Use of Patient-Reported Outcomes in Medical Product Development: A Report from the 2009 NEI/FDA Clinical Trial Endpoints Symposium

Rohit Varma,[1] *Elaine A. Richman,*[2] *Frederick L. Ferris III,*[3] *and Neil M. Bressler*[4]

On October 13, 2009, the third in a series of meetings to discuss endpoints in ophthalmic clinical trials was held at the National Institutes of Health (NIH). This event, about patient-reported outcomes (PROs), was an opportunity for the vision community and the U.S. Food and Drug Administration (FDA) to meet and discuss the FDA requirements for adding new endpoints to the evaluation of ophthalmic treatments and products. The meeting was attended by researchers, clinicians, policymakers, and representatives from industry and vision associations. The first two meetings, in 2006 and 2008, were held to discuss endpoints and clinical trial strategies for evaluating new treatments for age-related macular degeneration (AMD) and diabetic retinopathy and glaucoma treatments, respectively.[1,2]

The objectives of this conference were to provide definitions of PROs, to describe the importance of PROs, to assess what is known and what should be known about PROs in ophthalmology, to provide insights on how the FDA evaluates development and validation of PRO instruments, and to report on clinical trial design issues relevant to PROs. An ancillary issue was the use of PROs in labeling claims and patient-information materials.

The meeting was organized by the Association for Research in Vision and Ophthalmology (ARVO) and co-chaired by Rohit Varma, MD, PhD, Frederick Ferris, MD, and Neil Bressler, MD. Representing the FDA were Laurie Burke, RPh, PhD, Wiley Chambers, MD, Malvina Eydelman, MD, Danica Marinac-Dabic, MD, PhD, and Päivi Miskala. All participants and their affiliations are shown in the box on page 6096).

Growing evidence indicates the importance of vision-related PROs in clinical trials for evaluating medical drugs and devices, not only for medical product labeling, but also to expand the understanding of clinical trial outcomes.[3–7] Although the FDA has incorporated labeling for PROs in areas outside of ophthalmology, the issues and challenges relevant to ophthalmology are just beginning to be understood.

Most clinical trials in ophthalmology use visual acuity measured by an eye chart to assess changes in vision related to experimental conditions. However, as many clinicians and researchers are aware, objective measurement of visual acuity (and/or visual field) may not adequately describe the total impact of a treatment on a patient's visual world.

What Is a PRO?

A PRO is a measurement of any aspect of a patient's health status that is reported directly by the patient, free of interpretation by a physician, researcher, or other person. It is an account of how the patient functions or feels relative to a health condition or therapy. A PRO would measure any of the following:

- Symptoms
- Symptom impact and functioning
- Disability or handicap
- Adverse events
- Treatment tolerability
- Treatment satisfaction
- Health-related quality of life

The European Committee for Medicinal Products for Human Use also defines PROs:

> *Any outcome based on a patient's perception of a disease and its treatment(s) scored by the patient himself is called a Patient-Reported Outcome (PRO). PROs are a large set of patient-assessed measures ranging from single item (e.g., pain VAS, overall treatment evaluation, and clinical global improvement) to multi-item tools. Multi-item tools can be unidimensional measuring a single well-defined concept such as specific physical functioning or multi-dimensional questionnaires measuring broad concepts such as psychological function or health-related quality of life (HRQOL). In general terms, PROs provide information on the patient's perspective of a disease and its treatment.*[8]

Susan Vitale, PhD, MHS, describing PROs as a way to measure the impact of an intervention on aspects of patients' health (e.g., signs/symptoms, ability to perform activities of daily living, and psychological functioning), also emphasized the importance of distinguishing PROs from quality of life (QOL) and health-related quality of life (HRQOL). PRO is an umbrella term that includes both QOL and HRQOL.

A PRO instrument intended to measure treatment efficacy and to support medical product labeling claims generally requires development and validation. Not all PROs are useful for evaluating treatment efficacy for all conditions. The FDA assesses the instrument's appropriateness and competence to evaluate treatment efficacy as a key trial endpoint on the basis of a review of the clinical trial protocol and analysis plan, the targeted labeling claims, the instrument's content validity documentation, an evaluation of the other measurement properties (e.g., construct validity, reliability, and ability to detect

From the [1]Doheny Eye Institute and the Department of Ophthalmology, Keck School of Medicine, University of Southern California, Los Angeles, California; [2]Richman Associates, LLC, Baltimore, Maryland; the [3]National Eye Institute, National Institutes of Health, Bethesda, Maryland; and [4]Wilmer Eye Institute, The Johns Hopkins University, Baltimore, Maryland.

Supported by National Eye Institute Grant EY11753 and the National Center on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD.

Submitted for publication April 2, 2010; revised August 30, 2010; accepted September 27, 2010.

Corresponding author: Rohit Varma, MD, MPH, Doheny Eye Institute, Suite 4803, 1450 San Pablo Street, Los Angeles, CA 90033; rvarma@usc.edu.

DOI:10.1167/iovs.10-5627
Investigative Ophthalmology & Visual Science, December 2010, Vol. 51, No. 12
Copyright © Association for Research in Vision and Ophthalmology

change), and translation and cultural adaptation documentation for multinational studies.

Development of PRO Instruments and Validation Issues

The FDA realizes the value of the patient perspective when evaluating treatment benefit of a medical product. To support a labeling claim measured by a PRO instrument, the FDA would hold the PRO instrument to the same standards as any other labeling claim. According to FDA guidance, "…the determination of whether the PRO instrument supports an effectiveness endpoint includes an assessment of the ability of the PRO instrument to measure the claimed treatment benefit and is specific to the intended population and to the characteristics of the condition or disease treated."[9] PROs that are intended as key clinical trial endpoints to determine treatment efficacy and to support medical product labeling claims must be well-defined and reliable.

A claim, by FDA definition, is a statement of treatment benefit or comparative safety advantage. A *claim* can appear in any section of a medical product's FDA-approved label or in advertising or promotion of medical products.

A *treatment benefit* refers to the impact of treatment on a patient's survival, functioning, and/or on how the patient feels. Treatment benefit can be supported by comparing treatment efficacy (e.g., improvement of symptoms) or safety (e.g., delay in onset of treatment-related toxicity).

Päivi Miskala, MSPH, PhD, described some of the most salient features in the development and validation of PRO instruments. For labeling, as an example:

- The label must contain a summary of the essential information needed for safe and effective use of the drug or device.
- The information must be informative and accurate.
- The label must be updated as new information becomes available.
- It should contain no implied claims that have not been substantiated by the clinical development program for the particular medical product.

The FDA recognizes that some treatment effects are known only to patients and that improvement in clinical measures does not always translate into improvements in how a patient feels or functions. PROs have been and are being used as key endpoints in FDA-reviewed medical product trials, and they have been and are being used as a primary basis for a medical product approval in some clinical areas when appropriate.

A PRO measure to determine treatment efficacy requires instrument development and validation. Rating scales should be developed based on a hypothesis in which the numbers are shown empirically to quantify a specific concept in the target population. The concepts that are most often used to support labeling claims relate to patient's symptoms, signs, or an aspect of functioning directly linked to disease status.

PRO instruments that are intended as key trial endpoints require development and validation. The 2009 FDA Guidance for Industry describes how the FDA reviews and evaluates PRO instruments. This guidance outlines the specific instrument development processes and attributes that may be helpful for documenting an instrument's content validity. These include, item generation, collection method and instrument administration mode, recall period, and response options. The details of these are included in the 2009 FDA Guidance for Industry on PROs and are shown in an Appendix at http://www.iovs.org/content/51/12/6095/suppl/DC1.

Participants in the 2009 NEI/FDA Clinical Trial Endpoints Symposium [Commercial Relationship Disclosure]

Dimitri T. Azar, Head, Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, IL [Bausch & Lomb (C), Alcon Laboratories (C), Allergan (C), Foresight (C), Sarentis (C), Prism Ventures (C)]

Neil M. Bressler, Chief, Retina Division, Johns Hopkins University, Baltimore, MD [Allergan (F) Bausch & Lomb (F), Bristol-Meyers Squibb (F), Carl Zeiss Meditec (F), EMMES Corporation (F), Genentech (F), Lumenis (F), Notal Vision (F), Novartis (F), Othera (F), QLT (F), Regeneron (F), Steba Biotech (F)]

Laurie B. Burke, Director, Study Endpoints and Label Development, Office of New Drugs, Center for Drug Evaluation and Research (CDER), U.S. Food and Drug Administration (FDA), Silver Spring, MD [None]

Wiley A. Chambers, Acting Director, Division of anti-Infective and Ophthalmology Products, CDER, FDA [None]

Anne L. Coleman, University of California, Los Angeles [Alcon (C), Allergan (C), Pfizer (C), Science-based Health (C)]

Malvina Eydelman, Director, Division of Ophthalmic and Ear, Nose and Throat Devices, Center for Device and Radiologic Health (CDRH), FDA [None]

Frederick L. Ferris III, Clinical Director, National Eye Institute (NEI), Bethesda, MD [None]

Suber S. Huang, Director, Vitreoretinal Diseases and Surgery, University Hospitals/Case Western Reserve University, Cleveland, OH [Bausch & Lomb Surgical (C); Merck & Co., Inc. (C); SurModics, Inc. (C, I); Second Sight (C); Therapeutic Nanoparticle and Molecular Imaging (C); Digital Healthcare, Inc. (C, I); i2i Innovative Ideas, Inc. (E); as Director of the Case Western Reserve REDIARC (Retinal Diseases Image Analysis Reading Center): Neurotech (C), Lux Bio (C), Alcon (C), Pfizer (C), Shering Plough (C), Vitreo Retinal Technologies, (C) MacuSight (C)]

Paul P. Lee, Vice Chairman, Duke University Eye Center, Durham, NC [Alcon (C, R, F), Allergan (C, R, F), Pfizer (C, R, F, I), Genentech (C, F)]

Nancy K. Leidy, Senior Vice President, Scientific Affairs, United BioSource Corporation, Bethesda, MD [None]

Anne S. Lindblad, Executive Vice President, The EMMES Corporation, Rockville, MD [The EMMES Corporation (F)]

Danica Marinac-Dabic, Director, Division of Epidemiology, CDRH, FDA [None]

Peter J. McDonnell, Director, Wilmer Eye Institute, The Johns Hopkins University School of Medicine, Baltimore, MD [None]

Timothy T. McMahon, Director of Contact Lens Service, University of Illinois, Chicago, IL [Bausch & Lomb, Inc. (C), Vistakon, Inc. (C)]

Päivi Miskala, Study Endpoints Reviewer, Study Endpoints and Label Development, CDER, FDA [None]

William L. Rich III, Falls Church, VA [None]

Paul A. Sieving, Director, NEI [None]

Roger F. Steinert, Vice Chairman of Clinical Ophthalmology, and Director of Refractive, Cornea, and Cataract Surgery, Gavin Herbert Eye, Institute University of California, Irvine, CA [Abbott Medical Optics (C)]

Rohit Varma, Director, Glaucoma Service, Doheny Eye Institute, University of Southern California, Los Angeles, CA [Allergan (C), Pfizer and Alcon Laboratories (C), Optovue (C), Aquesys (C, I), Laboratorios Sophia (C), Bausch & Lomb Surgical (C), Merck (C), Replenish (C, I)]

Susan Vitale, Research Epidemiologist, Division of Epidemiology and Clinical Applications, NEI [None]

Robert N. Weinreb, Director, Hamilton Glaucoma Center, University of California, San Diego, CA [Aciex (C, R, F), Alcon (C, R, F), Allergan (C, R, F), GlaxoSmithKline (C, R, F), Merck (C, R, F), Optovue (C, R, F), Pfizer (C, R, F), Topcon (C, R, F), Carl Zeiss Meditec (C, R, F), Heidelberg Engineering (C, R, F), Novartis (C, R, F)]

The FDA reviews a PRO instrument by evaluating whether it measures a well-defined concept that is supported by empiric evidence from qualitative research and psychometric validation studies in the intended clinical trial target population. The FDA evaluates whether the instrument is specific for the clinical trial target population and for the target indication that the sponsor is planning to pursue. The adequacy of the measurement properties is important, especially content validity (see box).

Content validity documentation includes reports from the literature, expert opinion, and patient input derived from qualitative research (in-depth patient interviews or focus groups, and cognitive interviews with patients). Patients who are included in the qualitative research should represent the sociodemographic and clinical characteristics of the intended clinical trial target population to the extent possible.

The FDA evaluates content validity to determine whether the instrument measures the concept(s) it is intended to measure and whether the concept(s) measured by the instrument match the specific language targeted for a labeling claim.

The FDA, according to Dr. Miskala, generally recommends that pain intensity, for example, be assessed via a single-item, patient-reported, worst-pain-intensity measure with a 24-hour recall period in a daily diary. A patient-reported daily diary may be preferable over a longer recall period when one might expect considerable day-to-day variability in the symptom being measured. Information related to pain medication use (recorded separately) is important to determine, whether any improvement in pain is due to the investigational product or to increased use of other pain medications.

Development of assessments to evaluate children brings up a whole host of additional instrument development considerations and challenges. When a child is old enough to self-assess, the child's assessment is preferable over a caregiver's report. Assessments developed to evaluate young children who cannot self-assess should ensure that parents or caregivers evaluate and report only on the observable signs and/or behavior of children. Proxy assessments, in which parent or caregiver reports on things that are only known to the child (e.g., nonobservable symptoms), are generally not appropriate for the purpose of evaluating treatment efficacy. The same applies when evaluating incapacitated patients.

After content validity, in terms of importance, come other measurement properties such as construct validity, reliability, and ability to detect change. Construct validity demonstrates expected relationships with other measures or with scores produced in patient groups known to be similar or diverse. Reliability demonstrates stability of scores over time, internal consistency, and interinterviewer reproducibility. Ability to detect change demonstrates that the scores change in a predicted direction when there has been a notable change in the patient and that the scores are stable when there is no change in the patient. For international studies, documentation of instrument translation and cultural adaptation are very important review considerations, and those concerns should be taken into consideration early in PRO instrument development.

Sponsors should consult the FDA early in medical product development about the PRO instrument that is intended as a key endpoint, to evaluate treatment efficacy to support medical product labeling claims. These discussions can occur as early as during the pre-IND (Investigational New Drug) meeting with the FDA, if appropriate. In addition, the FDA can provide written comments on PRO submissions anytime during medical product development.

Items, Concepts, Domains, and Conceptual Framework in PRO Instruments

FDA Definitions

Item: an individual question, statement, or task that is evaluated by the patient to address a particular concept.
Concept: the "thing" being measured
Domain: a discrete concept within a multidomain concept
Conceptual framework: the expected relationship of items within a domain and of domains within a PRO concept

The conceptual framework of a PRO instrument may be straightforward if a single item is a reliable and valid measure of the concept of interest (e.g., pain intensity). If the concept of interest is general (e.g., physical function), a single-item PRO instrument does not provide a useful understanding of the treatment's effect, because a stand-alone single item does not capture the domains of the general concept. For this reason, single-item questions about general concepts that include multiple items or domains rarely provide sufficient evidence to support claims about that general concept. For example, in clinical trials of functional disorders defined by clusters of specific symptoms and signs, a PRO instrument consisting of a single-item global question usually would be inadequate as an endpoint to support labeling claims and would be uninformative about the effects on each specific symptom and sign. Instead, the effect of treatment on each of the appropriate symptoms and signs should be adequately measured.

Content Validity: The FDA Definition

"Evidence from qualitative research demonstrating that the instrument measures the concept of interest including evidence that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use. Testing other measurement properties will not replace or rectify problems with content validity" (Guidance for Industry[9, p31]).

For example, if a PRO instrument purports to assess symptoms of allergic conjunctivitis in adult patients, to support that indication, the content validity documentation should include a literature review, expert input, and empiric evidence from qualitative research in patients who reflect the intended clinical trial target population. This documentation should support the conclusion that the instrument items adequately capture symptoms of allergic conjunctivitis and that patients are able to understand and provide responses to the items in a way that adequately represents their allergic conjunctivitis symptoms. Qualitative research to support instrument development typically consists of concept elicitation interviews with patients, to determine instrument item content, and cognitive interviews with patients, to evaluate whether the instrument is interpretable to them.

Dr. Miskala summarized several key points to remember:

- Not all types of PROs are appropriate for evaluating treatment efficacy.
- PRO instruments intended as key trial endpoints to support medical product labeling claims typically require development and validation early in medical product development.
- The FDA evaluates PRO measures in the context of stated labeling goals. Content validity is an important review consideration.
- PRO endpoints should be well-established before proceeding to phase 3 drug trials or pivotal device trials.

The FDA's published guidance contains details about how it evaluates the PRO instruments used as primary or key secondary endpoints in clinical trials to support labeling claims. It will

consider alternative approaches if the approaches satisfy the requirements of applicable statues and regulations. Again, discussion with the FDA of any PRO measure should begin early in development process of the drug or medical device.

Practical Matters in Instrument Development: An Instrument Developer's Perspective

Developing a novel PRO instrument or selecting an existing instrument begins with a careful analysis of the target concept, population, purpose, and settings. It requires systematic design. Qualitative research must be conducted to inform the instrument structure and content of the item pool and quantitative research to pare down the number of items and test the questionnaire's psychometric performance. At the outset, it is critically important, says Nancy Leidy, PhD, to *clarify the target concept*:

- Who is the target population for this instrument? Is it patients with glaucoma, for example? What are the disease stage and age group?
- Is the clinical problem physiologic or behavioral?
- Is the planned intervention is planned pharmaceutical, device-driven, or behavioral?
- Is the intended outcome, concept, or claim to improve, stabilize, or prevent a disorder?
- What is wrong with the way the outcome is currently defined? In vision research, a problem is obviously that visual acuity and visual field measurements often tell only a portion of the story about the impact on patients of vision disorders and treatments.
- How is the outcome (concept) currently measured? What instrument or other means (e.g., clinical assessment) is being used?
- Is the instrument fit for the purpose and sufficiently sensitive?

Knowing the intended outcome, concept, or claim is an important first step in discussions about what instrument should be used and how the results should be evaluated. It is helpful to "begin with the end in mind" by clarifying the concept being measured and considering the options. HRQOL instruments can be especially difficult to develop and interpret because they address multidimensional topics (e.g., physical, functional, social, emotional, and spiritual) and differ across therapeutic areas. It is helpful to involve experts who specialize in the disease area and product line of interest and to consult FDA guidance (e.g., target product profile, endpoint model, and conceptual framework).

The development of an instrument involves two major phases: qualitative and quantitative. The qualitative phase involves transforming words and phrases from patient interviews into questions that will appear on the instrument. The effect of length of the recall period and language and cultural differences is evaluated. Patients are then asked to react to the instrument by telling what the items mean to them. The quantitative phase follows and involves reducing the number of items and testing the instrument that has been designed, keeping in mind such elements as content coverage, administrative methods, user instructions, stems and response options, data organization and analysis, and documentation. The documentation is studied carefully by FDA reviewers who render opinions about the instrument.

The best approach for moving the development process forward at an acceptable pace, according to Dr. Leidy, is to make it a team effort with persons or groups who are disease, product, and foreign language experts and with experts in data collection and analysis, psychometrics, statistical programming, and regulatory processes. It is a long process that can easily take two or more years. As Dr. Leidy pointed out, more than one study may be necessary to for validate an instrument. Validation grows as the instrument is tested over time and across populations and settings and in different administrative modalities (e.g., personal digital assistant, PDA). It must be used in a variety of different settings and in different populations. The FDA regularly advises sponsors about the development of PRO instruments. For companies with products in the pipeline but limited resources for PRO instrument development, the FDA suggests that they attempt to pool resources with other companies. The FDA is developing a system for qualifying instruments for use on specific populations, which may lighten the burden for individual sponsors. Another boon is that a single instrument could be used to assess different drugs for the same condition. The FDA would study documentation showing its content validity for the different uses.

Evaluation of Ophthalmic Devices

For purposes of clarity, Malvina B. Eydelman, MD, and Danica Marinac-Dabic, PhD, described a medical device as something that diagnoses, cures, mitigates, treats, or prevents a disease or condition or affects the function or structure of the body; does not achieve its intended use through chemical action; and is not dependent on being metabolized for its primary intended purposes. Recommendations for evaluation of devices, preclinical testing, and clinical trials come from available standards (e.g., International Organization for Standardization, ISO, and American National Standards Institute, ANSI) and FDA guidance. Patient-reported outcomes are an important part of the evaluation of ophthalmic devices.

PROs are being collected in many clinical trials of new devices. In general, PROs are *not* being used as primary endpoints for supporting marketing of ophthalmic devices, but they are being considered in FDA safety reviews of marketing applications and in forming recommendations regarding approval and clearance. PROs have been used as primary endpoints for postapproval studies. Examples include clinical trials and postapproval studies of phakic intraocular lenses, multifocal intraocular lenses, toric intraocular lenses, contact lenses, retinal prostheses, and refractive laser trials. The standards and guidance documents for these device categories recommend assessment of PROs, such as, glare, halos, double vision, spectacle and contact lens use, night driving, dry eye, and visual distortion, among others. The incorporation of PRO measures into clinical trials of implantable aqueous shunts for glaucoma is currently under discussion.

Because of the time and cost involved in administration of traditional paper-and-pencil PRO instruments, the FDA and NEI are currently collaborating on a study with the goal of trying to reduce this burden through computer-based administration of ophthalmic questionnaires. They will compare outcomes of previously validated paper-and-pencil questionnaires to those of computer-based versions of the same questionnaires. The FDA and NEI are also collaborating on development of an online PRO questionnaire for government-sponsored LASIK clinical trials. The CDRH perspective is to encourage development of validated PRO instruments that focus on the impact of treatment with ophthalmic devices.

The FDA has the authority to impose postmarket studies either at the time a device is approved or later if problems arise. The sponsor is required to continue evaluating and reporting to the FDA on the safety, effectiveness, and reliability of the device for its intended purpose.

The CDRH approach to PROs is to advocate for and support the use of existing validated instruments and the development and use of new ones in premarket activities, postapproval studies, and epidemiologic studies.

What We Know and Need to Know about PROs in Ophthalmology: The Clinical Perspective

We know that the concepts— defined by the FDA as the *things* being measured by a PRO instrument—that are important to patients may well be different, depending on whether the perspective is from, say, refractive surgery, cataract surgery, glaucoma, or retinal. Meeting presenters described characteristics of several ophthalmic conditions that may qualify as concepts. The following summaries are based on these presentations and approved for publication by each presenter.

Refractive Surgery

Over the years many patients treated with refractive surgery to improve visual acuity have reported problems with glare, halos, dry eye, and poor night vision. A question posed by presenter Peter McDonnell, MD, is why it can take many years before a medical field openly recognizes that a procedure has, for some patients, a negative impact on quality of life despite a good objective outcome (e.g., 20/20 visual acuity). He suggested that it is impossible to describe a visual event for which there is no accepted yardstick for comparison. He further suggested that contrast sensitivity, glare testing, or aberrometry results—if their ability to demonstrate refractive error correction-related quality of life were known— could be useful in clinical trials for assessing true patient outcomes.

Cataract Surgery

With the introduction of premium intraocular lenses (e.g., multifocal lenses, accommodating lenses, toric lenses, and light-adjustable lenses), cataract surgery is converging with visual acuity correction. Just as with refractive surgery, best corrected visual acuity after cataract surgery does not always correlate with complete patient satisfaction. PROs as endpoints in cataract surgery are needed to develop new products based on patients' report of their level of functioning and satisfaction. Roger Steinert, MD, believes that, to meet patients' expectations after cataract surgery, there is a need to establish objective PRO standards for visual functioning under different lighting conditions, for dysphotopsias, and for freedom from eyeglasses or contact lenses. Furthermore, the cost of premium implants represents a substantial expense to patients and insurers, and it behooves ophthalmologists to justify the cost on the basis of PROs. For the industry, the cost of developing new generations of products is worthwhile if feedback from patients supports effectiveness and acceptability.

Glaucoma

According to Robert N. Weinreb, MD, progress in glaucoma clinical research and practice is impeded by a reliance on visual field testing to evaluate the progression of glaucoma. Nevertheless, he believes there has never been a better time to translate discoveries from glaucoma research into clinical practice, and there has never been a better opportunity to enhance patient care and prevent glaucoma blindness by incorporating PROs into this research. Standard visual field testing provides limited data about what patients actually see and how they function. Patient-reported outcome measures could help overcome the problem, as illustrated with the following example:

A 70-year-old college professor with primary open-angle glaucoma has visual acuity of RE, 20/40, and LE, 20/30. The neuroretinal rims of her optic discs are narrow. Visual fields are diffusely depressed, and she has a mean deviation in both eyes of approximately 7 dB. Despite this objective assessment of her visual function, she is a satisfied patient, mainly because her work and other pastimes depend on her ability to read.

Another patient is a 62-year-old dentist with pigmentary glaucoma who has visual acuity in both eyes of 20/20. He has paracentral scotoma, but a mean deviation of ~0.0 dB in his visual fields. Despite good visual acuity, he is on disability related to optic nerve damage. His vision prevents him from engaging in his usual activity of practicing dentistry.

Understanding the impact of glaucoma and glaucoma treatments by using patient-reported measures could capture results beyond visual field test results and move therapies forward to improve patient satisfaction.

Retinal Diseases

The most common causes of vision loss related to retinal disease are AMD and diabetic retinopathy. In these conditions, change in visual acuity is the ultimate health-reported outcome, similar to refractive surgery and cataract surgery. Suber Huang, MD, MBA, stated that visual acuity testing does not address other functional aspects of vision (e.g., contrast, color, visual movement, central visual field, partial deficiencies, and dynamic or temporal fluctuations), nor does it capture difficulties in emotional and behavioral functioning across a continuum. He emphasizes that in developing treatments for patients with retinal disease, the best results depend on tools that sum the many interconnected aspects of visual and life functioning.

American Academy of Ophthalmology Perspective

William L. Rich, III, MD, reported on the point of view of the American Academy of Ophthalmology (AAO), saying that it is entirely appropriate that PROs become an integral part of NEI and FDA studies. This statement is in agreement with the viewpoints of the various specialists reported herein. No longer is the classic randomized clinical trial going to be the basis of adoption of new technologies or devices. Society is acutely aware of the deficiencies in the evaluation of outcome metrics, and no longer will industry be assured of coverage without addressing up front the efficacy and outcomes. One of the integral parts of outcome measure development is patient-reported impacts.

With the rapid growth in spending on technology come societal pressures to ensure that the adoption of new technologies leads to meaningful outcomes. One the major metrics of outcome measure development is the impact on the patient's health and lifestyle. The Institute of Medicine consistently emphasizes the importance of patient-centeredness. The National Quality Forum, charged with developing national outcomes measures, ensures that measures rely less on randomized clinical trials with risk-adjusted populations than on looking at population-based studies that incorporate patient-reported impacts on health and satisfaction accompanied by the aggregation of a large number of patients. If the ultimate successful adoption of new innovations in care require the demonstration of meaningful outcomes and one element of outcome measurement is PROs, shouldn't they be included?

The Perspective from Optometry

Reporting on the perspective from optometry about the place for PROs in clinical trials was Timothy McMahon, OD, He used an 8-year natural history study of patients with keratoconus as an example. The study revealed that despite visual acuity of 20/30 to 20/40 in the study participants, the score for many on visual function questionnaires was comparable to that of pa-

tients with moderate and even advanced AMD.[10,11] The observation reinforces the disconnect between classic measurements such as visual acuity and the actual experience of the patient as noted in the earlier discussions of refractive surgery, cataract surgery, and retinal diseases.

What We Know and Need to Know about PROs in Ophthalmology: The Research Perspective

Refractive Surgery and Cataract Surgery

Dimitri Azar, MD, reported the results of a study in which researchers compared refractive surgical procedures to correct myopia, hyperopia, astigmatism, and presbyopia, in which no clinically significant differences were found in patients' visual and refractive outcomes.[12] The study used FDA data and identified five outcome measures, none of which were PROs. Despite appearances of success we know, however, that there are differences. Patients report a range of post surgical symptoms, including poor distance vision, fluctuating vision, dry eye, redness, pain, glare, and halos, among others.

In several studies, instruments have been used for examining vision-related QOL relative to refractive surgery. The RSVP questionnaire, for example, measured pre- and postoperative concerns, expectations, physical and social functioning, and driving. People with higher refractive error scored higher in the concern category; men scored higher in expectations; contact lens wearers and people with greater refractive error showed more problems in physical and social functioning; and, with regard to driving, more problems were seen among females, older patients, people with greater refractive error, and those wearing contact lenses and glasses. Some of these factors contributed more than visual acuity to patients' overall ratings of postoperative vision. In other words, clinical measures do not always reflect a patient's satisfaction with his or her vision, and satisfaction may be influenced by sex, age, expectations, refractive error, and other factors. In fact, the RSVP questionnaire showed 15% dissatisfaction with refractive surgery.[13] Given that QOL is a subset of PROs and that QOL is affected by refractive surgeries, it follows that measuring visual acuity and HRQOL to compare refractive surgeries benefits patients.

Similarly, after cataract surgery, patient-reported assessments of functioning, satisfaction, and symptoms have identified aspects that traditional clinical measures fail to detect.[14,15] For example, satisfaction was greatest in patients with no preexisting ocular disease and in patients with more dense cataracts; the results of second-eye cataract surgery produced significant improvements in visual acuity, visual function, and psychosocial health status[16]; multifocal intraocular lens recipients showed greater satisfaction with distance, intermediate, near, and overall vision without glasses than did monofocal lens recipients.[17] A study using the VF-14 visual function index to compare HRQOL among patients receiving one of three types of multifocal lenses showed no statistical differences in visual acuity, glare, halo, or satisfaction with vision among the types.[18]

Additional PRO studies are needed to understand the reasons and predisposing factors that lead to different outcomes for patients who undergo refractive and cataract surgery, says Dr. Azar. Among the factors to consider is ethnicity. Of the Latino adults who undergo cataract surgery, a significant proportion have residual visual impairment.[19] Ideally, a determination of the proper lens or surgery for a patient would be based on the patient's characteristics derived on an analysis that includes PROs.

Glaucoma

This is not the first time the FDA and the ophthalmology community has considered QOL in discussions about glaucoma clinical trials. It figured into the discussions during the 2008 NEI/FDA CDER glaucoma clinical trials endpoints symposium.[2]

Paul Lee, MD, JD, reflected on the instruments that have been used to measure HRQOL in glaucoma patients. Some are vision specific; others are glaucoma specific; yet others emphasize general health. One, the VFQ-25, is a reliable and valid 25-item adaptation of the earlier 51-item NEI Visual Function Questionnaire (NEI-VFQ), and is especially useful in clinical trials. Some details about the VFQ-25:

- It was validated in people with eye disease and visual impairment.
- Test–retest reliability was between 0.68 and 0.91.
- The shorter VFQ-25 has demonstrated the same internal consistency and test–retest reliability as the VFQ-51.
- It has shown that people with eye disease and visual impairment have lower visual function scores than do groups without eye disease or visual impairment.

A key problem in glaucoma and other ophthalmic specialties is that visual field loss and visual acuity, which clinicians typically measure in eye examinations, are imperfect indicators of how patients are truly faring on their own or in response to treatment. A critical question, says Dr. Lee, is, "What degree of glaucomatous visual field change is necessary to observe meaningful change in the ability of adults to function independently or complete vision-related tasks?"

We have learned that visual function loss in glaucoma compromises abilities like reading, seeing details, outdoor mobility, and functional peripheral vision.[20] We know that worse vision imposes more limits in terms of instrumental activities of daily living (IADLs),[21] which include driving, preparing meals, doing housework, shopping, managing finances, managing medication, and using the telephone. Drivers with moderate to severe bilateral visual field loss (VFL) report significantly greater difficulty with night driving and tasks involving visual searching and visual processing speed than do drivers with less bilateral VFL.[22] These findings, based on the Visual Activities Questionnaire (VAQ) and the NEI-VFQ-25, speak to the ability of test instruments to respond to clinically meaningful differences in patient status and changes due to treatment interventions. Similarly, studies to determine the effect of glaucoma on reading speed in elderly subjects finds that 21.1% with unilateral glaucoma and 28.4% with bilateral glaucoma have impaired ability to reading.[23] This compares to a matched sample of subjects without glaucoma whose reading impairment was 16.0%. The authors note that race, cognitive ability, education, and visual acuity are important predictors of reading impairment. Others agree, based on Rasch analysis, that the NEI-VFQ 25 may not serve equally well as an index of the impact of therapeutic interventions and rehabilitation programs in all populations.[24]

The point is to use valid and reliable PRO measures in clinical trials of new drugs, devices, and biologics so that clinicians, in selecting the right treatment for patients, will have the advantage of research findings that take into account objective visual acuity and visual field tests and also demographics and patient reports. Some people have suggested that PROs be used as primary endpoints in clinical trials, whereas others, including Anne Lindblad, PhD, and Paul Sieving, MD, suggest that they are more appropriate as just one of several components. The situation could change as new measurement instruments are developed and subjected to scrutiny.

Retinal Diseases

A review paper published in 2002 reported on 22 vision-specific instruments for assessing HRQOL and visual functioning. Nine had some level of validation in retinal disease.[25] In 2008, six instruments specific to AMD were reviewed.[26] Some are well-validated and have been shown to differentiate between disease stages and to be responsive. Is one instrument superior across the board or does each measure a different function? Could they differentiate between diseases? The answers may be obtainable now as longitudinal studies of treatments have established the effectiveness of treatments and responsiveness of some of these measures. On top of these results, it may be possible to use one of these HRQOL instruments to measure clinically meaningful vision-specific changes, keeping in mind that some instruments are sensitive to changes in the better or the worse eye.

Available instruments for the retina are mostly able to discriminate disease characteristics in AMD, diabetic macular edema, diabetic retinopathy, branch retinal vein occlusion, and cytomegalovirus. They include the following, some of which are specific to retina, whereas others measure vision or contrast sensitivity in general:

Activities of Daily Living Scale (ADVS)
Daily Living Tasks Dependent on Vision (DLTV)
Impact of Vision Impairment (IVI)
Macular Disease Quality of Life Questionnaire
NEI-VFQ
Visual Function Index (VF-14)
Low-Luminance Questionnaire (LLQ)
Miedziak's instrument
Vision-specific sickness impact profile (SIPV)
Turano's instrument
Vision-Related Quality of Life Questionnaire
Retinopathy-Dependent QOL

There is reasonable consistency in these scales, especially for the NEI-VFQ, where one cross-sectional analysis from a prevalence study in which only 6% of the participants had a visual acuity of 20/40 or worse in the better-seeing eye showed a 5-point change corresponded to a 1- or 2-line vision change.[27] The ANCHOR and MARINA studies, in which more than half the participants had a visual acuity worse than 20/40 in the better-seeing eye, showed differences in VFQ in diverse treatment arms.[5,6] The combined scores of the ANCHOR and MARINA studies suggested that a 4- to 6-point change would correspond to approximately a 15-letter change in visual acuity.[28]

As Dr. Lindblad described, most instruments measure the ability of a person to do a task but not usually the impact of not being able to do the task or how important the task is to the individual. Also, divergent tasks (e.g., threading a needle and interacting socially) are often weighted equally, which may not be accurate. Good measurements should have unidimensionality, hierarchical order, and equal interval spacing.[29]

Another question concerns which instrument to use. Some are not vision specific and have not been shown to be sensitive to loss of vision. In others, depression and socioeconomic status are confounding variables. Is one instrument good for all retinal conditions, or do we need disease-specific instruments? Standardizing instruments within disease-specific trials is important to allow cross-study comparisons. Other issues relate to how a questionnaire would be administered and the point at which it is given in the course of a study.

Dr. Lindblad suggested that PROs may ultimately become useful as surrogates for visual acuity testing in patients who are no longer able to travel to the clinic during the course of a clinical trial. Missing data in studies of the elderly may be the consequence of health deterioration and worsening vision. Including PROs as a surrogate can help quantify the potential impact of these missing data on study results.

Whether these measures are suitable as key endpoints in clinical trials to evaluate treatment efficacy and to support medical product labeling claims is up for further debate The outcome of the discussion depends on the ophthalmic research community's providing empiric evidence to the FDA that demonstrates the instrument's content validity and the adequacy of other measurement properties in the intended clinical trial target population.

Clinical Trial Design Challenges Relevant to Patient-Reported Outcomes Intended to Support Medical Product Labeling

Trials of Ophthalmic Drugs and Biologics

As Wiley Chambers, MD, reminded the symposium participants, the mission of the CDER is to assure that safe and effective drugs are available to the American people. The FDA accomplishes its mission in three ways:

- By monitoring the drug development process during investigational stages. This process is mostly confidential and between only the FDA and sponsor.
- By approving new drug products that are safe and efficacious. This process is confidential until approval and then becomes deliberately transparent.
- By monitoring adverse events after approval, to assure that the risk–benefit ratio remains acceptable.

The most familiar endpoints in clinical trials are objective anatomic measurements. Examples are reattachment of the retina, prevention of retinal detachment, prevention of CMV retinitis progression, prevention of diabetic retinopathy progression, and re-epithelialization of the cornea with elimination of bacteria. Other objective endpoints include intraocular pressure, refractive power, conjunctival redness, pupil size, and tear production.

Visual function (visual acuity, color vision, visual fields, and contrast sensitivity) is considered a subjective endpoint because it requires a guided response. PROs are obviously subjective endpoints and may be single- or multidomain. It is easier to get a claim for a simple rating-scale–based endpoint that measures a single well-defined concept for the obvious reason that each item in a multidomain must be validated.

The FDA evaluates all products in the context of risk, recognizing that all drugs have some risk and that assessment is open-ended until a product has been on the market and used by a large number of people. The basis for approval is whether the benefits outweigh the risks when the product is taken by the intended population as labeled. The risk is expected to be demonstrated in *adequate and well-controlled studies* showing safety and efficacy. Uncontrolled studies may be used to provide corroborative evidence but are not acceptable on their own.

Seven primary components make up an adequate and well-controlled clinical trial:

1. A clear statement of the objectives of the investigation
2. A design that permits a valid comparison with a control to provide a quantitative assessment of the drug effect
3. A method of selecting subjects that provides adequate assurance that the subjects have the disease or condition

being studied or evidence of susceptibility and exposure to the condition against which prophylaxis is directed
4. A method of assigning patients to treatment and control groups that minimizes bias and is intended to assure comparability of the groups with respect to pertinent variables such as age, sex, severity of disease, duration of disease, and use of drugs or therapy other than the test drug
5. Adequate measures to minimize bias on the part of the subjects, observers, and data analysts
6. Methods of assessment of subjects' response that are well-defined and reliable
7. An analysis of the results of the study that is adequate to assess the effects of the drug

These seven components, as presented here, are described in the Code of Federal Regulations and are the characteristics of an adequate and well-controlled study.

Some aspects, according to Dr. Chambers, are subject to interpretation, such as whether one eye can serve as a control for the other. It often depends, he says, on the particular disease being studied. Some diseases follow the same pattern in both eyes, and one eye is predictive of the clinical course of the other eye. Allergic conjunctivitis is an example of a disorder in which it is best not to use the fellow eye as the control because the eyes respond to allergens independently of each other.

What Are Some Clinically Meaningful Challenges in this Field?

Anne Coleman, MD, PhD, reiterated several challenges in developing PROs as clinical trial endpoints in any field. One challenge is choosing the correct psychometric tests for validating a PRO instrument. Another challenge concerns whether to evaluate patients based on composite scores or specific domains. (The FDA defines a domain as a discrete concept within a multidomain concept. All the items in a single domain contribute to the measurement of the domain concept.) A composite score does not recognize that HRQOL is multidimensional; it includes both subjective well-being and objective functioning. Looking at composite scores would not reveal, for example, whether two patients with similar scores are the same or different in, say, a physical or mental domain.

Yet another challenge is that PRO instruments need continuous validation, correlation with other instruments, and assessment in new patient populations and in relation to co-morbidities and other factors like aging, depression, and adaptation to vision disorders over time. Looking at new associations across domains could lead to insights that could help patients and their physicians choose from among different treatments. As of now, many clinicians are unconvinced that HRQOL measurements are useful compared with clinical tests.

Early interaction of researchers and sponsors with the FDA is critically important for moving the field forward to understand what patients value most about HRQOL relative to vision disorders and how data can help in making informed decisions about treatments. Perhaps the ultimate challenge for advocates of PROs as primary or secondary outcomes in ophthalmic clinical trials and in labeling claims for new drugs and devices is to follow the FDA guidance document for establishing their scientific validity.

What's Next?

Whether PROs are suitable as primary endpoints in clinical trials is up for further debate and depends on the ophthalmology community providing data to the FDA that demonstrate instrument's content validity and adequate other measurement properties in the intended clinical trial target population.

The clear message from the FDA is that they recognize that PROs can be useful in the evaluation of drugs and devices. They also recognize the challenge in developing appropriate PRO instruments. The FDA encourages drug and device makers to schedule meetings with the agency before embarking on a phase 3 trial, to determine whether instruments are appropriate for the stated purpose. The agency acknowledges that it can be less difficult to gain approval of specific items than broad issues (e.g., ocular pain versus HRQOL).

The FDA indicated that to consider using items or domains from a PRO instrument such as the NEI VFQ-25—the well-studied instrument applied in numerous clinical studies—for medical product labeling, their content validity and other measurement property documentation for the intended clinical trial population would have to be submitted to the FDA for review. Content validity documentation, in particular, may not be described in sufficient detail in publications for the FDA review purpose and the FDA generally requests to see that documentation in detail (e.g., qualitative research protocols, interview guides, data summaries). In addition, to be considered for medical product labeling, the concept measured by the instrument must be clinically appropriate for evaluating treatment efficacy as a primary or key secondary clinical trial endpoint, multiplicity concerns must be taken into consideration in the statistical analysis plan, and a statistically and clinically important treatment benefit must be demonstrated in phase 3 trials.

PROs in labeling will help the patient understand how a drug or device may affect everyday life. PROs in labeling will help clinicians and surgeons make better decisions for treatment of patients based on the outcomes that patients value most for themselves.

References

1. Csaky KS, Richman EA, Ferris FL. Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci.* 2008;49:479–489.
2. Weinreb RN, Kaufman PL. The glaucoma research community and FDA look to the future: a report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci.* 2009;50:1497–1505.
3. Hyman LG, Komaroff E, Heijl A, Bengtsson B, Leske MC. Treatment and vision-related quality of life in the early manifest glaucoma trial: Early Manifest Glaucoma Trial Group. *Ophthalmology.* 2005: 112;1505–1513.
4. Janz NK, Wren PA, Lichter PR, et al. The Collaborative Initial Glaucoma Treatment Study (CIGTS): interim quality of life findings following initial medical or surgical treatment of glaucoma. *Ophthalmology.* 2001;108:1954–1965.
5. Bressler NM, Chang TS, Fine JT, Dolan CM, Ward J. Anti-VEGF Antibody for the Treatment of Predominantly Classic Choroidal Neovascularization in Age-Related Macular Degeneration (ANCHOR) Research Group. Improved vision-related function after ranibizumab vs. photodynamic therapy: a randomized clinical trial. *Arch Ophthalmol.* 2009;127:13–21.
6. Chang TS, Bressler NM, Fine JT, Dolan CM, Ward J, Klesert TR, MARINA Study Group. Improved vision-related function after ranibizumab treatment of neovascular age-related macular degeneration: results of a randomized clinical trial. *Arch Ophthalmol.* 2007;125:1460–1469.
7. Perry HD, Solomon R, Donnenfeld ED, et al. Evaluation of topical cyclosporine for the treatment of dry eye disease. *Arch Ophthalmol.* 2008;126:1046–1050.
8. PRO Labels. http://www.mapi-prolabels.org/more_details.php#pro_label. Lyon France: Mapi Research Trust; 2009. Accessed June 2009.
9. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, MD: U.S. Department of Health and Human Services.

Food and Drug Administration. Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH); December 2009. (http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf) Accessed December 2009.

10. Kymes SM, Walline JJ, Zadnik K, Gordon MO. Quality of life in keratoconus. *Am J Ophthalmol.* 2004;138(4):527–535.
11. Kymes SM, Walline JJ, Zadnik K, et al. Changes in the quality-of-life of people with keratoconus. *Am J Ophthalmol.* 2008;145(4):611–617.
12. Sakimoto I, Rosenblatt MI, Azar DT. Laser eye surgery for refractive errors. *Lancet.* 2006;367:1432–1447.
13. Schein OD, Vitale S, Cassard SD, Steinberg EP. Patient outcomes of refractive surgery: the refractive status and vision profile. *J Cataract Refract Surg.* 2001;27:665–673.
14. Steinberg EP, Tielsch JM, Schein OD, et al. National study of cataract surgery outcomes: variation in 4-month postoperative outcomes as reflected in multiple outcome measures. *Ophthalmology.* 1994;101:1131–1140, discussion 1140–1141.
15. Mangione CM, Phillips RS, Seddon JM, et al. Development of the 'Activities of Daily Vision Scale': a measure of visual functional status. *Med Care.* 1992;30:1111–1126.
16. Castells X, Alonso J, Ribó C, et al. Comparison of the results of first and second cataract eye surgery. *Ophthalmology.* 1999;106:676–682.
17. Javitt JC, Steinert RF. Cataract extraction with multifocal intraocular lens implantation: a multinational clinical trial evaluating clinical, functional, and quality-of-life outcomes. *Ophthalmology.* 2000;107:2040–2048.
18. Gierek-Ciaciura S, Cwalina L, Bednarski L, Mrukwa-Kominek E. A comparative clinical study of the visual results between three types of multifocal lenses. *Graefes Arch Clin Exp Ophthalmol.* 2010;248:133–140.
19. Barañano AE, Wu J, Mazhar K, Azen SP, Varma R; Los Angeles Latino Eye Study Group. Visual acuity outcomes after cataract extraction in adult Latinos. *Ophthalmology*. 2008;115:815–821.
20. Ramulu P. Glaucoma and disability: which tasks are affected, and at what stage of disease? *Curr Opin Ophthalmol.* 2009;20:92–98.
21. Lee P, Smith JP, Kington R. The relationship of self-rated vision and hearing to functional status and well-being among seniors 70 years and older. *Am J Ophthalmol.* 1999;127:447–452.
22. Janz NK, Musch DC, Gillespie BW, Wren PA, Niziol LM, for the Collaborative Initial Glaucoma Treatment Study (CIGTS) Investigators. Evaluating clinical change and visual function concerns in drivers and nondrivers with glaucoma. *Invest Ophthalmol Vis Sci.* 2009;50:1718–1725.
23. Ramulu PY, West SK, Munoz B, Jampel HD, Friedman DS. Glaucoma and reading speed: The Salisbury Eye Evaluation Project. *Arch Ophthalmol.* 2009;127:82–87.
24. Labiris G, Katsanos A, Fanariotis M, et al. Sychometric properties of the Greek version of the NEI-VFQ 25. *BMC Ophthalmol.* 2008;8:4.
25. Margolis MK, Coyne K, Kennedy-Martin T, Baker T, Schein O, Revicki DA. Vision-specific instruments for the assessment of health-related quality of life and visual functioning. *Pharmacoeconomics.* 2002:20;791–812.
26. Finger RP, Fleckenstein M, Holz FG, Scholl HPN. Quality of life in age-related macular degeneration: a review of available vision-specific psychometric tools. *Qual Life Res.* 2008;17:559–574.
27. Globe DR, Wu J, Azen SP, Varma R, Los Angeles Latino Eye Study Group. The impact of visual impairment on self-reported visual functioning in Latinos: The Los Angeles Latino Eye Study. *Ophthalmology.* 2004;111:1141–1149.
28. Suner IJ, Kokame GT, Yu E, Ward J, Dolan C, Bressler NM. Responsiveness of NEI VFQ-25 to changes in visual acuity in neovascular AMD: validation studies from two phase 3 clinical trials. *Invest Ophthalmol Vis Sci.* 2009;50:3629–3635.
29. Massof RW, Fletcher DC. Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Res.* 2001;4:397–413.