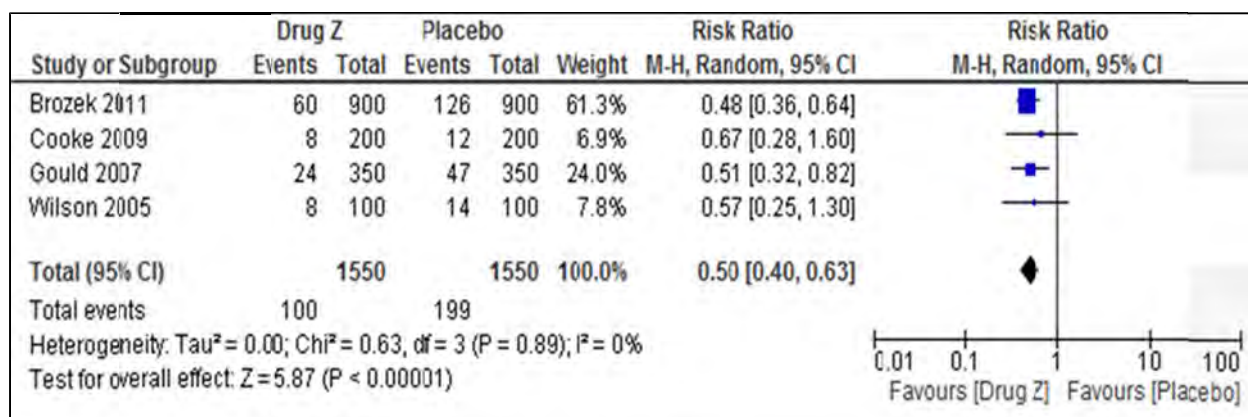
RevMan is a common meta-analysis program that can be downloaded for free from the Cochrane Collaboration's website, <http://tech.cochrane.org/revman/download>. A User's Guide is available at,

[http://tech.cochrane.org/sites/tech.cochrane.org/files/uploads/documents/revman/RevMan\\_5.2\\_User\\_Guide.pdf](http://tech.cochrane.org/sites/tech.cochrane.org/files/uploads/documents/revman/RevMan_5.2_User_Guide.pdf).

The result of a meta-analysis is the estimated effect of an intervention on an outcome, with 95% confidence intervals. There is no right or wrong way to save or display the results of meta-analyses; however, it is convenient to store or display the results for many outcomes as a table:

Outcome or Subgroup	Studies	Participants	Statistical Method	Effect Estimate
1.1 Asthma exacerbation (mostly need for doubling ICS or systemic steroid)	6	3116	Risk Ratio (M-H, Random, 95% CI)	0.72 [0.59, 0.86]
1.1.1 Adults and adolescents	7	2730	Risk Ratio (M-H, Random, 95% CI)	0.74 [0.60, 0.92]
1.1.2 Children	1	386	Risk Ratio (M-H, Random, 95% CI)	0.62 [0.48, 0.80]
1.2 Quality of life (AQLQ)	2	1257	Mean Difference (IV, Random, 95% CI)	Not estimable
1.2.1 Adults and adolescents	2	1257	Mean Difference (IV, Random, 95% CI)	Not estimable
1.2.2 Children	0	0	Mean Difference (IV, Random, 95% CI)	Not estimable
1.3 Quality of life (AQLQ), improvement of $\geq 0.5$ point	5	1886	Risk Ratio (M-H, Random, 95% CI)	1.19 [1.08, 1.30]
1.3.1 Adults and adolescents	4	1886	Risk Ratio (M-H, Random, 95% CI)	1.19 [1.08, 1.30]
1.3.2 Children	1	0	Risk Ratio (M-H, Random, 95% CI)	Not estimable
1.4 Asthma control (ACQ)	3	342	Mean Difference (IV, Random, 95% CI)	-0.87 [-1.14, -0.60]
1.4.1 Adults and adolescents	1	342	Mean Difference (IV, Random, 95% CI)	-0.87 [-1.14, -0.60]
1.4.2 Children	2	0	Mean Difference (IV, Random, 95% CI)	Not estimable
1.5 Need for systemic corticosteroid	5	2214	Risk Ratio (M-H, Random, 95% CI)	0.73 [0.61, 0.87]
1.5.1 Adults and adolescents	4	1828	Risk Ratio (M-H, Random, 95% CI)	0.73 [0.56, 0.94]
1.5.2 Children	1	386	Risk Ratio (M-H, Random, 95% CI)	0.68 [0.51, 0.89]
1.6 Daily dose of oral corticosteroids	3	95	Mean Difference (IV, Random, 95% CI)	Not estimable
1.6.1 Adults and adolescents	1	95	Mean Difference (IV, Random, 95% CI)	Not estimable
1.6.2 Children	2	0	Mean Difference (IV, Random, 95% CI)	Not estimable

And to display the results for one outcome as a Forest Plot:



## Summary of Findings and Quality Assessment Tables

“Summary of findings tables” provide the number of studies, number of patients in the intervention and control groups, number of events in the intervention and control groups, absolute effect, and relative effect for each outcome of an intervention. “Quality assessment tables” provide the number of studies, study design, assessment of the criteria for quality of evidence, and quality of evidence ratings for each outcome of an intervention. These tables collectively summarize the body of evidence and are helpful when the committee formulates its recommendations based upon the evidence. GradePro software combines summary of findings and quality assessment tables into a single table and is available for free download at <http://tech.cochrane.org/revman/other-resources/gradepro/download:>

Quality assessment							Summary of findings					Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	No of patients		Effect		Quality	
							Indefinite anticoagulant therapy	control	Relative (95% CI)	Absolute		
<b>Mortality</b>												
4	randomized trials	no serious limitations	no serious inconsistency	serious <sup>1</sup>	serious <sup>2</sup>	none	16/570 (2.8%)	28/550 (5.1%)	RR 0.56 (0.31 to 1.01)	23 fewer per 1000 (from 47 fewer to 1 more)	⊕⊕⊕⊕ LOW	CRITICAL
<b>Recurrent venous thromboembolism</b>												
4	randomized trials	no serious limitations	no serious inconsistency	serious <sup>1</sup>	no serious imprecision	very strong association <sup>3</sup>	20/570 (3.5%)	95/550 (17.3%)	RR 0.16 (0.06 to 0.41)	139 fewer per 1000 (from 174 fewer to 103 fewer)	⊕⊕⊕⊕ MODERATE	IMPORTANT
<b>Major bleeding</b>												
4	randomized trials	no serious limitations	no serious inconsistency	serious <sup>1</sup>	no serious imprecision	strong association <sup>4</sup>	19/570 (3.3%)	5/550 (0.9%)	RR 3.14 (1.27 to 7.77)	24 more per 1000 (from 7 more to 43 more)	⊕⊕⊕⊕ MODERATE	IMPORTANT

In the sample table above, the “Importance” column displays the priority assigned to each outcome as described in Chapter 5, the “Number of patients” columns show the frequency of events in the intervention and control groups for each outcome, and the “Effect” columns show the absolute and relative effects of the intervention on each outcome; the data for these

columns derive from the meta-analyses. The “Quality assessment” columns display the committee’s assessment of the criteria for rating the quality of evidence:

- Limitations: Limitations are factors that introduce bias into a study, including lack of concealment (blinding of the randomization process), lack of blinding (patient, caregiver, or assessor blinding), subjective outcomes, a large loss to follow-up, early termination of a trial due to benefit, failure to follow the intention-to-treat approach, baseline differences, and selection bias. If the body of evidence includes numerous studies with a risk for bias, the committee may choose to downgrade one level from “no limitations” to “serious limitations”. In contrast, if the risk for bias is extreme, the committee may elect to downgrade two levels to “very serious limitations”.
- Inconsistency: Inconsistency exists when there is substantial variation in the direction or size of the effect across studies. For example, inconsistency is present if some studies found a benefit and others found harm, some studies demonstrated an effect and others did not, or some studies found a large effect and others found a small effect. Inconsistency is often based on the committee’s general impression, but objective measures from the meta-analysis can also be used, including the  $I^2$  test ( $I^2 > 25\%$  suggests inconsistency) and the p-value for heterogeneity (p-value  $< 0.05$  suggests inconsistency). If the body of evidence is judged inconsistent, the committee may choose to downgrade one level from “no inconsistency” to “serious inconsistency”. In

contrast, if the inconsistency is severe, the committee may elect to downgrade two levels to “very serious inconsistency”.

- Indirectness: Indirectness exists when the population, intervention, comparator, or outcome of the clinical question differ from that in the studies. As examples, indirectness of the population exists if your question is about pneumococcal vaccination in the elderly but the relevant studies were conducted in adults of all ages, while indirectness of the intervention exists if your question is about static resistance training for pulmonary rehabilitation but the relevant studies looked at dynamic resistance training. If the body of evidence is significantly indirect compared with the clinical question, the committee may choose to downgrade one level from “no indirectness” to “serious indirectness”. In contrast, if the indirectness is more severe, the committee may elect to downgrade two levels to “very serious indirectness”. If indirectness is extreme, the committee may determine that the evidence is not relevant at all.
- Imprecision: To determine whether or not imprecision exists, the committee must first determine the threshold effect that warrants a change in clinical practice. The degree of benefit that warrants the initiation of therapy is smaller if the benefit is important, side effects are infrequent, and/or cost is low, than if the benefit is minor, side effects are major, and/or cost is high. Imprecision exists if the ends of the confidence interval, as determined by the meta-analysis, lead to different clinical decisions. In other words, imprecision indicates that the trial had too few events to definitively answer the clinical

question. However, imprecision is not synonymous with lack of statistical significance. For example, a statistically significant odds ratio can still be imprecise if the upper limit is large in magnitude but the lower limit is close to one. If the body of evidence is judged imprecise, the committee may choose to downgrade one level from “no imprecision” to “serious imprecision”. In contrast, if the imprecision is severe, the committee may elect to downgrade two levels to “very serious imprecision”.

- Other considerations: There are other criteria that also affect the quality of evidence. The quality of evidence may be downgraded for reporting bias (i.e., publication bias, lag bias, and selective outcome reporting bias). A large magnitude of effect warrants upgrading the quality of evidence one (2-5 times) or two (>5 times) levels, just as a dose-effect relationship warrants upgrading. Finally, if inclusion of all conceivable (but unmeasured) confounders would have biased the trial toward a more extreme effect (e.g., all confounders would have biased the trial toward an increased effect, but the trial found no effect), upgrading is justified.

Finally, the “Quality” column displays the quality of evidence for each outcome. The ATS uses the Grading, Recommendations, Assessment, Development, and Evaluation (GRADE) approach to rate the quality of evidence as high, moderate, low, or very low. The quality of evidence indicates the confidence that the committee has in the estimated effects. To determine the quality of evidence, a baseline assumption is made. Randomized trials are assumed to be high quality evidence as a starting point, while observational studies with control groups (retrospective and prospective cohort studies, case-control studies) are assumed to start as low

quality evidence, and case reports, case series, and clinical observations are assumed to be very low quality evidence. Once the underlying assumption is made, the final quality of evidence is based upon the number of criteria for which the quality of evidence was downgraded or upgraded. As examples, if you have randomized trials with serious indirectness of the population, you would downgrade from high quality evidence to moderate quality evidence. If you have randomized trials with serious limitations and serious indirectness of the population, you would downgrade to low quality evidence. And, if you have observational studies with a large effect, you would upgrade from low quality evidence to moderate quality evidence.

For more information on evidence synthesis or rating the quality of evidence, please contact:

- Charlie B. Strange III, MD (Chair, DDIC), [strangec@musc.edu](mailto:strangec@musc.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Documents & Patient Education), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)

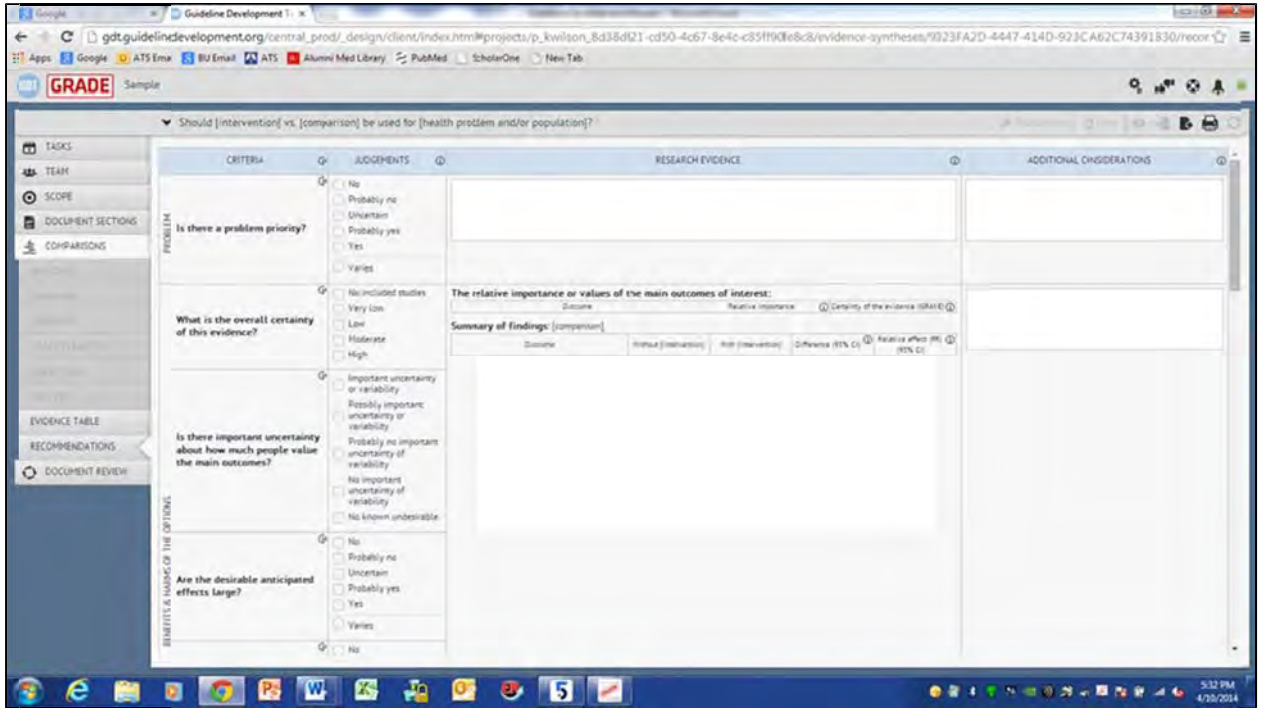


# **Chapter 7:**

## ***Formulating Recommendations***

The formulation of recommendations to answer each clinical question often begins during a face-to-face meeting of the full guideline development committee. During the meeting, participants' potential COI are disclosed and then the evidence for each clinical question is summarized. The following criteria are then used to determine whether to recommend for or against an intervention: 1) the balance of desirable consequences (i.e., benefits) versus undesirable consequences (i.e., harms and burdens), 2) the overall quality of evidence (i.e., the lowest quality of evidence among the critical outcomes), 3) patient values and preferences, and 4) costs and resource use. Based upon these considerations, the guideline developers choose to recommend for or against the intervention. If consensus on a recommendation cannot be achieved by discussion, voting may be necessary. The results of voting should be recorded and included in the guidelines, usually in an online supplement.

Use of the Evidence to Recommendations (EtR) framework may help guideline developers formulate recommendations. The online program asks guideline developers a series of questions related to the criteria described above, insuring that important considerations are not missed. Free access is available at <http://www.guidelinedevelopment.org/>.



The questions asked in the EtR framework include the following:

1. Is the clinical problem a high priority?
2. What is the overall quality of evidence?
3. How much uncertainty is there about how much patients value the main outcome?
4. Are the desirable anticipated effects large?
5. Are the undesirable anticipated effects small?
6. Are the desirable effects large relative to the undesirable effects?
7. Are the resources required small?
8. Is the incremental cost small relative to the net benefits?
9. What would be the impact on health inequities?
10. Is the option acceptable to key stakeholders?
11. Is the option feasible to implement?

Once these questions are considered and answered, the EtR framework asks guideline developers to decide whether the 1) undesirable consequences *clearly* outweigh the desirable consequences, 2) undesirable consequences *probably* outweigh the desirable consequences, 3) desirable consequences *clearly* outweigh the undesirable consequences, 4) desirable consequences *probably* outweigh the undesirable consequences, or 5) the balance is *uncertain*. Based upon these determinations, the guideline developers choose to recommend for or against the intervention.

For more information on formulating guideline recommendations, please contact:

- Charlie B. Strange III, MD (Chair, DDIC), [strangec@musc.edu](mailto:strangec@musc.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Documents & Patient Education), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)

## **Chapter 8:**

# ***Determining the Strength of Recommendations***

Once the recommendations are formulated, the strength of each recommendation is determined. The strength of the recommendation reflects the degree of certainty that the guideline development committee has that the recommended intervention is correct. A strong recommendation indicates that the committee is certain that the recommended intervention is the right thing to do, while a conditional (i.e., weak) recommendation indicates uncertainty, usually because either the desirable and undesirable consequences are finely balanced or the quality of evidence provides little confidence in that balance. If the strength of the recommendation cannot be determined by discussion and consensus, voting may be required. The results of such voting should be recorded and included in the guidelines, usually in an online supplement.

There are several additional ways to conceptualize the strength of a recommendation. First, a strong recommendation indicates that it is the right thing to do for more than 95% of patients, while a conditional recommendation indicates that it is the right thing to do for more than 50% of patients but not as many as 95%. Second, a strong recommendation says to the clinician, “just do it”, whereas a conditional recommendation says, “slow down, think about it, and

discuss it with the patient.” Finally, a strong recommendation says that the committee would be willing to tell a clinician who does not follow the recommendation, “you did the wrong thing”, and the recommendation would be an appropriate performance measure. A weak recommendation is one that is likely to change with the development of additional evidence.

It is important to realize that strong recommendations should usually be based upon high or moderate quality evidence. The reason is that if you have low or very low confidence in the estimated effects (i.e., low or very low quality evidence), how can you be certain that the desirable effects outweigh the undesirable effects? There are exceptions, of course, but these are infrequent. An example of an exception is an intervention that might improve a very important outcome, has few or only minor side effects, is inexpensive, and without which a poor outcome is inevitable. If tap water might cure lung cancer, then a recommendation for lung cancer patients to drink a glass of water each day would be a reasonable strong recommendation based upon low or very low quality evidence.

For more information on determining the quality of evidence, please contact:

- Charlie B. Strange III, MD (Chair, DDIC), [strange@muscc.edu](mailto:strange@muscc.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Documents & Patient Education), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)

# **Chapter 9:**

## ***Writing Recommendations***

By this point, the committee knows what it is going to recommend and the strength of the recommendations, so it is time to put the recommendations into writing using the Grading, Recommendations, Assessment, Development, and Evaluation (GRADE) approach.

Each recommendation concisely states the population, action, intervention, and comparator. The population is the group for whom the recommendation is intended. The action refers to whether the committee recommends administering or not administering, performing or not performing, etc.; it is preceded by “we recommend” for strong recommendations and “we suggest” is used for conditional (i.e., weak) recommendations. The comparator may be excluded if it is obvious. The written recommendations, therefore, fall into several general formats:

- For patients with X, we suggest Y rather than Z.
- For patients with X, we recommend Y rather than Z.
- For patients with X, we suggest NOT doing Y.
- For patients with X, we recommend NOT doing Y.

where X is the population, Y is the intervention, and Z is the comparator.

Finally, a grade consisting of the strength of the recommendation and quality of evidence follows each written recommendation. The strength of the recommendation is determined as described in Chapter 8. The quality of evidence for a recommendation is the lowest quality of evidence among the critical outcomes for the intervention. As an example, if you have an intervention with three critical outcomes, two of which were determined to be based upon high quality evidence and one of which was determined to be based upon moderate quality evidence, the quality of evidence for the recommendation would be moderate. Determining the quality of evidence for each outcome was described in the Summary of Findings and Evidence Tables section of Chapter 6.

Thus, a completed recommendation may look like the following: *“For patients with COPD whose exertional oxyhemoglobin saturation is less than 88%, we recommend supplemental oxygen therapy (strong recommendation, moderate quality evidence).”*

For more information on writing guideline recommendations, please contact:

- Charlie B. Strange III, MD (Chair, DDIC), [strangec@muscc.edu](mailto:strangec@muscc.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Documents & Patient Education), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)

# **Chapter 10:**

## ***Writing the Guideline Manuscript***

Clinical practice guidelines are documents that answer clinical questions by explicitly linking the related evidence to recommendations as described in the preceding chapters. Descriptions of epidemiology, pathophysiology, etc. are beyond the scope of clinical practice guidelines. The following format is recommended for ATS guidelines:

- Table of Contents
- Abstract (Background, target audience, methods, recommendations, conclusions)
- Overview (An introductory paragraph followed by a bulleted list of key conclusions and recommendations)
- Introduction
- Methods
- How to use these guidelines
- Question 1
  - Description of the evidence and its quality
  - Desirable consequences and their magnitudes
  - Undesirable consequences and their magnitudes
  - Rationale for the recommendation
  - What others are saying



- Recommendation
- Values and preferences
- Remarks
- Repeat same sections for additional questions
- Conclusions
- References

An <4,500 word Executive Summary may be published in the Am J Respir Crit Care Med and an <10,000 word Full-length version may be published on the journal's website. In addition, an online supplement is permitted.

For more information on writing the guideline manuscript, please contact:

- Charlie B. Strange III, MD (Chair, DDIC), [strangec@muscc.edu](mailto:strangec@muscc.edu)
- David J. Feller-Kopman, MD (Vice-chair, DDIC), [dfk@dartmouth.edu](mailto:dfk@dartmouth.edu)
- Kevin Wilson, MD (Chief, Documents & Patient Education), [kwilson@thoracic.org](mailto:kwilson@thoracic.org)