Addendum: Outcomes Measure Development

Purpose
The following addendum will guide American Academy of Neurology (AAN) Measure Development Work Groups in constructing outcome measures with greater consistency and standardization. As AAN Measure Development Work Groups develop outcome measures, the AAN Quality Measure and Reporting Subcommittee (QMR) provides the following guidance on criteria for outcome measure development and a process to advance their work. This document outlines the AAN’s guiding principles and steps by which AAN’s measure development groups will develop desired clinical outcome measures and intermediate outcome measures. The QMR Subcommittee will continue to oversee the process.

Goals of Addendum
- Provide background on the need for outcome measures and rationale for AAN work groups to consider outcome measures for all measure development efforts;
- Develop a set of guiding principles for neurology outcome measures;
- Provide a framework and methodological guidance for AAN measure work groups on the development and use of outcome measures.

Background Information

Overview of measurement
The Institute of Medicine (IOM) has defined healthcare quality as the extent to which health services provided to individuals and patient populations improve desired health outcomes. Care should be based on the strongest clinical evidence and provided in a technically and culturally competent manner with good communication and shared decision-making. To achieve high-quality care, the IOM identified six “aims” for improving the delivery of care—safety, effectiveness, patient-centeredness, efficiency, timeliness, and equity. The goals of AAN’s quality efforts are to increase neurologists’ focus on measuring and improving health care quality and value for their patients and to be the primary resource for neurologists integrating quality improvement activities into their clinical practice.

Clinicians have routinely measured the degree to which they follow evidence-based processes of care with the goal of improved patient outcomes. In 1966, Avedis Donabedian, MD, a noted health services researcher, proposed that one could assess whether high quality care is provided by examining the structure of the setting in which care is provided, by measuring the actual process of care,

1 Institute of Medicine, Crossing the Quality Chasm: A New Health System for the Twenty-first Century (Washington: National Academy Press, 2001).
2 Ibid.
and by assessing what are the *outcomes* of the health care service provided. Donabedian noted that when coupled with structure and process measures, outcomes measures provide useful and actionable information for physicians on the quality of care—if they are causally related—"structure leads to process and process leads to outcomes"³.

With increasing emphasis on accountable, transparent, and patient-focused care, health care quality measurement efforts are shifting from measuring a process, or series of processes, to measuring whether a patient achieved the desired health outcomes. Donabedian defined outcomes as "changes (desirable or undesirable) in individuals and populations that are attributed to health care services received"⁴. Outcomes measurement in health care is not new⁵. Individual clinicians have used biomedical measures, such as the results of laboratory tests, to determine whether a health intervention is necessary or successful. What is new to clinical practice are the emerging domains of outcomes measurement for which physicians are held accountable⁶. While the field of outcomes measurement is not fully mature⁷, health care is testing and implementing several types of outcome measures.

**Types of measures**

- **Desired clinical outcome measures**: Desired clinical outcome measures assess the long-term result of a patient’s interaction with health care system. Many neurological conditions are very complex and difficult to measure. In these cases, *proxy measures* are ways to measure what you can when you can’t measure exactly what you want or need to. Example: it may be difficult to access data in a medical record about a practice’s effectiveness in counseling for smoking cessation. Instead the proxy measure might be how many patients had “tobacco abuse” coded as a diagnosis and how many of those patients received prescriptions for Zyban or nicotine replacement.

- **Intermediate outcomes**: Intermediate outcomes precede and are closely related to a longer-range desired outcome. Often intermediate outcome measures are physiological or biochemical values. These measures often specify thresholds or other results of clinical care that are shown to affect the desired health outcome positively or adversely. The disease process of many neurological conditions can be very slow, often taking decades for a desired clinical outcome to manifest. In these cases, an intermediate outcome measure might be more appropriate. Example: blood pressure maintained at 120/80 or less, glycemic control at <8, and cholesterol level<180 are intermediate outcome measures on the trajectory to reduced complications in diabetic patients.

• **Patient reported outcome (PROs) measures**: Patient reported outcomes measures are defined as "any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else." PROs can provide important insights for providers into the way patients perceive their health and the impact that treatments have on their quality of life. PROs assess from the patient’s perspective:

  • **Health related quality of life**: Quality of Life assessments (QOLs) seek to evaluate how a treatment or disease affects a patient psychologically, socially, and physically. QOLs can identify specific healthcare needs, provide evidence that can lead to improved quality of care, and provide a rapid screening test to identify populations or individuals who might need a more detailed health needs assessment. Example: in patients with Multiple Sclerosis, PRO measures such as the Expanded Disability Status Scale or the Multiple Sclerosis Quality of Life-54 (MSQoL-54) assess the impact of illnesses on their day-to-day lives.

  • **Patient experience with care**: Patient experience reflects the interaction between an organization and a patient as perceived through the patient’s lens. Patient experience refers to the quality and value of all of the interactions—direct and indirect, clinical and non-clinical—spanning the entire duration of the patient/provider relationship. Example: the Consumer Assessment of Healthcare Providers and Systems (CAHPS) suite of surveys assesses patient perceptions in a variety of domains including timeliness of care, interactions with clinicians and office staff, and shared decision-making and patient engagement.

• **Economic Outcomes**: Economic outcomes include measures of health resource utilization and typically represent the payer and societal perspective. Cost measures usually describe the total costs of care for a particular population, the relative cost effectiveness of treatments, and provider resource use.

• **Adverse outcomes**: Adverse events are serious complications caused by medical treatment, devices, or medical advice. Examples of adverse outcomes include rates of Wrong Site Procedures, Wrong Side Procedures, Procedures on the Wrong Patient, Wrong Procedures, and Wrong Implants.

• **Morbidity and mortality**: Historically, most of the efforts to monitor and/or report outcomes have focused on morbidity or mortality. Indicators of morbidity, often indicative of the health of a community, include the prevalence of chronic diseases or disabilities. Mortality rates and ratios measure patient deaths associated with health care services.

---

9 IBID.
Measurement definitions
Clinical quality measures include a numerator, a denominator and indicate if there are any exclusions for the measure.

- **Numerator**: The numerator is the count who received the treatment or achieved the outcome out of those who were in the denominator.

- **Denominator**: The denominator is the “universe” of who is eligible to receive the treatment or achieve the outcome.

- **Exclusions**: Denominator exclusions are a patient characteristic that allows removal of that patient from the previously defined patient cohort. A numerator exclusion represents patient characteristics that indicate reasons individual patients should not receive the interventions (e.g., medications, procedures) expected in the numerator (e.g., a patient with allergy or serious adverse reaction to an expected medication).

- **Calculation**: Numerator: (# of patients meeting numerator criteria)
  
  \[
  \text{Divided by} \\
  \text{Denominator: (} \# \text{ patients in denominator) – (} \# \text{ patients with valid denominator exclusions)}
  \]

Drivers for outcome measurement
Increasingly purchasers, payers, government agencies, and patients are requiring that health care providers demonstrate the quality of the care they provide. This immediate national focus on outcomes is placing demands on medical specialty societies to develop outcomes measures rapidly that meaningfully and accurately demonstrate high-quality patient outcomes.

Drivers for outcomes measures include:

- Public and private purchasers of health care expect providers to show how their investments are realizing positive health outcomes.

- Payers are developing payment models that focus on accountability for improved outcomes and lower costs; and providing financial incentives to providers that achieve these aims.

- Providers use measures of health outcomes assess the effectiveness of treatments and make improvements to care.

- Patients use outcome measures to select high-quality providers.

Linking process, intermediate outcomes, and desired outcomes
As noted above, Donabedian specified that a causal relationship must exist between structure, process, and outcome\(^\text{11}\). In addition, outcomes measures should be based on the evidence and the links to structure and process should be evidence-based. The figures below illustrate the link from a process to desired clinical outcome measure and an example of the link between process, intermediate outcomes and desired clinical outcomes for multiple sclerosis.

Addendum: Outcomes Measure Development

Developing Desired Clinical Outcomes Measures and Intermediate Measures

Over the past seven years, the AAN has successfully developed several quality measure sets. The AAN’s measures primarily assess implementation of evidence-based processes and have been used to improve quality, advance professional certification programs, and have been considered for public payer incentive programs.

As health care quality measurement advances towards outcomes measurement so too must the AAN’s quality measure development efforts. AAN Measure Development Work Groups (“work groups”) will continue to develop quality measures sets using the process outlined in the full 2010 AAN Measure Development Process Manual. However, future work groups will develop process measures for quality improvement and seek to include desired clinical outcome measures, intermediate outcome measures, and/or patient-reported outcome measures in each measure set.

This document identifies guiding principles for outcome measures in neurology and provides a guiding framework for AAN work groups as they develop outcome measures.

### Example in the treatment of patients with Multiple Sclerosis (M.S.):

<table>
<thead>
<tr>
<th>Process</th>
<th>Intermediate Outcome</th>
<th>Desired Clinical Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documentation of offer of psychosocial support</td>
<td>• MSQoL-54 Score</td>
<td>• Disability (Days missed from work)</td>
</tr>
<tr>
<td></td>
<td>• Physical functional status</td>
<td>• Improvements to Health-related QOL</td>
</tr>
<tr>
<td></td>
<td>• Mental functional status</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1**: Link from process to intermediate to desired clinical outcome
Guiding Principles for AAN Desired Clinical Outcome Measures

- Outcome measures must be clinically relevant, meaningful, measureable, and actionable by the clinician.

- If assessing desired outcomes are impractical due to feasibility, data collection, availability of longitudinal data, and/or patient tracking then AAN work groups should begin with intermediate clinical outcomes measures\(^{12}\).

- Measure development work groups should distinguish between outcome measures that are around a defined episode (i.e., stroke), around a chronic condition (i.e., epilepsy), or around a deteriorating condition (i.e., dementia). Initially outcome measures will be developed around a clinical disease entity (e.g., Parkinson’s disease) or episode of care (e.g., stroke), rather than across multiple diseases\(^{13}\).

- Selecting desired clinical outcome must be based on a critical evaluation of the evidence base. QMR acknowledges that the evidence to support past quality measure development was rooted in AAN clinical practice guidelines. However, for outcomes measures there may be limited evidence and therefore a limited linkage between outcome measures and AAN’s clinical practice guideline recommendations.
  - There must be evidence supporting specific patient outcomes and those must be the result of patients’ interactions with the health care system.
  - The evidence must indicate with a high-level of confidence that the chosen outcome is appropriate for the disease and is the best measure for the disease.
  - There must be a body of evidence providing a link between the intermediate outcome measure or the desired clinical outcome measure and a process of care\(^{14}\).

- AAN desired clinical outcome and intermediate measures must be fully specified, including\(^{15}\):
  - A statement of the desired patient outcome;
  - The patient population and any exceptions to the population (denominator);
  - The definition of those who achieved the desired outcome (numerator);
  - Measure exceptions;
  - The measurement time window;
  - The frequency that the outcome measure should be calculated;
  - The level of measurement;

\(^{12}\)IBID.


\(^{14}\)IBID.

\(^{15}\)IBID.
• Provider responsible for the patient outcome: Outcome measures may be attributed to individual clinicians, teams of care or systems of care. However, ideally, the primary accountability should be with the ordering clinician who understands the clinical scenario best\textsuperscript{16}; and

• Data source: Measure specifications should clearly state from what data sources the measure components are derived (e.g., administrative data, chart abstraction, electronic health record)\textsuperscript{17}. Clinicians must be able to collect data for the measure. The data should be timely and must not create significant administrative or financial burden to practices. Feedback should be meaningful and ongoing. Data sources must not violate any standards of patient confidentiality.

• AAN desired clinical outcome and intermediate measures must be statistically sound. Measures must have the statistical power to detect variation in quality at the level of care for which they were developed. Measures must be:
  • Valid—they must assess what they measure is intended to.
  • Reliable—they must be repeatable across settings, raters, and patient populations.
  • Responsive—they must detect clinically meaningful changes in patients over time and remains stable when there is no change.

• Risk adjustment methodologies used will be unique to each measure set, depending on the type of measure (e.g., mortality, symptoms, or adverse events)\textsuperscript{18}. The denominator should incorporate dimensions of risk for the outcome, where applicable\textsuperscript{19} and must not lead to gaming or affect patient access to neurological care. To alleviate data burden, AAN measures should seek to avoid complex risk adjustment methodologies\textsuperscript{20}.

**Step 1: Define the desired outcome(s) for the patient population**

Neurology is a specialty of chronic and/or recurrent, incurable, or lethal disease. Due to the persistence and progression of some neurological diseases, measures should allow for outcomes on a continuum that includes cure, relief from disease-related symptoms, slower deterioration, prolongation of survival, and palliative care.

Work groups should define the desired clinical outcomes for the patient population and identify what interventions and services will lead to improved outcomes. Working from the desired outcome, work groups will then define the evidence-based intermediate outcomes and the evidence-based processes that lead to those outcomes. To guide the conversation, the work groups might ask the following questions:

\textsuperscript{16} IBID.
\textsuperscript{17} IBID.
\textsuperscript{18} IBID.
\textsuperscript{19} IBID.
\textsuperscript{20} IBID.
• What is the highest level of desired outcome that a particular patient population might achieve?
  • How can this be stated as a desired clinical outcome?

• Thinking of the typical trajectory of a disease and a patient populations’ response to treatment or non-treatment, what are the inflection points, clinical decision points, or clinical indicators that illustrate quality outcomes?
  • How might these help to identify potential intermediate outcomes measures?

• What services, processes, or interventions will need to be in place to help patients achieve the outcome?
  • How can the process(es) be measured?

• What is the intended use of the clinical measure? What decisions will clinicians, patients, and/or other stakeholders need to make?

**Step 2: Review the Evidence**

Outcome quality measures must be the result of a critical evaluation of the evidence. Some diseases or conditions have very well defined evidence-based guidelines, while other conditions have a smaller evidence base. Ideally, prior to establishing a work group the QMR Subcommittee will base their prioritization and selection of clinical areas for measure development on the availability of scientific literature supporting specific processes of care and desired clinical outcomes.

The work group will review the evidence for care processes, intermediate outcomes, and desired outcomes for a given condition. This includes evidence that the chosen outcome is appropriate, that it is the best measure for the disease or condition in question, and there is explicit evidence that the specific care process will lead to changes in clinical outcome.

National Quality Forum’s (NQF) provides guidance on the evidence needed to support intermediate and desired clinical outcomes measurement (Table 1).

---

21 The AAN’s evidence-base policy outlines the evidence requirements for measure development. Going forward this policy may need to accommodate different approaches to measure development and the evidence that is required for outcome measures. QMR recognizes there are some neurologic conditions for which there is not a high level of evidence either for processes of care, intermediate outcomes, or desired clinical outcomes. In these cases, the outcome measure should be based on the strongest evidence possible and work groups must provide a justification for the minimal link between proposed outcome measures and clinical guideline recommendations.
<table>
<thead>
<tr>
<th>Definition</th>
<th>Level of evidence</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired Clinical Health Outcome An outcome of care is a health state of a patient (or change in health status) resulting from healthcare – desirable or adverse.</td>
<td>Optimally, quantity, quality, and consistency for a body of evidence that the measured outcome (desirable or adverse) is influenced by at least one healthcare process or service. However, outcomes do not necessarily require evidence.</td>
<td>Mortality, Morbidity, Symptoms, Clinical Events Acute Myocardial Infarction 30-day Mortality, Oxford Knee score</td>
</tr>
<tr>
<td>Intermediate Outcome An intermediate outcome is a change in physiologic state that leads to a longer-term health outcome.</td>
<td>Quantity, quality, and consistency of a body of evidence that the measured intermediate clinical outcome leads to desired health outcomes in the target population.</td>
<td>Blood pressure levels, blood glucose levels Hemoglobin A1c Management</td>
</tr>
<tr>
<td>Patient Reported Outcomes Patient experience with care; knowledge, understanding, motivation; health risk status or behavior (including adherence); patient satisfaction or experience with care.</td>
<td>Characteristics for selecting PROS: Conceptual and measurement model, reliability, internal consistency, reproducibility, validity, content validity, construct and criterion related validity, responsiveness, interpretability of scores, burden, and alternative methods of administration. Generally, there is a limited evidence-base supporting patient experience of care and functional status.</td>
<td>Quality of life- health status as perceived by the individual; Functional measures- SF-36, PROMIS ; Pre- and post- treatment physical function; other measures of health status such as pain, vitality, perceived well-being, health risk status, etc. Patient Experience with Care: Consumer assessment of healthcare Providers and Systems (CAHPS); shared decision making; engagement of family and friends; patient knowledge and understanding.</td>
</tr>
</tbody>
</table>


23 The AAN has added an additional row to address the evidence required for patient reported outcomes.
Step 3: Identify the link between outcomes and regular processes of care

The work group should demonstrate through supporting evidence that there is a proximal link between the process of care and the desired clinical or intermediate outcome and that process of care is modifiable by health care providers through intervention. This means that the desired clinical or intermediate outcome relates directly to a specific process or clinical intervention and there is evidence of a causal relationship between the process and the desired clinical or intermediate outcome.

<table>
<thead>
<tr>
<th>Process</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>Intermediate Outcome</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Health/Clinical Outcome</td>
</tr>
<tr>
<td>Recommendation</td>
<td>Patient Reported Outcome</td>
</tr>
<tr>
<td>Intervention</td>
<td></td>
</tr>
</tbody>
</table>

Step 4: Define the type of outcome measure

Based upon the evidence supporting the desired clinical and intermediate outcome(s) for a patient population, work groups will determine what type of outcome measure is most appropriate to develop. Possible types of outcomes measures include:

<table>
<thead>
<tr>
<th>Type of Outcome Measure</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Outcomes</td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td>Post-surgery death rates</td>
</tr>
<tr>
<td>Morbidity</td>
<td>Reduce rates of gestational diabetes</td>
</tr>
<tr>
<td>Intermediate clinical outcomes</td>
<td>Blood glucose levels</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Reduction in pain</td>
</tr>
<tr>
<td>Clinical events</td>
<td>Stroke</td>
</tr>
<tr>
<td>Patient-Reported Outcomes</td>
<td></td>
</tr>
<tr>
<td>Health Status</td>
<td>PROMIS</td>
</tr>
<tr>
<td>Patient experience with care</td>
<td>CG-CAHPS</td>
</tr>
<tr>
<td>Economic Outcomes</td>
<td></td>
</tr>
<tr>
<td>Resource Use/Costs</td>
<td>Cost per episode of care</td>
</tr>
</tbody>
</table>

Table 2: Types of outcome measures

---


25 Ibid.
Step 5: Develop the candidate measure statements for the topic

After the work group identifies the desired clinical outcomes, intermediate outcomes, and process of care and links them to the evidence, the work group will draft candidate measures that define the measurement statement (i.e., numerator, denominator, exceptions, etc.). The work group should consider existing measures for a given condition. Measure specifications must be “harmonized” with other measures. Measure harmonization refers to the standardization of measure specifications and definitions so that they are uniform or compatible, including numerator, denominator, exclusions, data source, and data collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

The work group will precisely specify measures components considering the following criteria:

Specify the denominator

*What is the eligible patient population for this measure that could benefit from the intervention?*

**Inclusion criteria:** Denominator inclusion: diagnosis, diagnostic subgroup, acuity of diagnosis, age ranges, and other positive selection factors.

**Exception criteria** Denominator exceptions: medical, patient, or system causes that would remove a patient from the denominator if the numerator criteria cannot be met. Exceptions should be specific, applicable, and documented and should only include medical reasons (e.g., drug sensitivities, complicating co-morbid conditions), patient reasons (e.g., patient religious preference), and system reasons (e.g., imaging equipment not available, no insurance coverage for procedure).

Specify the numerator

*Which patients demonstrate the desired outcome? What must the clinician do to show they completed the measure?*

The numerator should reflect the clinical recommendation statements from the referenced clinical guidelines and/or other evidence that served as the basis for the measure.

**Time window:** What is the timeframe during which the patient population is expected to benefit from the intervention or to show the considered outcome? Often, the optimal duration can be difficult to assess when some outcomes manifest over long periods, such as the delay of or preventing worsening of a degenerative chronic disease. In these cases, the work group should identify key decision points or clinical inflection points of response to treatment or non-treatment and determine how these points might help to identify potential intermediate outcomes measures.

**Attribution:** To which provider is the desired or intermediate clinical outcome attributed? Most patients with neurologic conditions receive care from multiple clinicians. The work group should consider the appropriate level of accountability by assessing the decision-making responsibilities and utilization for a specific set of services for a given measure concept. This will also help the work group determine the level of measurement
Determine the level of measurement

What level of the health care system is accountable for the outcome?  
What changes to the outcome are in their control?

The level of measurement should include the level of accountability (e.g., individual, team, clinic, in-hospital caregivers, or noninvasive or procedural laboratory) and the lowest level of meaningful data collection. Neurology is often a specialty where practices and/or individual physicians see small numbers of patients with a given condition and one would need a very large sample to detect differences in the quality of care from neurologist-to-neurologist. The work group should ensure the level of measurement and the unit of surveillance results in valid data.

Classify risk adjustment variables and/or risk stratification methods:

Consider how risk would be addressed when writing the measures. Conducting multivariate modeling and regression analysis of risk variables for their relative impact on selected outcomes is time consuming and requires extensive resources.

- Ask whether risk adjustment is necessary. The answer to this question is complex and requires expert opinion from both clinical and epidemiological perspectives.

- Identify the variables or the “dimensions of risk” that will affect the outcome of interest and then devise a risk strategy to address these variables.

- Recommend a model or strategy
  - Risk stratification: Quality of care is assessed independently of patient case-mix. Risk stratification evaluates each patient’s current health status and morbidity burden, comparing this to the average of the population as a whole and stratifying the population into groups who have similar levels of healthcare requirements. Factors such as race, ethnicity, and socioeconomic status can be used for risk stratification. It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences. If a measure is stratified by a particular factor, it cannot then be used as a risk adjustment variable. *Example: Comparison of mammography rates between Caucasian women and African American women.*

  - Risk adjustment: Controls for the potential influence of patient characteristics that are generally not under the control of clinicians that can affect outcomes. The purpose is to “make the playing field level” across settings and sites. So as not to mask disparities in care, factors such as race, ethnicity, and socioeconomic status should not be included in risk adjustment models. *Example: The impact of a patient’s age on surgical outcomes.*

- Identify a list of risk predictive variables for the outcome of interest. Bear in mind that most risk models rely on readily available administrative data that can be used to assess risk factors relating to the patient’s diagnoses, surgical procedures, age, gender, and complications and comorbidities.
• Recommend the use of off-the-shelf methods that are readily available in existing software programs. Major risk adjustment developers and vendors include Johns Hopkins University Adjusted Clinical Groups (ACGs), Milliman (MRA Dx Adjuster, Rx Adjuster), or Verisk Health Diagnostic Cost Groups (DCGs). Each has its own proprietary algorithms and methodology, and offers a large number of application-specific risk adjustment models from which users can choose.

• Provide their rationale for selecting the risk model and methodology and describe its impact on the results for public reporting and quality improvement compared to other methods.

• The work group must also consider the availability of clinical data and its sources for risk adjustment or risk stratification.

Outline the data collection protocol (i.e., data sources, missing data procedures) including from which sources and how data will be collected?

Data used in quality assessment are obtained from diverse sources, such as billing records maintained by insurance companies to reimburse physicians, clinical records maintained by health care professionals, pharmacy and laboratory data, and surveys of patients. Once the clinical measures have been defined, the work group must state how the data should be obtained as well as the protocol for data collection. Potential data sources include:

• **Administrative data** (e.g. billing data, patient characteristics) is readily available and inexpensive to collect. Health care organizations, private insurers, and public insurance programs, such as Medicare and Medicaid, maintain administrative data.

• **Clinical data** or medical record data are the most complete sources of information on diagnosis, treatment, and clinical outcomes. They are, however, expensive to obtain. Medical records usually contain detailed information about a patient’s clinical history, current health status, and information on test results.

• **Clinical research databases** are often valid and reliable, if such databases are available in the chosen clinical area.

• **Survey data** from patients can be used to access attitudes, behavior, knowledge, and functional or other outcomes

Work groups should be attentive to:

• **Data Availability:** Some data elements necessary to assess and improve quality of care are simply not available to those responsible for quality measurement and improvement activities. These data gaps are due to a number of different factors, including the burden of data collection, technology barriers to data collection, and legal and technical barriers within and outside payer and/or care delivery organizations.

• **Data Completeness:** Data completeness refers both to the completion or documentation rates for existing data elements within a record and to the completeness of records in a dataset. Data completeness problems typically apply to medical records, administrative transaction data, and survey data.

• **Data Accuracy:** Certain types of data have been more prone to errors and inaccuracies. Administrative data are prone to accuracy issues including coding of procedures and diagnoses.
• **Data Timeliness:** Lags in administrative data submission and processing have rendered some types of data unusable for time-sensitive quality improvement purposes. Recent acceleration in adoption of EHRs has greatly reduced the lag times between the occurrence of healthcare events or orders and the availability of relevant clinical information for quality assessment and improvement.

*Technical specifications (e.g., coding, e-specifications, e-measures):*

The process for the development of technical specifications is being updated and will be specified in the full Measure Development Process Manual.

**Step 6: Assess measures for potential uses**

Once the measure set is specified, the work group will propose how the measures should be used. This has a significant effect on the AAN’s dissemination of the measure set and potential national endorsement efforts. Work groups, with the assistance of QMR Subcommittee, will rank measures into two tiers: Tier 1 identifies measures for quality improvement only and tier two proposes that the measure be considered for National Quality Forum (NQF) endorsement, public reporting, or accountability programs. Work groups will review the NQF measure evaluation criteria and aim to develop 1–2 measures that meet the NQF measure evaluation criteria. Only measures that meet the criteria and have a high potential to receive a positive initial review from the NQF will be beta tested.

**Step 7: Conduct preliminary testing**

Before outcomes measures are implemented, the work group will test the measures to identify areas that require further specification refinement and definition. The work group will test:

- **Reliability** of a clinical indicator expresses the extent to which repeated measurements by different providers, at different times and places, obtain similar results. Reliability is important for comparing groups or comparing the same group over time. Reliability can be tested as inter-rater reliability where different people or methods provide data on the same indicator. Measuring inter-rater reliability, internal consistency, and test–re-test reliability allows users to determine if the data collection methods are precise enough to provide reproducible results.

- **Validity** determines the degree to which an indicator measures what it is intended to measure. Validity can be tested by confirming that the scores of a measure are linked to specific outcomes, and that the measure can reflect good and bad quality.

### Developing Patient Reported Outcomes Measures

Historically, quality measurement has relied primarily on clinical measures. However, outcomes measures are beginning to include patient-reported outcomes and experience. Numerous programs have adopted patient-reported clinical outcomes and patient-reported experience measures. For example, the Hospital Value-Based Purchasing Program has incorporated 30-day condition-specific mortality measures as well as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) into its measure set, linking
Addendum: Outcomes Measure Development

clinical outcomes and patient-reported experience of care to provider payment. Because of the interest in patient reported outcomes, it is also important that the AAN pursue measuring the impact of neurological care on patient symptoms, functioning, and emotional well-being.

Patient-reported outcomes (PRO) are defined as "any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else." PRO is an umbrella term that covers a range of different types of outcome:

- Functional status: Functional status measures assess patient’s physical limitations (such as disability or activity limitations) and addresses issues that are most indicative of disease severity. These assessments should not be confused with clinical rating scales, where a clinician completes a form to rate disease severity or treatment effects.

- Health-related quality of life (HRQoL): Health related quality of life measures assess how a treatment or disease is affecting a patient psychologically, socially, and physically.

- Patient experience and satisfaction: Patient experience and satisfaction measures assess the patient’s perceptions of the process of treatment rather than its outcome, specifically a patient’s interaction with the health system or the extent to which care is “patient centered.”

The following are the principles and steps by which the AAN’s measure development groups, overseen by QMR, will develop patient reported outcomes measures.

Guiding Principles for AAN Patient Reported Outcome Measures

1) Selecting HR-QOL measures must be based on a critical evaluation of the evidence. QMR acknowledges several issues with PROs including a limited evidence-base. While there is often consensus that measure developers should undertake these activities, there is generally little evidence to support the use of PROs in specific diseases or conditions. However, the choice from among existing PROs should be based on a thorough search of the literature. There should be a body of evidence for the instrument in terms of its appropriateness, reliability, validity, responsiveness, precision, interpretability, acceptability and feasibility. There should be evidence that the construct is important, meaningful, can be improved to measure and the provider can influence any improvements.

2) PRO measures must be psychometrically and statistically sound. PRO measures ought to be based on a validated tool or instrument. In addition to the soundness of the tool, endpoints for patient reported outcomes must be tested to ensure that they are reliable, valid and responsive to change. Instruments and endpoints must be:

---


27 IBID.
• Valid — Assesses what we intended to 
• Reliable – Repeatable across patient populations 
• Responsive – Detects meaningful changes in patients over time and remain stable when there is no change.

3) PROs are prone to bias and error, just like any clinical measure to assess health status. QMR Subcommittee recommends work groups consider potential biases such as patients’ subjective experience, symptom severity and frequency, emotional and social well-being, and perceived level of health and functional ability.

4) Ideally, patient reported measures include a well-defined standardized format, well-documented procedures for administration and scoring, demonstrated reliability and validity, and guidelines for interpretation of scores. AAN patient reported outcome measures must be fully specified, including:

• The desired patient outcome;
• Measure exceptions;
• The measurement time window and the frequency that the outcome measure should be calculated;
• The specific validated tools/questionnaires/surveys, etc.;
• Administration (standard methods, modes);
• Rules for scoring

5) PRO measures must be simple to administer, complete and score. The data should also be timely PRO measures should accommodate several modes of administration and data collection. The data should be timely and must not create significant administrative or financial burden to neurology practices. Data sources must not violate any standards of patient confidentiality.

6) PRO assessment can be attributed to a specific visit or occur between visits. Ideally, the primary accountability for the experience, satisfaction, or health related quality of life of a patient should be with the facility or clinician who understands the clinical scenario best.

7) Patients’ views on content and format should be incorporated during measure development.

**Step 1: Identify the goals and outcomes that are meaningful to the target population.**

Patient-reported outcomes measures are useful to screen for functional problems, monitor disease progression or therapeutic response, improve doctor-patient communications, assess quality of care, or provide case-mix adjustment for comparing other outcomes between patient groups. The work groups, including consumers and patient stakeholders, should specify what change they expect to see in the patient population in terms of function, quality of life, and/or satisfaction.

---


29 IBID.
Step 2: Define actionable outcomes and relevant domains

All potential patient-reported outcomes should be based on evidence (expert recommendations, comprehensive evidence reviews and professional consensus) that the outcome is actionable and amendable to change. Most likely, the work group will select several domains (physical function, social function, role function, mental health, and general health perceptions) that are clinically useful, appropriate to the patient population, and are responsive to clinical intervention.

Step 3: Identify existing patient reported outcome tools or instruments

There are thousands of patient reported outcomes tools available. These tools include instruments, scales or other validated tools that have been tested (reliability, validity, responsiveness, and feasibility) in specific patient populations. Work groups will need to consider:

1) **Generic or disease-specific questionnaires:**
   - Generic instruments are used across different subgroups of individuals, and contain common domains that are relevant to almost all populations. They compare one population to another, or compare scores in a specific population to normative scores. Generic questionnaires may not be sensitive to changes in disease-specific health domains over time.
   - Disease specific instruments measure the patient’s perceptions of a specific disease or health problem. Multiple instruments are now available for common health problems. Disease-specific functional, health status, and quality of life instruments may be more sensitive to specific symptoms experienced by patients, but may miss domains affecting the patient but unrelated to the disease being treated.

2) **Existing or new instruments:** Given the resources required to develop and validate a new instrument, the QMR Subcommittee does not advise that work groups design new instruments for use in AAN’s measure sets. Thousands of PRO instruments and tools have been developed for use in clinical trials to test the efficacy of medications and therapeutics (see Tables 3 and 4). These tools often assess a variety of domains that are relevant to neurology patients. Work groups will review existing tools (e.g., from clinical trials, NeuroQoL, PROMIS) or they can suggest other tools for the work group to consider. Whenever possible, work groups should select reliable and valid instruments or items from validated and reliable “item banks.”

3) **Administrative burden, cost, and copyright:** Integrating patient reported outcome measures into a busy clinical practice is difficult because it require time, thought, recording, and follow-up. Their use should not require complex training or scoring algorithms and measures must be brief, easily incorporated into the clinical routine, and easy to interpret. Another practical consideration is the copyright status of an instrument. Some are entirely in the public domain and are free for use. Others are copyrighted and require permission and/or the payment of fees for use.
4) **Mode of administration**

- Self-administration either at home or in the clinical practice
  - Paper-based - Staff are needed to facilitate the administration process, check for missing data, and calculate scores.
  - Computerized questionnaires are less staff intensive and have a much lower cost.
- Interviewer- and telephone-administered
  - Interviewer facilitated questionnaires usually produce higher response rates and lower levels of missing data than self-administered postal questionnaires.

5) **Timing of administration**: A protocol should be set to ensure consistency in timing for the administration of the instrument. For example, for a surgical instrument, there should be agreement about whether it is done pre-operatively or post-operatively; on procedure day or at pre-admission clinic visit.

6) **Population**: It is important to understand the target population that will be completing the instrument. These may range from individuals who can self-report, to individuals requiring the assistance of a proxy or medical professional. With the potential for patients with cognitive impairment, caregivers or health care providers might give their assessment of the individual’s health. If non-English speaking populations will complete the instrument, it is necessary to have appropriate language/culturally adapted versions. Patients with greater levels of illness or disability are less able to complete lengthy questionnaires.

**Step 4: Specify score interpretation**

Instruments use a variety of scoring methods. At the most basic level, guidelines can provide information on score meaning for example, “higher scores mean better functioning”. Without knowledge of normal ranges, physicians may not know what cut-points of scoring indicate that action is warranted. Interpretations should be incorporated into any improvement tools. Whatever approach is used, training clinicians on the meaning of scores and on approaches to responding to issues raised is critical before implementing the instrument.

**Step 5: Test the measure for reliability and validity**

Measure groups will be encouraged to select from existing tested tools, questions, surveys and/or question banks. The instrument selected should be valid, reliable and internally consistent. Instruments should assess the domains that affect patients with the condition and target those domains that are amenable to change and that prevent a patient from meeting his/her needs.
### Table 3: Examples of existing instruments or item banks

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Key Domains</th>
</tr>
</thead>
</table>
| Patient-Reported Outcomes Measurement Information System (PROMIS) | A network of NIH-funded primary research sites and coordinating centers working collaboratively to develop a series of dynamic tools to measure PROs reliably and validly. PROMIS has developed short-form measures that allows for efficient, psychometrically robust assessment of PROs in clinical research and practice settings | - Physical health  
- Fatigue  
- Pain Intensity  
- Pain Interference  
- Physical Function  
- Sleep Disturbance/Sleep—Related Impairment  
- Pain Behavior  
- Sexual Function  
- Mental Health  
- Anxiety  
- Depression  
- Alcohol use  
- Anger  
- Cognitive function  
- Psychosocial Illness Impact  
- Satisfaction with and Participation in Social Roles and Activities  
- Activities  
- Social Isolation  
- Social Support |
| Neuro-QOL | A multi-site NINDS funded project that developed clinically relevant and psychometrically robust health-related quality of life measures for adults and children with a variety of neurological disorders. | - Ability to Participate in Social Roles and Activities  
- Executive Function  
- General Concerns  
- Anxiety  
- Depression  
- Emotional and Behavioral Dyscontrol  
- Fatigue  
- Lower Extremity Function—Fine  
- Positive Affect and Well-Being  
- Satisfaction with Social Roles and Activities  
- Sleep Disturbance  
- Stigma  
- Upper Extremity Function—Fine  
- Motor-ADL Communication |
<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Key Domains</th>
</tr>
</thead>
</table>
| SF-36v2 | A generic, easily administered, multi-purpose, short-form health survey with only 36 questions developed by RAND Corporation and the Medical Outcomes Study. | • Physical functioning  
• Physical role  
• Pain  
• General health  
• Vitality  
• Social functioning  
• Emotional role  
• Mental health |
| Consumer Assessment of Healthcare Providers and Systems (CAHPS) | A suite of surveys supporting the systematic collection of patient experience data in clinics, hospitals, and long-term care facilities. If patient experience is a desired outcome, work groups are encouraged to recommend from the CAHPS suite of surveys as their data are used in medical home certifications, and the American Board of Medical Specialties has endorsed CAHPS core communication items for Maintenance of Certification. When proposing non-CAHPS surveys the instrument should assess a specific visit, communication with all staff—not just physicians and be administered in a timely manner. | • Access  
• Information  
• Communication  
• Coordination of care  
• Comprehensiveness  
• Self-management support and shared decision making |
Table 4: Health-related QOL domains and subdomains assessed by Neuro-QoL.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Adult</th>
<th>Pediatric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical function</td>
<td>Lower Extremity (Mobility)</td>
<td>Lower Extremity (Mobility)</td>
</tr>
<tr>
<td></td>
<td>Upper Extremity</td>
<td>Upper Extremity</td>
</tr>
<tr>
<td></td>
<td>Fine Motor, ADLs</td>
<td>(Fine Motor, ADLs)</td>
</tr>
<tr>
<td></td>
<td>Sleep Disturbance</td>
<td>Pain</td>
</tr>
<tr>
<td></td>
<td>Fatigue</td>
<td>Fatigue</td>
</tr>
<tr>
<td>Mental function</td>
<td>Depression</td>
<td>Depression</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>Anxiety</td>
</tr>
<tr>
<td></td>
<td>Stigma</td>
<td>Stigma</td>
</tr>
<tr>
<td></td>
<td>Positive Affect and Well-being</td>
<td>Anger</td>
</tr>
<tr>
<td></td>
<td>Applied Cognition – General Concerns</td>
<td>Applied Cognition – General Concerns</td>
</tr>
<tr>
<td></td>
<td>Emotional and Behavioral Dyscontrol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Communication Difficulty</td>
<td></td>
</tr>
<tr>
<td>Social function</td>
<td>Ability to Participate in Social Roles and Activities</td>
<td>Social Relations – Interactions with Peers and Adults</td>
</tr>
<tr>
<td></td>
<td>Satisfaction with Social Roles and Activities</td>
<td></td>
</tr>
</tbody>
</table>