Evidence-based Health Policy: Literature Review Methodology for Local Coverage Determinations

Background

A local coverage determination (LCD), as defined in §1869(f)(2)(B) of the Social Security Act (SSA, known hereafter as the Act), is a determination by a Medicare Administrative Contractor (MAC) regarding whether a particular item or service is covered on a contractor–wide basis in accordance with section 1862(a)(1)(A) of the Act. The 21st Century Cures Act of 2016 added language to section 1862(l)(5)(D) of the Social Security Act directing the Secretary of the Department of Health and Human Services (HHS) to improve the transparency of the LCD development process.[1] The HHS Secretary promulgated this directive into regulation by revising Chapter 13 of the Program Integrity Manual (PIM). As a result, MACs must appraise the literature and develop LCDs that include a summary of the evidence appraisal informing the decision for coverage. This article provides stakeholders further insight into First Coast Service Options' (FCSO) evidentiary review process in support of more predictable and transparent evidence-based health policy development. This document is a user-friendly guide to assist stakeholders' ongoing questions related to evidence appraisal and is a companion to the CAG issued guidance finalized August 7, 2024.

Methodology

Evidence-to-Decision Framework

FCSO has formally adopted an evidence-to-decision (EtD) framework that offers a transparent, structured process for translating evidence into robust insights that influence health policy decisions. The GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) methodology was selected after the recognition of several key attributes that support policy development and facilitate MAC compliance with the 21st Century Cures Act and PIM, Chapter 13 requirements. GRADE methodology is a widely accepted system adopted by academic, government, and non-government organizations as one of the more effective rubrics for evaluating the certainty of evidence influencing coverage decisions in healthcare.

Systematic and Comprehensive Evaluation

The GRADE methodology facilitates a systematic approach to evaluating the certainty of evidence. It assesses evidence based on several factors, including study design, consistency of results, directness of evidence, precision of estimates, and the risk of bias. By incorporating these factors, GRADE methods ensure all relevant aspects of the evidence are considered when making coverage decisions.

Facilitating Evidence-Based Decision Making

The primary advantage of the GRADE system is its facilitation of evidence-based decision making. By providing a clear, transparent, and systematic evaluation of evidence, incorporating GRADE supports health policy makers in developing well-informed coverage decisions. This evidence-based approach is essential for developing effective and efficient health policies that improve health outcomes and ensure the best use of taxpayer resources (e.g., Medicare Trust Fund).

Transparency and Clarity

Another significant advantage of the GRADE methodology is its transparency. The criteria for rating the certainty of evidence are explicitly stated, which enhances the clarity and reproducibility of the evaluation process. Stakeholders are provided the rationale behind each determination, making it easier to understand the basis for decisions. This transparency also facilitates better communication with healthcare professionals, patients, and the public, fostering trust and acceptance of health policies.

Focus on Patient-Centered Outcomes & Balancing Benefits and Harms

A key strength of the GRADE methodology is its emphasis on patient-centered outcomes. It prioritizes outcomes that are important to patients, such as quality of life, symptom improvement, and functional status, rather than solely focusing on clinical or proxy endpoints. This patient-centered approach aligns health policy decisions with the needs and preferences of the population, leading to more relevant and effective health interventions. The GRADE system emphasizes the importance of balancing the benefits and harms of healthcare interventions. It systematically evaluates both positive and negative outcomes, helping policy makers consider the full spectrum of health technology's impact, leading to more nuanced health policy decisions.

Improving Health Equity[2-5]

Use of the EtD framework assists in identifying and addressing health disparities by systematically considering the impact of interventions on different population groups (e.g., rural (resource poor) vs. suburban (resource rich)). The focus on equity ensures that health policies do not disproportionately benefit certain groups while neglecting others. By promoting fair and equitable healthcare, well-informed LCDs contribute to the overall goal of improving health outcomes for all populations.

Enhancing Consistency and Comparability

By using a standardized approach, GRADE methodology allows for the comparison of evidence and coverage decisions across various conditions, diagnostics, and interventions. This consistency is vital for FCSO health policy makers, who often need to prioritize and allocate healthcare resources (e.g., diagnostics and/or interventions) across multiple health issues.

Enhancing Stakeholder Involvement

In addition to using the GRADE methodology, FCSO may include members from our Contractor Advisory Committees (CACs), clinicians, researchers, Medicare beneficiaries, patient advocates, and any interested parties in the decision-making process. By incorporating diverse perspectives, EtD ensures that insights are comprehensive and considerate of different viewpoints. This inclusiveness enhances the relevance and acceptability of health policies, leading to greater stakeholder support and implementation success.

Assessing Individual Studies

FCSO may apply distinct methodological strategies for coverage determinations depending on the service assessed (e.g., diagnostic versus therapeutic services). However, we systematically utilize GRADE methodology to assess the certainty of individual studies, evaluating factors such as certainty of evidence, magnitude of effect, potential benefits and harms, and relevance to the Medicare beneficiary population.

In principle, the traditional hierarchy of research design is based on the ability of each design to minimize biases. Although there is some debate about the specific order within this hierarchy, there is consensus that randomized controlled trials (RCTs) and systematic reviews of RCTs are highly preferred. This is because RCTs are uniquely effective in minimizing bias through random treatment assignments and controlling for unknown or unrecognized variables. Numerous examples demonstrate starkly contradictory findings between lower-level evidence (e.g., small RCTs or studies evaluating surrogate outcomes) and large RCTs or meta-analysis of large RCTs. For example, in a study of 12 patients evaluating milrinone, the authors demonstrated improvement in left ventricular function during exercise. However, a RCT of 1088 patients and a meta-analysis of several RCTs demonstrated a 28% relative increase in mortality with milrinone compared to placebo. The contrast can be even more stark when comparing observational studies and RCTs.[6]

While RCTs are powerful tools, they are not without limitations. Randomization alone does not guarantee that the results accurately reflect the truth, and there are many potential procedural pitfalls in the design and execution of RCTs. Therefore, not only is the hierarchy of research design considered, in policy decision-making, but also other factors, such as the severity and clinical importance of outcomes, consistency of results, risk of bias, and other methodological features, too. Consequently, RCTs, often considered level 1 evidence, may not always provide sufficient certainty to instruct policy decisions on their own, while alternative research designs may in fact support confidence in the investigator's conclusions due to the methodological rigor employed. The GRADE methodology enables an equitable analysis of various trial designs by considering key research features such as observational versus experimental, prospective, or retrospective, double-blinding, randomization techniques, sample sizes, follow-up procedures, handling of missing data, attrition rates, population characteristics (list not all-

inclusive). Analyzing these factors using the EtD framework helps determine the certainty of literature reporting net health outcome(s).

Certainty of Individual Studies

Risk of Bias

Risk of bias refers to the possibility that the design, conduct, or analysis of a study has introduced systematic errors or deviations from fact, which can distort the study's findings. This distortion reduces internal validity and confidence (i.e., certainty) that differences in measured outcomes are attributable to the treatment under investigation.[6] This means the study's findings may not accurately reflect the true effects of the intervention or diagnostic being investigated. As a result, a body of evidence based on studies with a considerable risk of bias would generally be considered weak or unreliable evidence.[7,8]

Individual studies that employ the following strategies to minimize biases are given more weight when incorporated in coverage decisions:

1. Study Design[9,10]

    a. **Randomization** – Randomly assigning participants to different study groups helps ensure that the groups are similar in all respects except for the intervention or diagnostic. Using stratified randomization ensures that certain characteristics (e.g., age, gender) are evenly distributed across groups, further reducing selection bias. This minimizes selection bias by preventing systematic differences between groups.
    b. **Blinding** – Blinding participants, investigators, and outcome assessors to the group assignments can reduce performance and detection biases. Double blinding (blinding both participants and researchers) is particularly effective.
    c. **Control Groups** – Including a well-defined control group (e.g., placebo or standard treatment) optimizes objective comparisons with the technology under review and helps reduce bias when interpreting its effects.
    d. **Prospective Design** – Conducting a study prospectively helps reduce bias associated with selective reporting and recall. Moreover, the temporal sequence of introducing the technology before the observed outcome is critical for proving that it caused the outcome, rather than merely being associated with it.

2. Participant Selection

    a. **Inclusion and Exclusion Criteria** – Clearly defining who can and cannot participate in the study helps ensure that the study population is representative of the target population. This reduces selection bias.

b. **Reduces Confounding Variables** – Clear selection criteria help minimize variability among participants, ensuring that differences in outcomes are more likely due to the technology under investigation rather than extraneous factors, thus increasing the validity and confidence of the findings.

3. Data Collection[11,12]

    a. **Standardized Protocols** – Using standardized procedures for data collection minimizes variations and ensures consistency across study sites and investigators, reducing performance and detection biases.
    b. **Objective Outcome Measures** – Whenever possible, use objective and validated measures (e.g., lab results, imaging studies) rather than subjective assessments to reduce detection bias.
    c. **Complete Data Collection** – Ensuring that data is collected from all participants reduces the risk of attrition bias, where differences in dropout rates could skew results.

4. Data Analysis[13,14]

    a. **Intention-to-Treat Analysis** – Analyzing participants in the groups to which they were originally assigned, regardless of whether they completed the intervention, helps mitigate bias related to dropout or non-compliance.
    b. **Pre-Specified Analysis Plan** – Defining the analysis plan before the study begins and sticking to the plan helps reduce reporting bias by preventing selective analysis or data dredging.

5. Reporting and Publication[15-18]

    a. **Comprehensive Reporting** – Following guidelines like CONSORT (Consolidated Standards of Reporting Trials) ensures that all relevant aspects of the study are reported, including methods, results, and potential limitations. This reduces the risk of reporting bias.[10]
    b. **Transparency and Registration** – Registering the study protocol in a public database before the study begins (e.g., ClinicalTrials.gov) promotes transparency and helps prevent selective reporting of outcomes.

6. External Review and Replication

    a. **Peer Review** – Submitting the study to peer-reviewed journals allows independent experts to scrutinize the methodology and results, helping to identify and correct potential biases.
    b. **Replication Studies** – Independent reproducibility of study results help confirm the findings and reduces the impact of any single study's bias.

Real-World Evidence

Real-World Evidence (RWE) is defined as clinical evidence regarding the usage, benefits, and risks of medical products derived from the analysis of real-world data (RWD), namely data routinely collected from a variety of sources.[19] This is in comparison to randomized controlled trials (RCTs), whose data originates from highly controlled environments. While RCTs are traditionally emphasized due to their high validity, real-world evidence (RWE) provides valuable insights from real-world settings that can complement the evidence from RCTs used to guide local coverage determinations. RWE often includes more diverse populations than those typically enrolled in RCTs, providing data on subgroups that may be underrepresented in clinical trials. Additionally, RWE can enhance generalizability of recommendations by providing evidence from routine clinical practice, as well as feasibility of integrating the subject of inquiry into real-world health systems. RWE also can offer insights into long-term safety, adherence, and effectiveness that RCTs with shorter follow-up periods might miss. In limited instances, the United States Food and Drug Administration (FDA) has accepted RWE to support drug product approvals, primarily in the setting of oncology and rare diseases when assignment to a placebo or non-treatment (control) arm is unethical or unfeasible. Typically, this may occur when the effect size is expected to be large based on preliminary data.[19] Therefore, while lower in the hierarchy of influencing health policy decisions, RWE has a significant role in the body of evidence used to appraise health procedures and technology.

Causality

Regardless of whether the design of a study is an RCT, a non-RCT, a cohort study, or a case-control study, the primary criterion for methodological strength or certainty of outcomes is defined in the extent to which differences between intervention and control groups can be attributed to the intervention or diagnostic studied. When there are only associations but not causal relationships between a study's variables and outcomes, it is important not to draw causal inferences. Confounding refers to independent variables that systematically vary with the causal variable. [9]This distorts measurement of the outcome of interest because its effect size is mixed with the effects of other extraneous factors. For observational, and in some cases randomized controlled trials, the method in which confounding factors are managed (either through stratification or appropriate statistical modeling) are of particular concern.[9,21]

Effect Size and Direction of the Potential Benefits and Harms

Precision of Estimates

Precision refers to variance in the estimate of (treatment) effect[9]. Precision is typically judged based on the width of the confidence interval. 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. [9]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. 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.[9] 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. FCSO can draw more confident conclusions about effectiveness when multiple studies report findings in the same direction for a particular health outcome.[9] Substantial inconsistency in the direction of results or the magnitude of effect estimates may weaken the strength of evidence for a conclusion.

Generalizability to the Medicare Beneficiary Population

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 may be considered, but its 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. When assessing applicability, FCSO routinely considers whether the studied population was representative of the Medicare beneficiary population (e.g., age, sex, ancestry, ethnicity, the severity of disease, presence of co-morbidities, etc.) 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 notable elements in assessing a study's external validity. [9] 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 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.[9]

FCSO routinely considers studies that are performed in whole or in part outside of the United States (US). Whether outcomes from non-US 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, a non-US study designed to show positive results may not be generalizable to the Medicare beneficiary population. For example, a non-US 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.[9]

Conclusion

The FCSO evidence-based, data-driven LCD development process utilizes a structured transparent approach that relies on high-quality research and rigorous data analysis as well as input from our stakeholders to guide policy decisions and coverage determinations. Using GRADE methodology, FCSO balances benefits and harms, patient-oriented outcomes, decisional transparency, stakeholder contributions, and advocate concerns. The EtD approach ensures that LCDs are developed using the best available evidence, enhancing their effectiveness, efficiency, and equity. By integrating scientific research with real-world evidence, FCSO policymakers can address health challenges more accurately, allocate Medicare resources more effectively, and be sustainably responsive to Medicare beneficiaries.[24-28]

Further details on analysis of evidence review can be found in this CMS document at: https://www.cms.gov/medicare-coverage-database/view/medicare-coverage-document.aspx?mcdid=37.

Reference List

1. Centers for Medicare and Medicaid. Program Integrity Manual, Chapter 13.1.1. February 2, 2019. Accessed July 12, 2024. https://www.cms.gov/regulations-and-guidance/guidance/manuals/downloads/pim83c13.pdf

2. Welch VA, Akl EA, Guyatt G, et al. GRADE equity guidelines 1: considering health equity in GRADE guideline development: introduction and rationale. *J Clin Epidemiol*. Oct 2017;90:59-67. doi:10.1016/j.jclinepi.2017.01.014

3. Akl EA, Welch V, Pottie K, et al. GRADE equity guidelines 2: considering health equity in GRADE guideline development: equity extension of the guideline development checklist. *J Clin Epidemiol*. Oct 2017;90:68-75. doi:10.1016/j.jclinepi.2017.01.017

4. Welch VA, Akl EA, Pottie K, et al. GRADE equity guidelines 3: considering health equity in GRADE guideline development: rating the certainty of synthesized evidence. *J Clin Epidemiol*. Oct 2017;90:76-83. doi:10.1016/j.jclinepi.2017.01.015

5. Pottie K, Welch V, Morton R, et al. GRADE equity guidelines 4: considering health equity in GRADE guideline development: evidence to decision process. *J Clin Epidemiol*. Oct 2017;90:84-91. doi:10.1016/j.jclinepi.2017.08.001

6. Devereaux, P.J. and Yusuf, S. The evolution of the randomized controlled trial and its role in evidence-based decision making. Journal of Internal Medicine, 254: 105-113. https://doi.org/10.1046/j.1365-2796.2003.01201.x

7. Page MJ, Boutron I, Hansen C, Altman DG, Hrobjartsson A. Assessing risk of bias in studies that evaluate health care interventions: recommendations in the misinformation age. *J Clin Epidemiol*. May 2018;97:133-136. doi:10.1016/j.jclinepi.2018.01.004

8. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. Oct 18 2011;343:d5928. doi:10.1136/bmj.d5928

9. Centers for Medicare and Medicaid. CMS National Coverage Analysis Evidence Review. August 7, 2024. Accessed September 18, 2024. https://www.cms.gov/medicare-coverage-database/view/medicare-coverage-document.aspx?mcdid=37

10. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. Aug 2010;63(8):834-40. doi:10.1016/j.jclinepi.2010.02.005

11. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? *BMJ*. May 1 1999;318(7192):1209. doi:10.1136/bmj.318.7192.1209

12. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials*. May 2006;1(1):e9. doi:10.1371/journal.pctr.0010009

13. Gupta SK. Intention-to-treat concept: A review. *Perspect Clin Res*. Jul 2011;2(3):109-12. doi:10.4103/2229-3485.83221

14. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ*. Nov 13 2001;165(10):1339-41.

15. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. May 26 2004;291(20):2457-65. doi:10.1001/jama.291.20.2457

16. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. Aug 18 2009;151(4):W65-94. doi:10.7326/0003-4819-151-4-200908180-00136

17. De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med*. Sep 16 2004;351(12):1250-1. doi:10.1056/NEJMe048225

18. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. *N Engl J Med*. Mar 3 2011;364(9):852-60. doi:10.1056/NEJMsa1012065

19. Framework for FDA's Real-World Evidence Program. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence. (Accessed July 12, 2024)

20. Local Coverage Determination (LCD) Process Modernization Qs & As, Accessed July 12, 2024. https://www.cms.gov/medicare/coverage/determinationprocess/downloads/lcd_qsas.pdf.

21. Higgins, J. P., & Green, S. (Eds.). (2011). Cochrane Handbook for Systematic Reviews of Interventions(Version 5.1.0). The Cochrane Collaboration. Accessed July 12, 2024 https://training.cochrane.org/handbook/current.

22. Drummond, M. F., Schwartz, J. S., Jönsson, B., Luce, B. R., Neumann, P. J., Siebert, U., & Sullivan, S. D. (2008). Key principles for the improved conduct of health technology assessments for resource allocation decisions. International Journal of Technology Assessment in Health Care, 24(3), 244-258.

23. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.AD.

24. Schünemann HJ, Wiercioch W, Brożek J, et al. GRADE evidence to decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT. J Clin Epidemiol. 2017;81:101-110. doi:10.1016/j.jclinepi.2016.09.009.

25. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016;353:i2016. doi:10.1136/bmj.i2016.

26. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ. 2016;353:i2089. doi:10.1136/bmj.i2089.

27. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013;66(7):719-725. doi:10.1016/j.jclinepi.2012.03.013.

28. Alonso-Coello, P., Oxman, A. D., Moberg, J., Brignardello-Petersen, R., Akl, E. A., Davoli, M., ... & Schünemann, H. J. (2016). GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*, 353, i2016. Accessed July 12, 2024 https://www.bmj.com/content/353/bmj.i2016