



CMS National Coverage Analysis Evidence Review

Proposed Guidance Document

June 22, 2023

Contents

Preamble	2
Background	3
Methodological Principles	3
Review of Individual Studies	4
Risk of Bias.....	5
Precision of Estimates.....	6
Consistency in Direction of Findings.....	6
Applicability / External Validity.....	6
Strength of Evidence Assessment	7
References	9

Preamble

Section 1862(l)(1) of the Social Security Act (the Act) requires the Secretary of Health and Human Services (the Secretary) to make available to the public the factors that are considered in making national coverage determinations (NCDs) of whether an item or service is reasonable and necessary. The Centers for Medicare & Medicaid Services's procedures for issuing guidance documents under this authority are set forth in 69 Fed. Reg. 57325 (September 24, 2004).

NCDs concerning whether a particular item or service is reasonable and necessary under section 1862(a)(1)(A) of the Act are based on information including clinical experience, and medical, technical, and scientific evidence.¹ The NCD process also considers public comments. The public is afforded the opportunity to comment on a proposed determination as set forth in section 1862(l) of the Act. When we make an NCD, we provide a clear statement of the basis for the NCD as well as responses to the comments received from the public.

To encourage innovation and accelerate beneficiary access to new items and services, CMS is proactively publishing this proposed guidance document to provide a framework for more predictable and transparent evidence development.

This proposed guidance represents the Centers for Medicare & Medicaid Services' (CMS') current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind CMS or the public. Where warranted by circumstances, CMS may consider an alternative approach if it satisfies the requirements of the applicable statutes and regulations. Individuals interested in discussing an alternative approach are encouraged to contact CAGInquiries@cms.hhs.gov and reference this guidance.

Questions may be submitted to CAGInquiries@cms.hhs.gov.

CMS is seeking public comment on this proposed National Coverage Analysis (NCA) Evidence Review Guidance. CMS will review the public comments and respond to the comments in the final document.

Alternatively, written comments may be submitted to the Coverage and Analysis Group, Centers for Medicare and Medicaid Services, mailstop: S3-02-01, 7500 Security Blvd. Baltimore, MD. 21244. Please refer to this proposed guidance document when submitting comments.

To ensure consideration, comments must be received by **August 21, 2023**.

¹ § 1862(a) in the material following (25). (“[I]n making the [national coverage] determination, the Secretary has considered applicable information (including clinical experience and medical, technical, and scientific evidence) with respect to the subject matter of the determination[.]”)
https://www.ssa.gov/OP_Home/ssact/title18/1862.htm

Background

When making NCDs, CMS generally evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service falling within a benefit category is reasonable and necessary for the diagnosis or treatment of an illness or injury or to improve the functioning of a malformed body member under section 1862(a)(1)(A) of the Act. The overall objective for the critical appraisal of the evidence is to determine to what degree we are confident that the specific assessment questions raised in an NCA can be answered conclusively.

When conducting NCAs for an item or service under the reasonable and necessary statute, CMS generally makes three kinds of assessments: (1) The quality of relevant individual studies; (2) What conclusions can be drawn from the body of the evidence on the direction and magnitude of the intervention's potential harms and benefits; and (3) The generalizability of findings from relevant studies to the Medicare beneficiary population.

Methodological Principles

The methodological principles described below represent a broad framework of the issues we consider when reviewing clinical evidence. However, it should be noted that each NCD has its unique methodological aspects.

Methodologists have developed criteria to assess the weaknesses and strengths of clinical research. Strength of evidence generally refers to the scientific validity of study findings regarding causal relationships between health care interventions and health outcomes, including the reduction of bias.

In general, some of the methodological attributes of clinical studies that are associated with stronger evidence include the following:

- Use of randomization (in allocation of patients to either an intervention or a control group) to reduce bias.
- Use of contemporaneous control groups (rather than historical controls) to ensure comparability between the intervention and control groups.
- Prospective (rather than retrospective) studies to ensure comparability through a more thorough and systematic assessment of factors related to outcomes.
- Larger sample sizes in studies to help ensure adequate numbers of patients are enrolled to demonstrate both statistically significant and clinically meaningful improvements in health outcomes.
- Masking (blinding) to ensure, at a minimum, patients and investigators do not know to which group patients were assigned (intervention or control). Blinding is especially important for subjective outcomes, such as pain or quality of life, where beliefs about an intervention may lead to a perceived outcome improvement by either the patient or investigator.

For both interventional and observational study designs, methodological rigor is needed to support causal inference – that is, the extent to which any differences in the health outcomes of interest in the intervention group versus the control group can be properly attributed to the intervention studied. This is known as internal validity. Various types of bias can undermine internal validity. (Cochrane, 2022; Philips et al., 2022) These include:

- Systematic differences in characteristics between patients participating and those theoretically eligible for study but not participating (selection bias);
- Deviations (e.g., co-interventions or different care provisions) from the intervention under evaluation (performance bias);
- Differential collection of data or assessment of outcomes in study groups (detection bias);
- Systematic differences between study groups in the number and reasons that participants do not complete or withdraw from a study (attrition bias); and
- Systematic differences between published versus unpublished study findings (reporting bias).

In addition, confounding is a threat to the internal validity of a study, resulting in statistical associations that suggest a causal relationship between an intervention and outcome when in fact such a relationship does not exist. Confounding occurs when risk factors for the outcome of interest are systematically different between the intervention and control group, which makes it difficult to disentangle the extent to which the outcome is “caused” by the intervention (vs. a confounding factor). CMS carefully considers how potential confounding factors are accounted for in observational and, in some cases, randomized controlled trials (e.g., use of appropriate study design and statistical methods). For example, to interpret and generalize conclusions to the Medicare beneficiary population, studies may need to match or stratify their intervention and control groups by patient age, co-morbidities, disabling conditions, etc.

Methodological rigor is a multidimensional concept related to a clinical study's design, analysis, and conduct. Thorough documentation of a study, particularly the patient selection criteria, the data collection process, and the attrition rate, is essential for CMS to adequately assess the evidence produced. Additionally, when reviewing individual studies, CMS carefully considers both the funding source and potential conflicts of interest for study investigators.

Review of Individual Studies

CMS NCDs have historically been based on a systematic review of findings reported in peer-reviewed literature. However, high-quality findings from other publicly reported results, such as pre-market studies the FDA has reviewed, may also be used. In its review of the evidence, CMS considers the direction and magnitude of study findings, as well as the balance of harms and benefits implied by those findings. Additionally, CMS considers the quality of the overall body of reviewed evidence, which speaks to the confidence or certainty of conclusions that can be drawn from that evidence.

CMS endorses the concept that studies should be fit-for-purpose (FFP). That is, the study design, analysis plan, and data source(s) should be sufficient to credibly answer the question(s) it intends to answer. In general, but not absolutely, the hierarchy of evidence dictates that randomized controlled trials (RCTs) represent the most credible evidence because they are the least subject to biased estimates of outcomes; even so, observational studies can be more representative of actual clinical practice and may answer questions that RCTs cannot answer. Newer study design approaches and analytic techniques that handle bias continue to evolve and may improve the reliability and validity of observational study results. FFP observational studies aim to emulate the strengths of RCTs by taking advantage of these newer approaches and techniques. ([Hernan 2016](#)) Coverage with Evidence Development (CED) observational studies may also meet high standards of credibility through a review of study proposals with AHRQ and CMS before study execution, complete transparency, faithful execution, and clear public reporting.

CMS generally judges the evidence according to the four criteria recommended by the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) Working Group: Risk of Bias, Precision (95% confidence interval), Consistency (in the direction of findings), and Directness (sometimes referred to as *applicability*). ([Guyatt 2008a](#); [Guyatt 2008b](#); [GRADE Home](#)) These criteria are also consistent with principles recommended by AHRQ. ([Berkman-AHRQ 2015](#); [Higgins-Cochrane 2022](#))

The following criteria are generally applied to the body of evidence for health outcomes of interest:

Risk of Bias

In most cases, a credible study design involves comparing a treatment group to one or more comparison groups. Risk of bias, in the context of the methodological quality of studies, refers to the possibility that study design and conduct differentially affects one or the other of the two patient groups being compared. ([Page, 2018](#)) This occurs, for example, when patient characteristics, treatment circumstances, or measurement/data collection differ in ways that may affect estimates of treatment effects. This imbalance detracts from a study's internal validity and reduces confidence that differences in measured outcomes are attributable to the treatment under investigation. A body of evidence based on studies with a considerable risk of bias would generally be considered weak evidence. The risk of bias in individual studies can be minimized by study features such as randomized treatment assignment; for nonrandomized studies matching or balancing between treatment and control groups as well as application of the target trial design approach ([Hernan, 2016](#)) and new users designs ([Franklin & Schneeweiss, 2017](#)); and efforts to prevent a substantial difference in study completion rates. Efforts to maintain data integrity, such as testing variable definitions or handling missing data, are also ways to reduce bias. ([PCORI, 2021](#)) When designing new studies, investigators should keep in mind that CMS uses these tools to assess the threats to internal validity both in published studies and in study protocols submitted

under the CED program: CLARITY in Randomized Controlled Trials ([Guyatt 2011](#)) and USPSTF Criteria for Assessing Internal Validity of Individual Studies. ([USPSTF 2017](#))

Precision of Estimates

Precision refers to variance in the estimate of (treatment) effect. Precision is typically judged based on the width of the confidence interval and, more importantly, whether that interval encompasses values that suggest no meaningful effect and values that indicate a meaningful effect. A wide confidence interval does not permit a confident conclusion regarding the effects of treatment. A key determinant of precision is often the convergence of effect estimates across multiple individual studies or a meta-analysis of numerous studies investigating a particular causal relationship between an intervention and an outcome. Meta-analyses can sometimes provide helpful evidence on overall precision from several studies but cannot substitute for close analysis of individual studies.

Ensuring adequate sample size so that a study has sufficient power to detect clinically meaningful outcome differences between the treatment and control group, with acceptable precision (e.g., acceptably narrow confidence interval) is also important in evaluating studies. The precision of effect estimates for individual studies is generally a function of the sample size, attrition rate, event rate, and the magnitude of change expected for each outcome. CMS recommends that study proposals with a comparison group include a power analysis. While non-comparative studies may not be as useful, they may help demonstrate that treatments can be provided safely in particular settings, and they may allow for longer-term follow-up than is often possible in RCTs. For example, a case series may demonstrate that procedures can be safely performed in a community hospital and may indicate that a device continues to function within acceptable limits with longer-term follow-up. The appropriate minimum sample size for a case series depends on context.

Consistency in Direction of Findings

The reproducibility of studies and their findings is a major principle that underpins the scientific method. CMS can draw more confident conclusions about effectiveness when multiple studies report findings in the same direction for a particular health outcome. Substantial inconsistency in the direction of results or the magnitude of effect estimates may weaken the strength of evidence for a conclusion.

Applicability / External Validity

In making NCDs, CMS generally considers the applicability, or generalizability, of study findings to the clinical situation of interest. Even well-designed and well-conducted trials may not supply the relevant evidence if the results of a study do not apply to the Medicare beneficiary population. Evidence that provides accurate information about a

population or setting not well represented in the Medicare program might be considered, but their applicability may suffer from limited generalizability.

RCTs may have limited generalizability to the Medicare population because of small sample sizes, limited inclusion of Medicare-eligible patients, insufficient enrollment of women and underrepresented portions of the Medicare beneficiary population, or study inclusion and exclusion criteria not reflective of the Medicare population. (National Academies of Sciences, 2022) When assessing applicability, CMS generally considers whether the studied population was representative of the Medicare beneficiary population (e.g., age, sex, race/ethnicity, the severity of disease, presence of co-morbidities, and disability status) and whether the comparison group received treatment that credibly reflects current practice (e.g., dosage, timing, and route of administration; co-interventions or concomitant therapies).

The level of care and the providers' experience in the study are also important elements in assessing a study's external validity. Trial participants in an academic medical center may receive more or different attention than is typically available in non-tertiary settings. For example, an investigator's lengthy and detailed explanations of the potential benefits of the intervention, use of advanced testing, or access to specialty care may point to positive results that may not be consistently replicated in the community setting.

CMS routinely considers studies that are performed in whole or in part outside of the United States (OUS). Whether outcomes from OUS studies may be generalized to the Medicare beneficiary population depends on multiple factors, but an important consideration is whether the study outcome depends on the care delivery context. To the extent that health systems and practice standards differ between countries, an OUS study designed to show positive results may not be generalizable to the Medicare beneficiary population. For example, an OUS study that aims to demonstrate that an intervention reduces hospitalizations may not be generalizable to the US if there are substantial differences in the types of (and coverage provided by) health insurance, hospital bed availability, and practice patterns between the US and the study country. Studies that include outcomes that may be sensitive to care delivery context (whether across different sites in the US or multi-country studies) should be appropriately designed and analyzed, potentially incorporating clustering or stratification into their statistical analysis plan.

Strength of Evidence Assessment

In making NCDs, CMS considers the totality of the evidence across multiple dimensions, including study design and conduct. The evidence for some outcomes, populations, or clinical settings may be of higher quality than evidence for others. Additionally, when CMS reviews evidence for NCD reconsiderations, CMS-approved CED studies may generally be more persuasive than other observational studies because the study design, analysis plan, and data sources will generally be prespecified and posted on clinicaltrials.gov. Reporting study results offers an assurance of quality because, generally, public access to information incentivizes a higher level of accountability in the accurate reporting of the clinical study protocol and results, and in

the conduct of the trial itself. This accountability derives both from public access to information about studies and from the potential risk of penalty for submitting false or misleading clinical trial information in some trials.² Case series and case reports generally have the lowest evidentiary value, and CMS does not typically focus on evidence in these categories.

An intervention's benefits should generally be clinically meaningful and durable rather than marginal or short-lived. When making NCDs, CMS generally places greater emphasis on health outcomes important to patients and their caregivers, such as quality of life, functional status, duration of disability, morbidity, and mortality, and less emphasis on outcomes in which patients often have a less direct interest, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses.

In reviewing the evidence base, CMS aims to make well-founded judgments about the evidence and clearly link it to coverage policy. The direction, magnitude, and consistency of the risks and benefits across studies are important considerations. The evidence is graded for the most important outcomes, and CMS generally conducts qualitative syntheses when drawing conclusions. Based on the analysis of the strength of the evidence, CMS typically assesses the relative magnitude of an intervention or technology's harms and benefits to Medicare beneficiaries. Generally, an intervention is not reasonable and necessary if its harms outweigh its benefits.

² See e.g., Public Health Service regulation at 42 CFR 11.6..

References

Higgins, J. P. T. et al. (2019). Cochrane Handbook for Systematic Reviews of Interventions, Wiley.

Phillips, M. R., et al. (2022). "Risk of bias: why measure it, and how?" Eye **36**(2): 346-348.

Hernan, M. A. and J. M. Robins (2016). "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available." Am J Epidemiol **183**(8): 758-764.

Guyatt, G. H., et al. (2008). "What is "quality of evidence" and why is it important to clinicians?" BMJ **336**(7651): 995-998.

Guyatt, G. H., et al. (2008). "GRADE: an emerging consensus on rating quality of evidence and strength of recommendations." BMJ **336**(7650): 924-926.

GRADE Working Group. Accessed at: <https://www.gradeworkinggroup.org/>. (Accessed 03/13/23)

Berkman, N. D., et al. (2015). "Grading the strength of a body of evidence when assessing health care interventions: an EPC update." J Clin Epidemiol **68**(11): 1312-1324.

Page, M. J., et al. (2018). "Assessing risk of bias in studies that evaluate health care interventions: recommendations in the misinformation age." Journal of clinical epidemiology **97**: 133-136.

Franklin, J. M. and S. Schneeweiss (2017). "When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?" Clin Pharmacol Ther **102**(6): 924-933.

Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee. The PCORI Methodology Report. PCORI; 2021.

Guyatt, G. H., et al. (2011). "GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias)." J Clin Epidemiol **64**(4): 407-415.

Appendix VI. Criteria for Assessing Internal Validity of Individual Studies. U.S. Preventive Services Task Force. July 2017. (Accessed 03/13/23)

National Academies of Sciences, E. and Medicine (2022). Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups. Washington, DC, The National Academies Press.