Published in final edited form as:

Ophthalmology. 2009 April ; 116(4): 710–71782. doi:10.1016/j.ophtha.2008.12.065.

For which glaucoma suspects is it appropriate to initiate treatment? Article Type: AAO Meeting Paper Section/Category

Abstract

Objective—Because uncertainty exists about which glaucoma suspects should be treated, we sought to identify glaucoma suspects whom an expert panel could agree would be appropriate or inappropriate to treat.

Design—RAND/UCLA appropriateness method, a well-established procedure to synthesize the scientific literature with expert opinion to resolve uncertainty on a health topic.

Participants—Eleven-member panel composed of recognized international leaders in the field of glaucoma.

Corresponding author: Eric Cheng, MD, MS, VA Greater Los Angeles Healthcare System, B 500, ML127, Los Angeles, CA 90073, eric.cheng@va.gov.

Eric M. Cheng, MD, MS	none
JoAnn A. Giacon, MD,	none
Anne L. Coleman, MD, PhD	none
Brian L. Lee, MD	none
Robert N. Weinreb, MD *	Alcon, Allergan, Merck, Pfizer, Zeiss-Meditec
Sydne J. Newberry, PhD	none
Marika J. Suttrop, MS	none
Matthias Schonlau, PhD	none
Murray Fingeret, OD	Alcon, Allergan, Carl Zeiss Meditec,
Neeru Gupta, MD, PhD *	Alcon, Allergan, Merck, Pfizer,
Roger A. Hitchings, MD *	Alcon, Allergan, Merck, Pfizer, Santen
Henry D. Jampel, MD *	Allergan, Glaukos, Pfizer
Thomas L. Lewis, OD *	
Stefano Miglior, MD	Alcon, Merck, Pfizer, Zeiss, Heidelberg Engineering
Richard P. Mills, MD, MPH *	
Norbert Pfeiffer, MD, PhD *	Alcon, MSD, Pfizer,
Fotis Topouzis, MD *	Alcon, Allergan, Pfizer,
Thomas W. Samuelson, MD *	Alcon, Allergan, Alcon, Pfizer, I Science, Glaukos, Ivantis Denali Medical, QLT Inc

served on the 11-member panel to rate scenarios of glaucoma suspects

Appropriateness of treating glaucoma suspects RAND study group (a list of members of the study group is available at <http://aaojournal.org>)

The American Academy of Ophthalmology, American Academy of Optometry, American Optometric Association, American Glaucoma Society, Canadian Ophthalmological Society, European Glaucoma Society, and European Ophthalmologic Society nominated individuals to serve on our expert panel.

Methods—Based on a systematic review of the literature on potentially important factors to consider when deciding to initiate treatment, we initially created over 1000 scenarios of glaucoma suspects. The panel formally rated the appropriateness of initiating treatment for glaucoma suspects through a two-round modified Delphi method, a technique that preserves the confidentiality of individual panelists' ratings but allows panelists to compare their own ratings to those of the entire panel.

Main Outcome Measures—Final ratings for scenarios were categorized as appropriate, uncertain, or inappropriate to treat according to typical pre-specified statistical criteria previously used in projects using the RAND/UCLA appropriateness method. Tools were developed to help clinicians approximate the panel ratings of glaucoma suspects.

Results—The panel chose age, life expectancy, intraocular pressure (IOP), central corneal thickness, cup/disc ratio, disc size, and family history as the variables to consider when deciding whether to treat glaucoma suspects. Permutations of these variables created 1800 unique scenarios. The panel rated 587 (33%) scenarios as appropriate, 585 (33%) as uncertain, and 628 (35%) as inappropriate for treatment initiation. Analysis of variance determined that IOP had greater impact than any other variable on panel ratings. We created a point system with 96% sensitivity and 93% specificity for predicting panel ratings of appropriateness for a glaucoma suspect.

Conclusions—An expert panel can reach agreement on the appropriateness and inappropriateness of treatment for glaucoma suspects.

Introduction

Glaucoma is a leading cause of irreversible blindness worldwide.¹ Glaucoma suspects are defined as individuals who possess clinical findings or risk factors that indicate an increased likelihood of developing glaucoma.² However, not all glaucoma suspects will develop glaucoma. The Ocular Hypertensive Treatment Study (OHTS) randomized 1,636 individuals with ocular hypertension, a subgroup of glaucoma suspects with elevated intraocular pressure (IOP), to topical glaucoma medication versus observation.³ This study showed that whereas treatment reduces the 5-year risk of developing glaucoma by half (9.9% vs. 4.4%), fewer than 10% of subjects in either group developed glaucoma during the 5-year study period.

No explicit recommendations have been made regarding which glaucoma suspects should be treated. The American Academy of Ophthalmology guidelines on glaucoma suspects state that the decision to begin treatment of the glaucoma suspect is “complex and depends on ocular, systemic, medical, and psychosocial factors.”² A validation study for developing a glaucoma risk calculator acknowledged that factors other than risk of glaucoma development itself — such as a patient's health, life expectancy, and preferences—must be considered in any decision to initiate treatment for glaucoma.⁴

We employed the RAND/UCLA Appropriateness Method (RAM) to identify which glaucoma suspects should and should not receive treatment to delay the onset of glaucoma.⁵ The aim of this consensus method is to determine the extent to which an expert panel agrees about a given issue. The four core features of the RAM are: anonymity, iteration, controlled feedback, and statistical group response (see Table 1).⁶ It is designed to prevent a single individual from dominating a meeting, as is sometimes seen in decision-making by committees and also gives participants the opportunity to change their mind if they see their ratings as an outlier. This method has demonstrated validity and reliability for assessing the appropriateness of a wide variety of medical procedures that lacked a large evidence base.^{7–9} Most importantly, results of this technique have predictive validity. For example, the RAND/UCLA appropriateness method was used to determine appropriate indications for cataract surgery.¹⁰ A subsequent study of patients undergoing cataract surgeries found that at least a 2 Snellen line improvement

in visual acuity occurred in 627 of 701 (89%) patients when the indication for cataract surgery was considered appropriate by the prior expert panel, compared to improvements in visual acuity in 36 of 53 (68%) patients when the indication for cataract surgery was considered uncertain, and in only 5 of 14 (36%) patients when the indication for cataract surgery was considered inappropriate.¹¹

Methods

Overview of RAND/UCLA Appropriateness Method

A brief overview of the RAND/UCLA Appropriateness Method is presented here. Each of these steps is described in more detail below. First, the research team conducted a systematic review of the literature to identify factors to consider when deciding whether to treat glaucoma suspects. Based on this literature review, we chose a number of key variables and generated clinical scenarios based on all permutations of these key variables. An expert panel rated the appropriateness of initiating treatment for each of the scenarios on a risk-benefit scale through a two-round Delphi process. Panelists were given opportunities to revise the selection and value ranges of variables, so that the number of scenarios increased in number between rounds of rating. Final ratings were analyzed to identify glaucoma suspects for which initiating treatment was considered appropriate. Note that this study analyzes the collective judgment of an expert panel on the existing medical literature about whether to treat simulated patients. It does not present data from actual patients.

Systematic Review

In November 2006, we conducted searches in Pubmed and the Web of Science for articles related to glaucoma development, risk factors, treatment issues, quality of life, and glaucoma progression. We also mined pertinent references from these articles. We reviewed the titles of 3314 articles, retrieved 756 articles, and performed detailed data abstraction on 141 articles. The systematic review explored the following questions: risk factors for developing glaucoma, quality of life, implications of glaucoma treatment, and life expectancy.

Expert panel

An 11-member expert panel was appointed, based on the nominations of international professional ophthalmologic and optometric societies and the input of our technical advisory committee (Drs. Coleman, Giacon, Lee, and Weinreb; Dr. Weinreb also served on the 11-member expert panel, but he did not have access to the individual panelist ratings at any time during the study). The nominating organizations were informed that expert panelists would use their own personal judgments in this research study, and that their ratings did not necessarily represent official opinions from the nominating organizations. The expert panelists were selected because of their clinical expertise, professional reputation, and geographic location, and represented both academic and community medicine.

Developing the initial set of scenarios

We iteratively drafted a set of scenarios that included variables identified by our systematic review and the technical advisory committee as being associated with the risk of developing glaucoma. Each scenario represented a homogenous group of glaucoma suspects who might benefit from medical treatment. Glaucoma suspects were broadly defined as individuals at risk for developing glaucoma, but who currently do not meet the clinical definition of glaucoma, that is, glaucomatous optic neuropathy and visual field loss. An initial set of 1080 scenarios included all possible combinations of the following variables: age, IOP, central corneal thickness (CCT), cup-to-disc ratio, an estimate of life expectancy based on the Charlson Index, and two proxy measurements for functional status (driving and ability to self-administer eye

drops). For each variable, the technical advisory committee also recommended the range of values that reflect patients typically seen in clinical practice. The values of IOP ranged from normal to ocular hypertension.

First round ratings

The expert panel members were sent the first-round scenarios and the report of the systematic literature review for the variables included in the scenarios. They were instructed to rate, independently, the appropriateness of treatment for each scenario without discussion or contact with the other panelists, using their own best clinical judgment rather than their perceptions of what other experts might say. A scale of 1–9 was used, where 9 was defined as extremely appropriate (benefits greatly exceed risks), 5 was defined as uncertain (benefits and risks about equal), and 1 was defined as extremely inappropriate (risks greatly exceed benefits). “Appropriate” was defined to mean that the expected health benefit (such as reduction in risk of developing glaucoma, reduction in risk of developing complications from glaucoma, reduction in anxiety while remaining untreated) should exceed the expected negative consequences (such as adverse events, increase in anxiety while under treatment, difficulty or need for help in administering eye drops) by a sufficiently wide margin that initiating treatment is regarded as worthwhile. The definition of appropriateness excluded consideration of costs.

The rating forms contained 30 scenarios per page. For one scenario on each page, we provided the calculated glaucoma risk using a publicly available glaucoma risk calculator^{12, 13} based on the combined data from OHTS and the European Glaucoma Prevention Study (EGPS); this calculator uses the age, IOP, CCT, and cup-to-disc ratio. Because the glaucoma risk calculator also includes Pattern Standard Deviation (PSD), a variable that was not included in our final set of scenarios, we used a value of 1.9 dB (mean value of the OHTS and EGPS participants) for all calculations. We instructed the panelists that these calculations were provided solely as a reference point and that they were not required to take these risk calculations into consideration in their rating decisions.

The panelists mailed their ratings of the initial scenarios back to the research team for processing.

Second-round ratings

The panel convened in Chicago, Illinois on October 11, 2007 for the second-round ratings. Gaps in the literature were discussed, and questions and perceptions were shared among panel members to assist in the second-round ratings. Each panel member was provided with a personalized computer printout that showed the distribution of first-round ratings and a mark for that panelist’s own first-round ratings; however, no panelist could determine how another individual panelist rated a scenario. Presenting first round results in this manner preserved the confidentiality of individual panelists’ ratings but allowed panelists to see and compare their own ratings to those of the entire group. A moderator directed the discussion toward the scenarios for which the first-round ratings had shown considerable disagreement among panelists.

After discussion, the panelists revised the initial set of variables to reflect their concepts of the critical factors in decision-making for glaucoma suspects. The final set of variables and definitions is provided in Table 2. The two proxy measures for functional status were dropped. A three category variable of disc size was added because the panelists stated that the interpretation of the cup-to-disc ratio was ambiguous without it. Family history was also added. In addition, the definition of life expectancy and value ranges for the age and IOP categories were changed. We did not provide glaucoma risk calculations to panelists during the second round. The panelists then confidentially rated the 1800 revised scenarios during this second

round. Panelists returned their second-round ratings to the research team by the close of the face-to-face meeting.

Analysis

We calculated the median rating of each scenario and rounded up to the next highest integer if necessary (for a small number of scenarios, missing ratings resulted in a total of only 10 ratings instead of 11). We adopted definitions of disagreement and appropriateness used in prior research studies employing the RAND/UCLA Appropriateness method.⁵ Adjusting these definitions for an eleven-member panel, we considered ratings for a scenario to be in disagreement when three or more ratings were in the 1–3 range and three or more were in the 7–9 range. Ratings for a scenario were considered appropriate if the median rating was ≥ 7 , and they did not meet criteria for disagreement. Ratings for a scenario were considered to be inappropriate if the median rating was ≤ 3 , and they did not meet criteria for disagreement. Ratings for a scenario were considered uncertain when the median rating was >3 and <7 or when the scenarios met criteria for disagreement. We used the risk calculator described above¹² to generate the 5-year risk of glaucoma for all scenarios, and we compared it to ratings of scenarios.

As in prior analyses of ratings, we constructed an “appropriateness index” to create a continuous measure for further analyses described below.¹⁴ This index is a weighted sum of the median rating and the number 5; the effect of the weighting moves the index away from the median rating and towards 5 (the point at which the benefits and risk are deemed to be equal) as variation increases among the eleven ratings.

We ran analyses of variance (ANOVA) to determine the fraction of variance in the appropriateness index explained by each variable. We also calculated the mean of the appropriateness index for each of the 7 variables. The Bonferroni test for multiple comparisons was used to test for pairwise differences in the means.

Finally, we developed shorthand tools that clinicians can use to predict panel ratings of appropriateness. We first used recursive partitioning, a non-parametric method for exploratory data analysis and summarization to create a tree that shows graphically the best sequence of data splits to determine appropriateness and inappropriateness.¹⁴ We also developed a point system using the adjusted appropriateness index means for each value range of a variable.

Results

For each of the 1800 scenarios, the median ratings, mean absolute deviation from the median, and whether ratings met criteria for disagreement are shown in Appendix 1 (available at <http://aaojournal.org>). To facilitate interpretation of the data, we color coded the scenarios according to the consensus appropriateness of treatment: green for ratings that meet criteria for appropriateness, yellow for uncertain ratings (do not meet criteria for appropriateness or inappropriateness), and red for ratings that meet criteria for inappropriateness (Appendix 2, available at <http://aaojournal.org>).

The summary statistics for the panel ratings are shown in Table 3. By happenstance, approximately one-third of the scenarios met criteria for each of the categories for appropriateness, inappropriateness, and uncertainty. Only 24 (1.3%) scenarios met criteria for disagreement. Therefore, most scenarios categorized as uncertain had median scores in the middle of the scale rather than meeting criteria for disagreement.

The ANOVA results are shown in Table 4. All variables were significantly correlated with the appropriateness index ($p < 0.001$). IOP explained more of the variation in ratings than did any

other variable ($r^2=0.53$). The means for each value range of a variable are significantly different from each other ($p<0.001$) with the exception of the means for the two youngest age groups (<55 and 55–79). The values associated with a higher predicted risk of glaucoma (greater IOP, greater cup/disc ratio, thinner CCT)¹³ are associated with higher mean appropriateness scores, with the notable exception that the oldest age group is associated with lower appropriateness scores even though older age is associated with a higher risk of developing glaucoma according to the risk calculator.

Using recursive partitioning as a shorthand tool to predict appropriateness

Based on the graphical tree generated by recursive partitioning, we produced a simple criteria for appropriateness: “IOP >26 mm Hg and <90% 5-year risk of mortality.” These simple criteria correctly identified 338 of the 587 scenarios rated as appropriate by the expert panel (58% sensitivity) and 1119 of the 1213 scenarios that did meet criteria for appropriateness (92% specificity). Recursive partitioning also produced simple criteria for inappropriateness: “IOP $\leq$ 23 mm Hg and cup/disc ratio <0.9”. These criteria correctly identified 448 of the 628 scenarios that the expert panel rated as inappropriate (71% sensitivity) and 1080 of the 1172 scenarios that did not meet criteria for inappropriateness (92% specificity).

Using a point system based on adjusted appropriateness index means as a shorthand tool to predict appropriateness

We examined the mean appropriateness ratings for each variable to develop a point system for predicting appropriateness (Table 5). Each value range of a variable was assigned points that reflected the distribution of mean scores for that variable. A cutoff of ≥ 7 points in this system identified 561 of the 587 scenarios rated as appropriate by the expert panel (96% sensitivity) and 1129 of the 1213 scenarios that did not meet criteria for appropriateness (93% specificity). A cutoff of ≤ 3 points in this system identified 586 of the 628 scenarios that the expert panel rated as inappropriate (93% sensitivity) and 1128 of the 1172 scenarios that did not meet criteria for inappropriateness (96% specificity).

According to this point system, for example, a 65-year old person (0 points) with normal life expectancy (0 points), IOP of 28 mm Hg (6 points), cup/disc ratio of 0.6 (1 point), CCT of 550 μ m (0 points), family history of glaucoma (1 point), and medium disc size (0 points) has a total of 8 points; based on this point system’s thresholds, this person would meet the panel’s criteria for appropriateness to treat. These characteristics correspond to scenario #134 on page 17 in Appendix 1 (available at <http://aaojournal.org>), and the expert panel rated this scenario as appropriate to treat.

Discussion

Although current guidelines do not provide explicit recommendations for identifying which glaucoma suspects should be treated, this expert panel reached agreement on appropriateness and inappropriateness of treating glaucoma suspects for the majority of scenarios presented to them. Similar to the purpose of guidelines, our intention was to produce a tool for guiding clinicians in their decision-making when they are uncertain on whether to initiate treatment. We did not intend that the tool should govern care when clinicians have already made a decision on whether to treat; clinicians are required to incorporate additional factors such as other ocular measurements and patient preferences that are not included in our scenarios. We encourage clinicians who want to know how a particular glaucoma suspect’s scenario was rated by the expert panel to review the actual ratings for that scenario in Appendices 1 and 2 (available at <http://aaojournal.org>). Of the two shorthand tools we developed, recursive partitioning produced simple rules that are easy to remember, but the point system had far greater sensitivity for the same level of specificity.

While both the panel ratings in our study and the glaucoma risk calculator based on OHTS and EGPS data can guide medical decision-making, we note the following differences. First, the glaucoma risk calculator uses pattern standard deviation (PSD); however, our technical advisory committee stated that clinicians may not know this value for a patient. Therefore, we did not include it in our scenarios of glaucoma suspects. Second, our scenarios included life expectancy, family history, and disc size, variables that are not included in the glaucoma risk calculator. Third, whereas the calculator predicts the 5-year risk of developing glaucoma, there is no consensus on the appropriate level of risk for which to initiate treatment. Our ratings are a direct measure of appropriateness (based on expert opinion) that can be used clinically.

As noted in the results, the values associated with a higher predicted risk of glaucoma are associated with higher mean appropriateness scores, with the exception of age. However, we do not believe that the risk calculator had a strong influence on ratings. First, the risk calculator was only presented for 1/30 of the scenarios in round one and not for any in round two. Second, during the discussion at the face-to-face meeting, it was evident that not all panelists used the calculator. Third, even among those that used the calculator, there was no consensus on the risk threshold considered appropriate to treat. Fourth, when IOP was set at the highest and lowest values, ratings of appropriateness did not correlate with level of calculated glaucoma risk. For example, the panel rated three scenarios (#12, 22, 27 on page 14 of Appendix 1, available at <http://aaojournal.org>) as inappropriate to treat. However, using the variable values in this scenario of Age=62 years, IOP=22 mm Hg, C/D=0.3, CCT=500 μ m, and PSD= 1.9 dB, the calculated 5-year risk of glaucoma is 22.8%. On a post-panel survey, the expert panel members stated that the efficacy of lowering IOP in this range for preventing glaucoma development has not been established (EGPS inclusion criteria required an IOP of ≥ 22 mm Hg, OHTS inclusion criteria required IOP of ≥ 24 mm Hg), that IOP is the only available glaucoma treatment even for those without elevated IOP, and that they might have underestimated the impact of a low CCT. On the other extreme, the panel rated three scenarios (#70, 80, and 90 on page 3 in Appendix 1, available at <http://aaojournal.org>) as appropriate to treat. Using values of Age=50 years, IOP=30 mm Hg, C/D=0.3, CCT=600 μ m, and PSD=1.9dB, the five-year calculated risk of glaucoma is only 6.5%. In this case, it appears that the high IOP value accounted for the panel ratings of appropriateness. On a post-panel survey, the panelists stated that IOP is given a high level of importance because it is the only modifiable glaucoma risk factor. In addition, because younger patients will likely live many multiples of the 5-year periods used for prediction by the glaucoma risk calculator, it is likely that such patients would eventually develop glaucoma. However, some panelists also acknowledged that a common practice of “treating all IOPs 30 and above” may need to be re-examined in a glaucoma suspect who also has a thick CCT or a small cup-to-disc ratio.

Some limitations of our study should be noted. First, prior studies of the RAND/UCLA Appropriateness method showed that panelists who perform the treatment being rated (in this case, initiating treatment of a glaucoma suspect) tend to rate more scenarios as being appropriate for treatment than do panelists who do not perform the treatment.^{15–17} In other words, compared with our panelists, an internist might rate fewer scenarios of glaucoma suspects as appropriate to treat. Another potential limitation is that if the value ranges for some variables (such as CCT) were too broad, the scenarios would not represent a homogenous group of patients. However, further splitting of scenarios might have increased their number beyond what the panel could feasibly discuss and rate. Third, we did not consider type of treatment (medication, laser treatment, surgery) because at the time this research project was designed, the general consensus from the literature and our advisory committee was that medication was the first line treatment for glaucoma suspects. If laser or surgery treatment becomes more accepted as initial treatment or if newer medications are substantially better, then the threshold for initiating treatment may change. Therefore, ratings should be interpreted in the context of the current medications available for glaucoma suspects. Fourth, appropriateness rating

exercises typically exclude cost considerations; because our international panel practice in widely varying healthcare systems that impose different levels of cost-sharing among patients, the need to consider cost might make it much more difficult to focus on clinical risk factors. Finally, we considered potentially important factors including patient race, binocular versus monocular involvement, health status, and retinal nerve fiber layer loss; however, we ultimately did not include these variables to prevent the number of total scenarios from increasing beyond what could feasibly be rated.

We also recognize that some difficulties will be encountered when fitting these scenarios to actual patients. First, the panelists did not provide physiologic definitions for small, medium, and large disc size, so these categories would need to be defined by the clinician. Second, determining which patients might have a short life expectancy (90% risk of death within 5 years) may be difficult without an explicit measure, and life expectancy tools are still being perfected.^{18–20} Although we used the Charlson Index, it does not include patient race/ethnicity and does not consider residence in developing countries – factors that are clearly related to life expectancy.

Future quality-of-care studies might use appropriateness scores to assess overuse and underuse of treatment among glaucoma suspects. After assessing the quality of care for glaucoma suspects (i.e., high quality care being defined as offering treatment when ratings state that it is appropriate to treat and withholding treatment when ratings state that it is inappropriate to treat), researchers can analyze the potential role of factors such as geography, practice setting, insurance status, patient race/ethnicity, and subspecialty of the provider in predicting quality of glaucoma care.

Finally, scenarios for which our panel's determinations of appropriateness were uncertain or were at odds with the risk of developing glaucoma based on existing calculators define ideal candidate populations for future clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was sponsored by Pfizer Inc. Dr. Cheng is also supported by a Career Development Award from NINDS (K23NS058571).

We thank Paul Shekelle, MD, PhD and Robert Brook, MD, ScD for providing their expertise in designing the project; Beth Ann Griffin, PhD for performing glaucoma risk calculations; and Patty Smith and Breanne Johnson for their work in setting up the panel meeting.

References

1. American Academy of Ophthalmology Glaucoma Panel. Primary Open-Angle Glaucoma. San Francisco, CA: American Academy of Ophthalmology; 2005 [Accessed December 5, 2008]. Preferred Practice Pattern; p. 1-26. Available at: <http://one.aao.org/asset.axd?id=2e86ca2e-9db0-43c9-b605-c004a82b6ea5>
2. American Academy of Ophthalmology Glaucoma Panel. San Francisco, CA: American Academy of Ophthalmology; 2005 [Accessed December 5, 2008]. Preferred Practice Pattern. Primary Open-Angle Glaucoma Suspect; p. 1-16. Available at: <http://one.aao.org/asset.axd?id=17bd140e-8f59-408b-be73-5cce8020300f>
3. Kass MA, Heuer DK, Higginbotham EJ, et al. Ocular Hypertension Treatment Study Group. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive

medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;120:701–13. [PubMed: 12049574]

4. Ocular Hypertension Treatment Study Group, European Glaucoma Prevention Study Group. Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. *Ophthalmology* 2007;114:10–9. [PubMed: 17095090]
5. Shekelle P. The appropriateness method. *Med Decis Making* 2004;24:228–31. [PubMed: 15090107]
6. Jones J, Hunter D. Qualitative Research: Consensus methods for medical and health services research. *BMJ* 1995;311:376–80. [PubMed: 7640549]
7. Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998;338:1888–95. [PubMed: 9637810]
8. Hemingway H, Crook AM, Feder G, et al. Underuse of coronary revascularization procedures in patients considered appropriate candidates for revascularization. *N Engl J Med* 2001;344:645–54. [PubMed: 11228280]
9. Merrick NJ, Fink A, Park RE, et al. Derivation of clinical indications for carotid endarterectomy by an expert panel. *Am J Public Health* 1987;77:187–90. [PubMed: 3799858]
10. Lee, PP.; Kamberg, CJ.; Hilborne, LH., et al. Cataract Surgery: A Literature Review and Ratings of Appropriateness and Cruciality. Santa Monica, CA: RAND; 1993 [Accessed December 5, 2008]. p. 42-50. JRA-06. Available at: http://www.rand.org/pubs/joint_reports-health/2006/JRA06.pdf
11. Tobacman JK, Zimmerman B, Lee P, et al. Visual acuity following cataract surgeries in relation to preoperative appropriateness ratings. *Med Decis Making* 2003;23:122–30. [PubMed: 12693874]
12. European Glaucoma Prevention Study (EGPS) Group. Predictive factors for open-angle glaucoma among patients with ocular hypertension in the European Glaucoma Prevention Study. *Ophthalmology* 2007;114:3–9. [PubMed: 17070596]
13. Washington University in St. Louis School of Medicine. Glaucoma 5-year Risk Estimator. Based from the Ocular Hypertensive Treatment Study (OHTS) and the European Glaucoma Prevention Study (EGPS). [Accessed October 31, 2008]. Available at: <http://ohts.wustl.edu/risk/calculator.html>
14. Gale RP, Park RE, Dubois R, et al. Delphi-panel analysis of appropriateness of high-dose chemotherapy and blood cell or bone marrow autotransplants in women with breast cancer. *Clin Transplant* 2000;14:32–41. [PubMed: 10693633]
15. Kahan JP, Park RE, Leape LL, et al. Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Med Care* 1996;34:512–23. [PubMed: 8656718]
16. Herrin J, Etchason JA, Kahan JP, et al. Effect of panel composition on physician ratings of appropriateness of abdominal aortic aneurysm surgery: elucidating differences between multispecialty panel results and specialty society recommendations. *Health Policy* 1997;42:67–81. [PubMed: 10173494]
17. Bernstein SJ, Lazaro P, Fitch K, et al. Effect of specialty and nationality on panel judgments of the appropriateness of coronary revascularization: a pilot study. *Med Care* 2001;39:513–20. [PubMed: 11317099]
18. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol* 2003;56:221–9. [PubMed: 12725876]
19. Krahn MD, Bremner KE, Asaria J, et al. The ten-year rule revisited: accuracy of clinicians' estimates of life expectancy in patients with localized prostate cancer. *Urology* 2002;60:258–63. [PubMed: 12137823]
20. Wilson JR, Clarke MG, Ewings P, et al. The assessment of patient life-expectancy: how accurate are urologists and oncologists? *BJU Int* 2005;95:794–8. [PubMed: 15794785]

Appendix

Appropriateness of treating glaucoma suspects RAND study group

Eric M. Cheng, MD, MS

JoAnn A. Giauconi, MD

Anne L. Coleman, MD, PhD

Brian L. Lee, MD

Robert N. Weinreb, MD*

Sydne J. Newberry, PhD

Marika J. Suttorp, MS

Matthias Schonlau, PhD

Murray Fingeret, OD*

Neeru Gupta, MD, PhD*

Roger A. Hitchings, MD*

Henry D. Jampel, MD*

Thomas L. Lewis, OD, PhD*

Stefano Miglior, MD*

Richard P. Mills, MD, MPH*

Norbert Pfeiffer, MD, PhD*

Fotis Topouzis, MD*

Thomas W. Samuelson, MD*

Author affiliations

Eric M. Cheng, MD, MS, RAND Health, University of California, Los Angeles (UCLA)

JoAnn A. Giaconi, MD, UCLA

Anne L. Coleman, MD, PhD, UCLA

Brian L. Lee, MD, Southern California Kaiser

Robert N. Weinreb, MD*, Hamilton Glaucoma Center, University of California, San

Diego. Shiley Eye Center, La Jolla, CA

Sydne J. Newberry, PhD, RAND Health

Marika J. Suttorp, MS, RAND Health

Matthias Schonlau, PhD, RAND Health

Murray Fingeret, OD, * St. Albans VA Hospital, St. Albans, NY

Neeru Gupta, MD, * PhD, St Michaels Hosp Cardinal Carter Wing, University of

* served on the 11-member panel to rate scenarios of glaucoma suspects

Toronto, Toronto, ON, Canada

Roger A. Hitchings, MD, * Moorfields Eye Hospital, London, England

Henry D. Jampel, MD, * Johns Hopkins University, Baltimore, MD

Thomas L. Lewis, OD, * Pennsylvania College of Optometry, Elkins Park, PA

Stefano Miglior, MD, * Clinica Oculistica, Policlinico di Monza, Monza (MI), Italy

Richard P. Mills, MD, MPH, * Swedish Medical Center, Seattle, WA

Norbert Pfeiffer, MD, PhD, * Universitäts-Augenklinik, University of Mainz, Mainz, Germany

Fotis Topouzis, MD, * Aristotle University of Thessaloniki, Thessaloniki, Greece

Thomas W. Samuelson, MD, * Minnesota Eye Consultants, P.A., Minneapolis, MN

Table 1
Four features of the RAND Appropriateness Method

Feature	Purpose
Anonymity	Individual panelist ratings are made in private, thus limiting the ability of a single individual from dominating a meeting, as is sometimes seen in decision-making by committees.
Iteration	Multiple rounds of ratings are conducted, allowing panelists to change their minds
Controlled feedback	Panelists are shown the ratings from the entire group, so that they can privately determine whether their own ratings are an outlier and should be reconsidered
Statistical group response	All scenarios are analyzed to determine the extent of agreement; in contrast, committees sometimes publicly present only the scenarios with unanimous consensus. In addition, the ratings from each panelist is given equal weight

Table 2

Variable definitions in the final set of 1800 scenarios of glaucoma suspects

Variables used to construct scenarios of glaucoma suspects	Range of values for each variable
Age and life expectancy	Age <55 Age 55–79, life expectancy not short (<90% 5-year risk of death) Age 55–79, short life expectancy (≥90% 5-year risk of death) Age ≥ 80, life expectancy not short (<90% 5-year risk of death) Age ≥ 80, short life expectancy (≥90% 5-year risk of death)
Intraocular pressure	≤19 mm Hg 24 to 26 mm Hg 27 to 29 mm Hg 30 to 34 mm Hg
Cup/disc ratio	<0.5 0.5 to <0.7 0.7 to <0.9 ≥0.9
Central corneal thickness	<520 μm 520–580 μm >580 μm
Family history	present (1 st degree family member with known glaucoma) absent (no 1 st degree family member with known glaucoma)
Disc size	small disc size medium disc size large disc size

Table 3
Summary statistics of ratings of 1800 scenarios of glaucoma suspects

Summary statistics of ratings	
Panel median	5.00
Panel mean	4.85
Mean absolute deviation from median	0.87
Scenarios whose ratings meet criteria for:	
	N (%)
Appropriateness, n (%)	587 (32.6%)
Uncertainty, n (%)	585 (32.5%)
Inappropriateness, n (%)	628 (34.9%)

1–9 scale, where 9 was defined as extremely appropriate [benefits greatly exceed risks], 5 was defined as uncertain [benefits and risks about equal], and 1 was defined as extremely inappropriate [risks greatly exceed benefits].

Ratings for a scenario considered to be in disagreement when three or more ratings were in the 1–3 range and three or more were in the 7–9 range. Ratings for a scenario were considered to be appropriate if the median rating was 6.5, and they did not meet criteria for disagreement. Ratings for a scenario were considered to be inappropriate if the median rating was ≤ 3.0 , and they did not meet criteria for disagreement. Ratings for a scenario were considered to be uncertain when the median rating was >3 and <6.5 or when the scenarios met criteria for disagreement.

Table 4

Analysis of variance (ANOVA) for determining the impact of variables on mean appropriateness index

Variable used to generate scenarios of glaucoma suspects	R ²
Intraocular pressure	0.53
Cup/disc ratio	0.16
Central corneal thickness	0.08
Life expectancy	0.07
Disc size	0.05
Age	0.04
Family history	0.01

All variables are associated with appropriateness index p-value <0.001

Point system for predicting appropriateness of initiating treatment for a glaucoma suspect according to the expert panel

Table 5

Variable	Points for values for a variable			
Age (years)	<55 0 points	55 to 79 0 points	>79 -1 point	
Life expectancy	≥90% 5-year risk of death -2 points <90% 5-year risk of death 0 points			
Intraocular pressure (mm Hg)	≤ 19 0 points	24 to 26 +4 points	27 to 29 +6 points	30 to 34 +8 points
Cup/disc ratio	<0.5 0 points	0.5 to <0.7 +1 point	0.7 to <0.9 +2 points	≥0.9 +4 points
Central corneal thickness (µm)	<520 +1 point	520-580 0 points	>580 -1 point	
Family history (Documented 1 st degree relative)		no 0 points	yes +1 point	
Disc size	small +1 point	medium 0 points	large -1 point	