NATIONAL QUALITY FORUM
Driving measurable health improvements together

Best Practices for Developing and Testing Risk Adjustment Models

WEB MEETING 1 DISCUSSION GUIDE

December 15, 2020

This report is funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I HHSM-500-0000.

Discussion Guide

Purpose: The purpose of this discussion guide is to build on the key elements of the environmental scan. NQF staff have identified key inputs for the TEP to consider but given limited time during TEP convenings, this discussion guide is intended to be the central repository of NQF staff and TEP input. TEP members are encouraged to provide tracked changes, comments for consideration. NQF staff will use this discussion guide to outline areas of consensus and key questions for discussion. This discussion guide will serve as the foundation for the environmental scan report.

Environmental Scan

The environmental scan was conducted using three interrelated approaches. First, a literature review was conducted to identify how risk adjustment model development has considered social or functional risk adjustment. Second, and closely related, an examination of risk adjustment methods used in a sample of performance measures submitted to the National Quality Forum (NQF) for endorsement was conducted. Finally, federal and non-federal value-based performance measurement programs were evaluated to understand how social and/or functional risk was considered. Each of these approaches is outlined below.

Literature Review

Methods

A PubMed search was conducted of available literature published in English from the last six (6) years (since NQF's 2014 report on Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors) to identify studies reporting risk adjustment model development that consider social risk and/or functional status-related risk factors within quality performance measurement. Search terms included a series of terms identified through PubMed's Medical Subject Headings (MeSH), which is the National Library of Medicine-controlled vocabulary thesaurus used for indexing articles for PubMed.[1] The following MeSH terms were used: outcome and process assessment, health care; quality indicators, health care; quality of health care; and risk adjustment.

A reference review was also conducted of NQF's 2014 report on Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors[2] and the 2016 and 2020 Assistant Secretary for Planning and Evaluation (ASPE) reports.[3,4] A forward search of these reports was performed using Google Scholar to identify additional relevant articles. Lastly, NQF consulted experts in the field, including the Technical Expert Panel, to identify additional literature for inclusion.

Studies were screened for relevance based on the following inclusion and exclusion criteria (Figure 1):

Inclusion Criteria:

- Literature focused on U.S. healthcare system
- Literature that included empirical testing Literature focused on risk adjustment of social and/or function status-related risk factors within the context of quality performance measurement

[1] https://www.nlm.nih.gov/bsd/disted/meshtutorial/introduction/index.html

[2] NQF. Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=77474

[3]ASPE. Social Risk Factors and Performance Under Medicare's Value-Based Payment Programs. First Report to Congress https://aspe.hhs.gov/system/files/pdf/253976/RTCAppendices.pdf

[4] ASPE. Social Risk Factors and Performance Under Medicare's Value-Based Payment Programs. First Report to Congress. Second Report to Congress. https://aspe.hhs.gov/system/files/pdf/263676/Second-IMPACT-SES-Report-to-Congress.pdf

- Literature focused on risk adjustment guidance within quality performance measurement

Exclusion Criteria:

- Literature published prior to 2014
- Literature not focusing on or not inclusive of U.S. healthcare system
- Literature focused on approaches to risk adjustment modeling not within the context of quality performance measurement
- Literature not focusing on or inclusive of social and/or functional-status related risk adjustment within quality performance measurement
- Literature that are opinion papers, blogs, comments
- Literature that does not include empirical testing

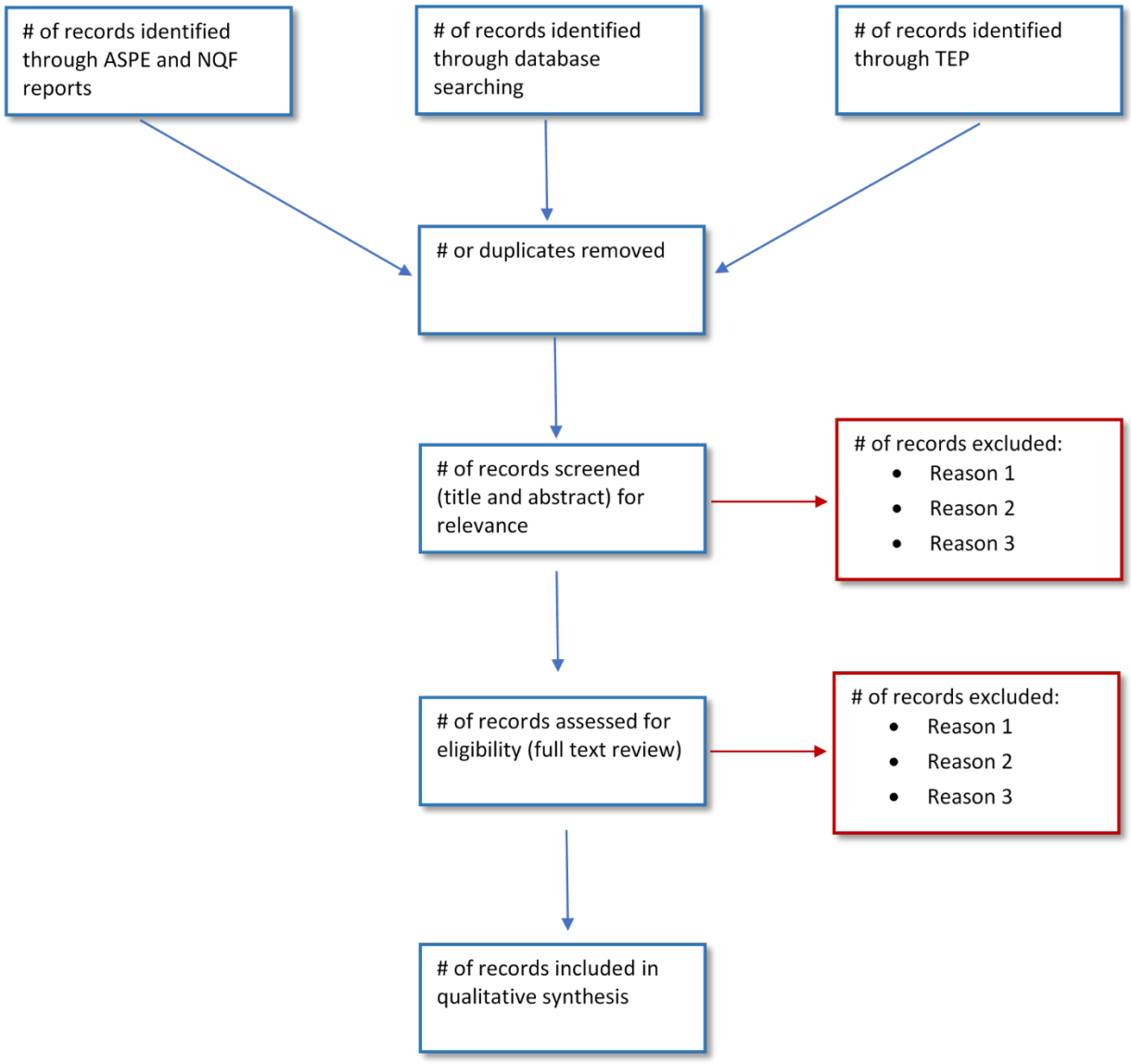
The environmental scan prioritized outcome and cost and efficiency measures. Studies focusing on solely on patient experience as an outcome were not included, as social risk adjustment has been generally accepted in this area of measurement.[5]

After screening for relevance, NQF abstracted data from included studies to capture the following:

1. Datasets used: the datasets used for risk adjustment and measure specifications
2. Functional or social risk factors available for testing: the functional or social risk factors available for use in measure specification and testing
3. Approaches to conceptual and statistical methods: the conceptual and statistical methods and criteria to select patient factors (functional risk or social risk factors) as well as community-, plan-, and facility-level factors.
4. Approaches to inclusion of functional and social risk factors: the myriad of approaches to examining inclusion of risk factors, including but not limited to: prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, and assessment of between-unit effects and within-unit effects

[5] O'Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L, Cleary PD. Case-mix adjustment of the CAHPS Hospital Survey. Health Serv Res. 2005;40(6 Pt 2):2162-2181. doi:10.1111/j.1475-6773.2005.00470.x

Figure 1. Literature Flow Diagram



Discussion Questions 12/15:

1. Is there any feedback on the approach for literature search (i.e., appropriateness of inclusion/exclusion criteria, appropriateness of data elements for abstraction)?
2. In your review of the summary table, are any additional data elements in the results table that should be collected?
3. Are there any other reports or studies that should be considered?

Measure Review

Methods

Measures from the NQF-endorsed measure portfolio and candidate measures submitted for NQF endorsement were reviewed for potential inclusion as illustrative measures for the environmental scan. NQF prioritized illustrative measures with novel or robust approaches to measure testing in NQF measure submissions forms (also called testing attachments) previously submitted to the Consensus Development Process (CDP) projects to identify 10 measures that showcase datasets used, functional or social risk factors available for testing, approaches to conceptual and statistical methods, and finally considerations for inclusion of functional and social risk factors in the final measure specifications.

In all, ten illustrative measures were selected for presentation and analysis of these considerations. Regarding approaches to conceptual and statistical methods, NQF examined the "ordering" of risk factor inclusion (e.g., are social risk factors added before or after all clinical factors). NQF also examined the relationship between functional risk adjustment and social risk adjustment by measure type and intended use.

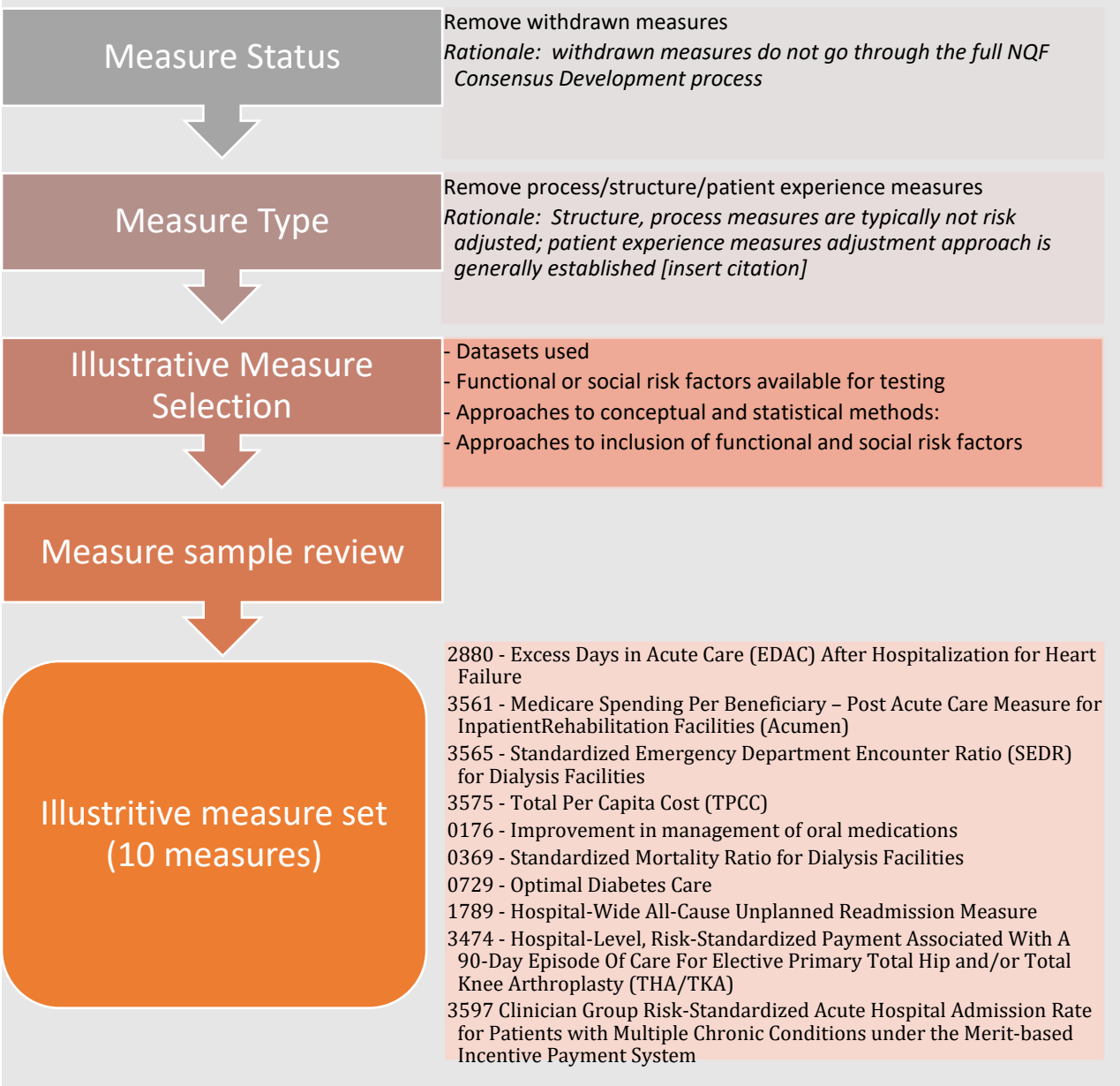
Utilizing the measure selection logic presented below, NQF identified all measures across multiple CDP projects under evaluation in the 2017-2021 NQF social risk trial. First, all measures withdrawn from NQF endorsement consideration were removed from consideration in this project. All process and structure measures were removed from consideration for inclusion in this project since these measures should only be adjusted in particular circumstances.[6] Namely, the process or structure is indicated for all patients within the denominator and adjustment is rarely required.[7]

The environmental scan prioritized outcome and cost and efficiency measures, for which a conceptual rationale for adjustment was demonstrated. NQF conducted a preliminary review of this subset of submitted testing attachments to identify an illustrative set of 10 measures. Figure 2 below illustrates this process of removal and selection.

[6] Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. NQF, 2014. Page 87. Retrieved November 19 from: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=77474

[7] Measures Management System. CMS. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwizi4vEnPjsAhVooXIEHYJoDqsQFjACegQIAxAC&url=https%3A%2F%2Fwww.cms.gov%2FMedicare%2FQuality-Initiatives-Patient-Assessment-Instruments%2FMMS%2FDownloads%2FRisk-Adjustment.pdf&usg=AOvVaw1J7CVsRimaXQA95_TA4QgQ

Figure 2. Measure Flow Diagram



Discussion Questions 12/15:

1. What is your feedback on the approach for inclusion of illustrative measures?
2. In your review of the summary table, are there any additional data elements in the results table that should be collected for the illustrative measures?
3. Are there any other illustrative measures that should be considered?

Program Review

Methods

The environmental scan also examined various federal and non-federal programs[8] that include quality measures for value-based payment and/or public reporting purposes, and the program adjusts or stratified results for social and functional risk factors at the program level, instead of or in addition to adjustment conducted at the individual measure level. The program review does not intend to examine all the federal and non-federal payment or public reporting programs, but rather to identify examples that illustrate different approaches in which social and functional risk factors are accounted for at the program level. NQF seeks input from the federal liaisons and the TEP on potential programs and approaches. Program examples may include:

a. Medicare Advantage Star Ratings Categorical Adjustment Index (CAI). The CAI was introduced in 2017 to address the average within-contract disparity in performance among beneficiaries who receive a low-income subsidy, are dual eligible (LIS/DE), and/or are disabled.
b. CMS Hospital Readmissions Reduction Program (HRRP): hospital performance is assessed relative to the performance of hospitals within the same peer group. Hospitals are stratified into five peer groups, or quintiles, based on proportion of patient that are dual eligible for both Medicare and full-benefit Medicaid. The median ERR of hospitals within the peer group is used as the threshold to assess hospital performance on each measure.[9]
c. Utilizing the program selection and data extraction flowchart (Figure 3) presented below, NQF identified programs that meet the following inclusion criteria:

Inclusion criteria:

- The program is used for value-based payment and/or public reporting
- The program includes quality measures
- The program adjusts or stratifies for social and/or functional risk factors at the program level, in addition to or instead of adjustment at the individual measure level

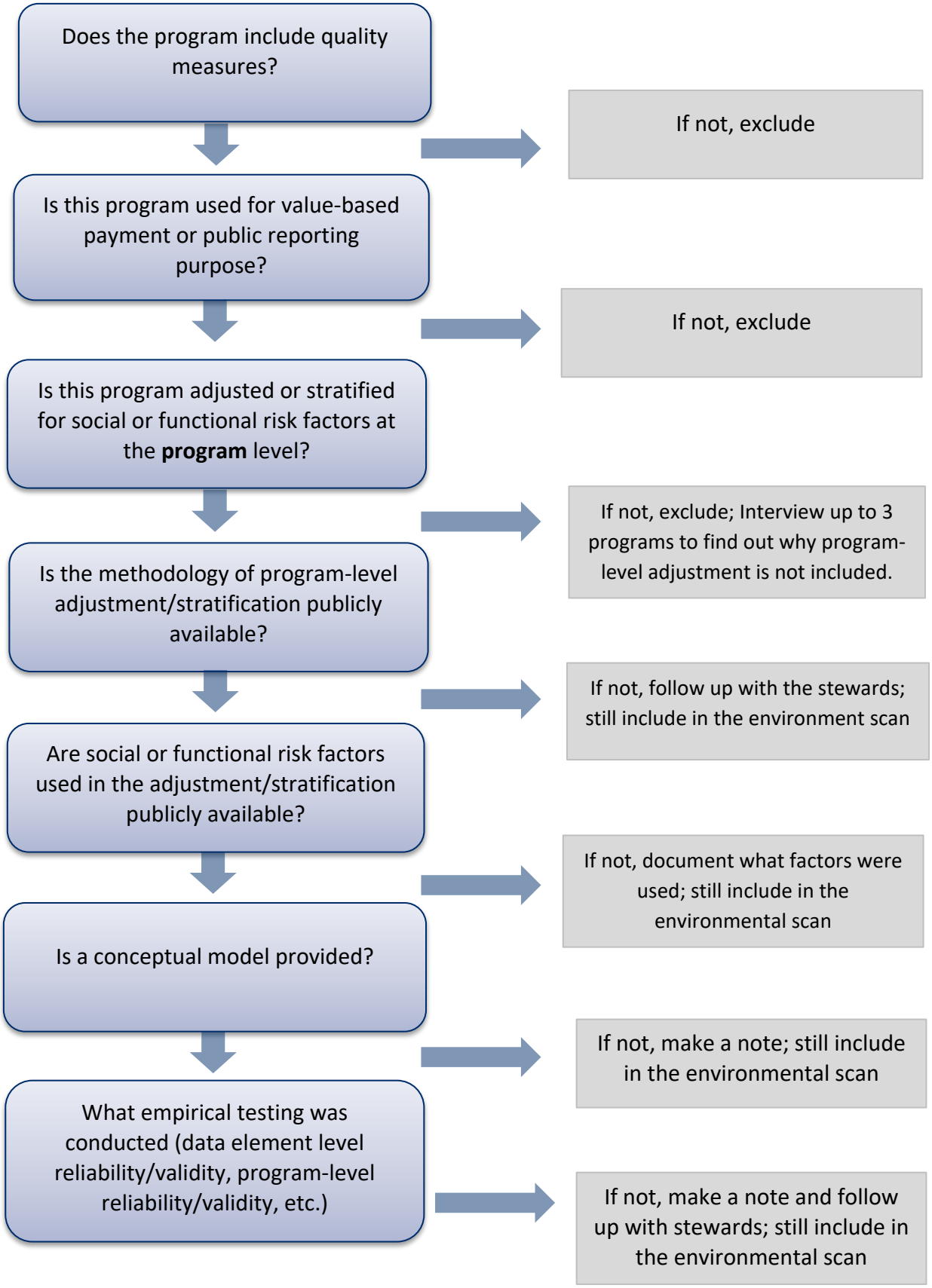
Exclusion criteria:

- The program is only used for internal quality improvement
- The program does not include quality measures
- The program does not adjust for social and/or functional risk factors at the program level

[8] We use "program" in this report in a broad way to refer to any payment or public reporting program, system or model that contains quality measures, such as ACO, Star Ratings, Hospital Compare, etc.

[9] CMS: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program

Figure 3. Flowchart and Data for program review



Discussion Questions:

- What is your feedback on the flowchart's logic?
- Are we capturing the right information in the summary table?
- Are there other non-CMS (both federal and nonfederal) programs that fit into this flowchart?

For future discussion:

- If a program has both payment and public reporting components, should risk adjustment approach vary depending on different purposes? Are there real-life examples?
- If a program adjusts for social and functional risk factors at both the individual measure and the program level, how do we assess whether there is an "over-adjustment?" Is there a preference on which level the adjustment should be conducted?