

# Use of the Delphi process for defining successful outcomes for strabismus surgery



Massimiliano Serafino, MD,<sup>a</sup> David B. Granet, MD, FACS,<sup>b</sup> Burton J. Kushner, MD,<sup>c</sup> Linda R. Dagi, MD,<sup>d</sup> Ramesh Kekunnaya, MD, FRCS,<sup>e</sup> and Paolo Nucci, MD, FEBO<sup>a</sup>

## SUMMARY

The purpose of this review was to identify areas of consensus and disagreement among experts for the definition of success following strabismus surgery using the Delphi process. Three rounds of electronic questionnaires were sent to a panel of 28 strabismus experts. Throughout the process, members of the panel were masked to one another's identities to minimize the possibility of influence among members. Prior to data collection, we defined consensus as an 85% agreement on the answer to each question. Questions for which there was no consensus were reworded, and the resultant new questions were used in each subsequent round of questioning. We arrived at consensus for 23 of the 36 questions (64%). Consensus was obtained for recommending unique criteria for the definition of success for certain specific strabismus conditions. In addition, it was considered important that stereopsis and the range of single binocular vision be included in the definition of success for certain types of strabismus. (J AAPOS 2019;23:309-312)

Successful outcome after strabismus surgery is not well defined in the literature. As previously summarized in detail,<sup>1</sup> some studies define success based on motor alignment, whereas others stress sensory outcome; several combine both sensory and motor results. Concepts such as the range of the single binocular visual field, the need for prisms, or the manipulation of accommodation by reducing the plus power of spectacle lenses might also be considered in the definition of a successful outcome.

The Delphi process<sup>2</sup> is a widely used method for achieving consensus among experts in controversial areas. It provides an adaptable method for gathering and analyzing data on practice patterns based on the views of a panel of experts. Unlike unstructured group discussions, the Delphi process allows for equal input from each participant and avoids undue influence of any individual. In the Delphi process, a group of experts is asked to respond anonymously to two or more rounds of questionnaires.

In each round after the first, the deidentified results of the previous round are revealed to each participant.

Previously the Delphi process has been widely used to delineate guidelines for the treatment of disorders where there is little or no evidence in the published literature.<sup>3-6</sup> The process is well suited for identifying areas for future research in areas of nonconsensus. Recently, Jabri and colleagues<sup>7</sup> stressed the necessity for international agreement in the field of strabismus as to what constitutes a successful outcome after strabismus surgery. The purpose of the present study was to identify areas of consensus and disagreement among experts in the definition of success after strabismus surgery using the Delphi process.

## Methods

The Delphi<sup>2</sup> process was used as a structured, rules-driven communication technique. The two original organizers of this study (MS, DGB) selected an executive committee (EC) with attention to international geographic representation, expertise, and quality of peer-reviewed publications. Three members of the EC (MS, DGB, BJK) were appointed as study leaders. The members of the EC collectively invited 29 potential participants, who were chosen using the following criteria: number and quality of peer-reviewed publications and presentations at international meetings, teaching roles, and surgical experience. Additional considerations included geographic diversity, with an attempt to include representation from every continent (except Antarctica) and proficiency in English to facilitate collaboration. Of the 29 invited, 25 accepted and, along with the 6 members of the EC, comprised the Strabismus Success Definition Delphi Study Group (a full listing appears at the end of this article).

Three rounds of electronic questionnaires were sent to 28 members of the study group. The three study leaders (MS, DG,

*Author affiliations:* <sup>a</sup>Department of Clinical Sciences and Community Health, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Multimedica, Eye Clinic San Giuseppe Hospital, University of Milan, Milan, Italy; <sup>b</sup>Ratner Children's Eye Center of the Shiley Eye Institute and Viterbi Family Department of Ophthalmology, University of California San Diego, La Jolla, California; <sup>c</sup>Department of Ophthalmology & Visual Sciences University of Wisconsin, Madison, Wisconsin; <sup>d</sup>Department of Ophthalmology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts; <sup>e</sup>Child Sight Institute, Jasti V Ramanamma Children's Eye Care Center, L V Prasad Eye Institute, Hyderabad, India

Submitted April 22, 2019.

Revision accepted July 3, 2019.

Published online October 3, 2019.

Correspondence: Massimiliano Serafino, MD, University Eye Clinic, San Giuseppe Hospital, University of Milan, Via San Vittore, 12, 20123, Milan, Italy (email: [massserafino@yahoo.com](mailto:massserafino@yahoo.com)).

Copyright © 2019, American Association for Pediatric Ophthalmology and Strabismus. Published by Elsevier Inc. All rights reserved.

1091-8531/\$36.00

<https://doi.org/10.1016/j.jaaapos.2019.07.006>

BK) did not participate in answering the questions, because they alone were not masked to the identities of the other participants. The EC created multiple-choice questions that focused on identifying which strabismus conditions should or should not have their own separate and unique set of outcome criteria for defining success. The EC believed that questions with categorical outcomes, in most cases yes or no, were optimum. Each resulting questionnaire was then beta-tested by all of the EC members. The final, revised first-round questionnaire was circulated electronically to the group using Google Survey. Before the data were analyzed, consensus was defined by the EC as an 85% agreement on the answer for any given question. In some cases, if a consensus was not reached on a given question on the initial round, the question was reworded for subsequent rounds. Specifically, if the three lead investigators felt that there was any ambiguity in the wording of a question that may have resulted in a lack of consensus, the question was rephrased, subject to approval by the other EC members. For example, consensus was not reached initially for the question, "Should partly accommodative ET have its own separate set of criteria for success?" The question was subsequently reworded to "Should partly accommodative ET have its own unique and separate set of criteria for success?" This modified question did attain consensus. In other situations, questions without consensus were reworked to narrow their focus. For example, the question, "Should cases of torsional diplopia have their own separate set of criteria for success?" did not achieve consensus. By splitting this question into three separate questions asking about the presence of diplopia in the primary position, diplopia in downgaze, and the presence of stereopsis, consensus

was achieved. Because some questions were split, there were more questions on subsequent rounds than on the first. Each participant was given 2 weeks to respond to each rounds of questions. Participants only saw results from the prior rounds of questions in a deidentified form in order to maintain anonymity.

## Results

At the third round of questions, one of participants dropped out of the study, leaving 27 respondents. After three rounds, consensus was obtained for 23 of the 36 questions (64%). For the remaining 13 questions (36%), there was no consensus even after multiple rounds of questioning. The questions and answers for which we obtained consensus are provided in [Table 1](#); the questions and answers for which there was no consensus, in [Table 2](#).

## Discussion

There is no standard accepted definition of success when evaluating or reporting outcomes of strabismus surgery. The Delphi process is a useful tool for identifying consensus among experts when no consensus has emerged though published research. The ideal number of participants in a Delphi process is not uniformly agreed upon, with reported numbers ranging from 10 to 1,685.<sup>8</sup> Some authors highlight the importance of the diversity of panelists,<sup>9</sup> which we addressed in part by selected panelists from diverse geographical regions. Although we arbitrarily defined consensus as 85% of panelists being in agreement

Table 1. Questions with consensus and relative percentages

Questions with consensus	Answers (%)
1. Should SO palsy have its own separate set of criteria for success?	Yes (86)
2. Should 6th nerve palsy have its own separate set of criteria for success?	Yes (86)
3. Should 3rd nerve palsy have its own separate set of criteria for success?	Yes (86)
4. Should the outcome criteria for alignment of an intermittent deviation consider the control?	Yes (89)
5. Should a sensory outcome include evaluation of stereopsis?	Yes (96)
6. Should a sensory outcome include presence of diplopia in primary position?	Yes (93)
7. Should a sensory outcome include presence of diplopia in near downgaze (reading position)?	Yes (86)
8. Should DVD have its own separate set of criteria for success?	Yes (93)
9. Should strabismus associated with Graves disease have its own separate set of criteria for success?	Yes (93)
10. Should Duane syndrome have its own separate set of criteria for success?	Yes (89)
11. Should Brown syndrome have its own separate set of criteria for success?	Yes (86)
12. Should nystagmus (fusional maldevelopment nystagmus excluded) associated with strabismus and AHP (anomalous head posture) have its own separate set of criteria for success?	Yes (96)
13. Should nystagmus (fusional maldevelopment nystagmus excluded) associated with strabismus without AHP have its own separate set of criteria for success?	Yes (89)
14. Should partly accommodative ET have its own unique and separate set of criteria for success?	Yes (89)
15. Should ET with either high AC/A or nonaccommodative convergence excess of 10 <sup>Δ</sup> or more have its own unique and separate set of criteria for success?	Yes (85)
16. Should acquired nonparalytic ET in adults have its own unique and separate set of criteria for success?	Yes (85)
17. Should intermittent XT have its own unique and separate set of criteria for success?	Yes (93)
18. Should torsional diplopia have its own unique and separate set of criteria for success?	Yes (93)
19. Should strabismus fixus associated with high myopia have its own unique and separate set of criteria for success?	Yes (85)
20. Should patients with diplopia have their own unique and separate set of criteria for success?	Yes (89)
21. Should the presence or absence of diplopia in the primary position be included in the evaluation of sensory outcome?	Yes (96)
22. Should the presence or absence of diplopia in the near downgaze reading position be included in the evaluation of sensory outcome?	Yes (89)
23. Should the range of single binocular vision be included in the evaluation of sensory outcome of at least <i>some types</i> of strabismus? (Please note, we are considering a general outcome and not ones for a specific entity.)	Yes (89)

Table 2. Questions without consensus and relative percentages

Questions without consensus	Responses (%)
1. Should infantile ET operated upon prior to 1 year of age have its own unique and separate set of criteria for success?	Yes (70)
2. Should infantile ET operated upon at 1 year of age or older have its own unique and separate set of criteria for success?	Yes (59)
3. Should constant XT have its own unique and separate set of criteria for success?	Yes (81)
4. Do you prefer as the outcome criteria for alignment the manifest deviation only? (Please note, we are considering a general outcome and not one for a specific entity)	Yes (48)
5. Do you prefer as the outcome criteria for alignment the manifest plus latent deviation only? (Please note, we are considering a general outcome and not one for a specific entity)	Yes (30)
6. Do you prefer as the outcome criteria for alignment the manifest, and manifest plus latent, separately? (Please note we are considering a general outcome and not one for a specific entity)	Yes (59)
7. Do you consider cosmetic outcomes differently from functional outcomes? (eg, consider a patient with strabismus and diplopia: a good cosmetic result could be obtained, but diplopia persist.)	Yes (78)
8. Should 4-dot testing be included in the evaluation of sensory outcome of at least some types of strabismus? (Please note, we are not asking what those specific entities would be.)	Yes (74)
9. Should the Bagolini lens testing be included in the evaluation of sensory outcome of at least some types of strabismus? (please note we are considering a general outcome and not one for a specific entity)	Yes (63)
10. How long after surgery do you consider your result sufficiently stable to be used for an outcome determination? (Please consider your answer from a scientific point of view as opposed to intuition; if there are no published data, rely on your personal experience.)	A. 1 week (0) B. 6 weeks (11) C. 6 months (15) D. A combination of one or more short term (eg, 6 weeks and/or 6 months) and then some defined longer-term periods (eg, 1 year, 5 years, etc) (67) E. A combination of one or more short term (eg, 6 weeks and/or 6 months) and then last exam (7)
11. What do you consider as a motor success after surgery? (ie, a final deviation that could be considered sufficiently stable. Consider scientific published data, and if you think there are no data or no reliable data, consider your personal experience. Please note, we are considering a general outcome and not one for a specific entity.)	A. $\pm 10^\Delta$ from straight (22%) B. Up to $10^\Delta$ of undercorrection and no manifest overcorrection (7%) C. Up to $10^\Delta$ of undercorrection and a small manifest overcorrection, eg, $2^\Delta$ (7%) D. Up to $10^\Delta$ of undercorrection and no manifest overcorrection and no more than $10^\Delta$ of latent overcorrection (15%) E. Up to $10^\Delta$ of undercorrection and a small manifest overcorrection, eg, $2^\Delta$ and no more than $10^\Delta$ of latent overcorrection (11) F. No manifest deviation and up to $10^\Delta$ of asymptomatic latent over or undercorrection (15) G. Other (22)
12. Do you feel a patient meets a predefined outcome criteria for success, if obtaining that angle is only obtained by manipulation of accommodation (eg, cutting plus or overminusing to reduce an unsatisfactory outcome)?	Yes (74)
13. Do you feel a patient meets a predefined outcome criteria for success if obtaining that angle is only obtained by the use of ground in prisms?	Yes (30)

on a specific question, there is no agreed-upon standard for the consensus threshold with the Delphi process. Other studies have used  $>50\%$ ,<sup>10</sup>  $\geq 60\%$ ,<sup>11</sup>  $\geq 66\%$ ,<sup>12</sup> or  $>80\%$ <sup>13,14</sup> for defining consensus.

The value of the Delphi process for ultimately achieving consensus by successively refining questions for which

there was initially no consensus is evident from the observation that we attained consensus on 7 of 32 (22%) questions in the first round; 6 of the remaining 25 (24%), in the second round; and 10 of the remaining 24 (42%) questions, in the third round. At the end of the process, we reached consensus in a total of 23 of 36 questions (64%).

We believe that areas of genuine nonconsensus shed light on topics that could benefit from further research. For example, there was no consensus on how long after surgery results be considered sufficiently stable to evaluate outcome, what magnitude of deviation should be considered a motor success after surgery, and what deviation should be considered (eg, the manifest deviation or the manifest plus latent deviation). These questions were formulated to address strabismus in general and not specific entities. The next phase of our study is to ascertain whether consensus can be achieved on a condition-specific basis. For example, perhaps there would be agreement on whether the manifest plus latent deviation should be considered in cases of intermittent exotropia, where there was no consensus on that issue for strabismus in general.

If the strabismus community can agree on what outcomes should be considered a successful result, and if future studies will be based on those agreed-upon criteria, results obtained by different treatment approaches can be more readily compared. We believe that this study is a first step toward achieving that goal.

There are some limitations to this study. By its very nature, the Delphi process achieves its goal by relying on consensus. It does not rely on data from clinical studies, except indirectly to the degree that clinical studies influence participants' opinions. Also, it does not give weight to the opinions of outliers, which may be of value. We recognize that consensus may not be the best way to the "correct" answer; however, compared to the diverse outcome criteria currently used in the strabismus literature, we believe it is a good starting point. Subsequently, if good data show that better criteria than those obtained by the Delphi process should be used, researchers can modify the criteria accordingly. The Delphi process starts with a set of questions and ends up with questions that may be a subset of the original ones. As indicated earlier, this phase of our use of the Delphi process to address strabismus outcomes relied on categorical questions. We anticipate that our follow-up study, which will focus on the specific criteria that should constitute a successful outcome for each condition identified in this present study, will involve questions with continuous or weighted answers. Finally, the EC was self-selected.

In conclusion, we describe the first use of a Delphi process to identify areas of consensus and nonconsensus in the evaluation of outcomes of strabismus surgery. We identified important points for which there was no consensus, including the following: (1) length of time after surgery for determination of success, (2) magnitude of deviation that can be considered consistent with success, and (3) which deviation should be considered for the outcome (manifest, manifest plus latent, or both separately). The ongoing study will strive to achieve consensus for these questions within the context of specific strabismus conditions. This present study did achieve consensus for identifying strabismus conditions that should have their own unique and separate set of outcome criteria for defining

success. It also identified the importance of stereopsis and the range of single binocular vision as being included in the definition of success for some types of strabismus.

## Acknowledgments

*The authors gratefully acknowledge the assistance of Francesco Bon-signore, orthoptist, in formatting the electronic questionnaire.*

**Strabismus Surgery Delphi Process Working Group** (members of the EC are denoted with an asterisk)

*Gillian Adams (United Kingdom), Adedaio Adio (Nigeria), Boon Long Quah (Singapore), Igor Aznaurjan (Russia), Linda Dagi\* (United States), Shuan Dai (New Zealand), Angela Maria Fernandez (Colombia), John Ferris (United Kingdom), Mauro Goldschmit (Brazil), David Granet\* (United States), Richard Hertle (United States), Jonathan Holmes (United States), Jeong-Min Hwang (South Korea), Ramesh Kekunnaya\* (India), Arif Khan (United Arab Emirates), Lionel Kowal (Australia), Burton Kushner\* (United States), Jun Hong Li (Taiwan), Andrea Molinari (Ecuador), Yair Morad (Israel), Samir El-Mulki (Jordan), Paolo Nucci\* (Italy), Scott Olitsky (United States), Seyhan B. Özkan (Turkey), Matthew Piblblad (United States), Stacy Pineles (United States), Richard Saunders (United States), Emin C. Sener (Turkey), Massimiliano Serafino\* (Italy), Meenakshi Swaminathan (India), Federico Velez (United States).*

## References

1. Kushner BJ, Fisher M. Is alignment within 8 prism diopters of orthotropia a successful outcome for infantile esotropia surgery? *Arch Ophthalmol* 1996;114:176-80.
2. Hsu C, Stanford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval* 2007;12:1-8.
3. Armon K, Lakhnpaul M, Stephenson T. An evidence and consensus based guideline for acute diarrhoea management. *Arch Dis Child* 2002;86:138.
4. Campbell SM, Hann M, Roland MO, Quayle JA, Shekelle PG. The effect of panel membership and feedback on ratings in a two-round Delphi survey: results of a randomized controlled trial. *Med Care* 1999;37:964-8.
5. Washington DL, Bernstein SJ, Kahan JP, Leape LL, Kamberg CJ, Shekelle PG. Reliability of clinical guideline development using mail-only versus in-person expert panels. *Med Care* 2003;41:1374-81.
6. Mathis R, Doyle S. A quality mix: using evidence and experience to evaluate new technologies. *J Health Qual* 2003;25:4-6. quiz 6-7.
7. Al Jabri S, Kirkham J, Rowe FJ. Development of a core outcome set for amblyopia, strabismus and ocular motility disorders: a review to identify outcome measures. *BMC Ophthalmol* 2019;19:47.
8. Powell C. The Delphi technique: myths and realities. *J Adv Nurs* 2003;41:376-82.
9. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;2:1-88. i-iv.
10. Evans C. The use of consensus methods and expert panels in pharmaco-economic studies. *Practical applications and methodological shortcomings.* *Pharmacoeconomics* 1997;12:121-9.
11. Verkerk K, Van Veenendaal H, Severens JL, Hendriks EJ, Burgers JS. Considered judgment in evidence-based guideline development. *Int J Qual Health Care* 2006;18:365-9.
12. Behrens A, Doyle JJ, Stern L, et al. Dysfunctional tear syndrome study group. Dysfunctional tear syndrome: a Delphi approach to treatment recommendations. *Cornea* 2006;25:900-907.
13. Morris CJ, Cantrill JA. Preventing drug-related morbidity—the development of quality indicators. *J Clin Pharm Ther* 2003;28:295-305.
14. Serafino M, Trivedi RH, Levin AV, et al. Use of the Delphi process in paediatric cataract management. *Br J Ophthalmol* 2016;100:611-15.