

McKinsey Healthcare Systems and Services Practice

# Risk adjustment for retrospective episode-based payment: Guiding principles and proposed methodology

Bowen Garrett, Friederike Haass, Josh Haselkorn, Adi Kumar,  
Tom Latkovic, Daniel Tsai, and Tim Ward



## Chapters

1. Introduction .....	1
2. Context for clinical exclusions and risk adjustment .....	2
3. Guiding principles for clinical exclusions and risk adjustment .....	4
4. Clinical-exclusion and risk-adjustment process.....	5
5. Statistical and technical Q&A .....	14
6. References.....	16

# 1. Introduction

Many healthcare payors and providers consider “episodes of care” to be an appealing approach to measure provider performance and reward successful outcomes. An episode of care is any clinical situation with relatively measurable start and end points and with clear patient outcomes, such as procedures, hospitalizations, acute outpatient care, and some treatments for cancer and behavioral health conditions. Recently, momentum has been building for “Retrospective Episode-based Payment” (REBP) as one method to reward providers who consistently deliver high quality and/or favorable costs for specific episodes.<sup>1,2</sup> REBP identifies one or more providers who are in the best position to affect the clinical outcomes and costs associated with an episode of care (herein referred to as Principal Accountable Providers, or PAPs). REBP then assesses (through retrospective analysis of claims data) the outcomes achieved and costs incurred during each episode, on average, over a specific period of time (e.g., quarterly). The PAP(s) for each episode are assessed based on their average performance across episodes they treated over that period, and can be rewarded, penalized, or neither accordingly.<sup>3</sup>

For episode-based performance measurement and payment, including REBP, to function correctly (and fairly), it is critical to account appropriately for clinical factors that affect the expected cost of delivering a specific episode of care for a specific patient. An effective approach should account for the fact that clinical factors affect the cost of different types of episodes to different degrees. An effective approach should also function correctly for dozens of different types of episodes, some of which may occur infrequently. For these reasons, current approaches to population-based risk adjustment (e.g., Johns Hopkins University’s Adjusted Clinical Groups, 3M’s Clinical Risk Groups, UCSD’s Chronic Illness and Disability Payment System) are not adequate for the task.

In this paper, we discuss the challenges posed by accounting for clinical factors in episodes of care and describe a scalable methodology we believe helps fairly measure episode performance even for relatively infrequent, or “low-volume,” episodes.

## 2. Context for clinical exclusions and risk adjustment

REBP relies on accurately, fairly, and consistently measuring the average medical spend for which a PAP is responsible when delivering an episode of care (from here on referred to as “episode spend”). We suggest several mechanisms that may be employed to achieve this goal:

- **Targeted inclusion of spend in the episode:** Include only medical spend judged to be relevant to the episode in the calculation of episode spend. For example, during an asthma acute exacerbation episode, an inpatient admission related to the asthma acute exacerbation would be included, while an inpatient admission because of a broken leg would not.
- **Exclusion of episodes for business reasons:** Take into account only those episodes for which the available claims information is comparable and complete. For example, episodes should be excluded for the following reasons: payment or eligibility rules (e.g., inconsistent enrollment, dual eligibility when only one payor’s claims data is available), select provider characteristics (e.g., certain provider types with special or exceptional payment rules, PAP not identifiable), or missing or exceptional claims information (e.g., long hospitalizations, missing DRG, incomplete claims).
- **Exclusion of episodes of certain patients:** Include only those episodes that involve comparable patients. Exclude episodes of patients who, for example, are long-term-care residents, left against medical advice, died, or fall within certain age groups (unless the episode is defined around patients with these characteristics).
- **Exclusion or winsorizing of episodes with very high episode spend:** Exclude or winsorize episodes with very high episode spend, which may indicate a patient had a unique or unusual event.

In addition, we suggest mechanisms that address clinical factors that affect the cost of delivering an episode of care but for which the PAP cannot be held accountable. There are at least four instances where claims data may indicate that a PAP should not be held accountable for episode spend: (1) some episodes may be affected by unique and/or severe clinical factors; (2) episode spend is too highly variable; (3) there are too few episodes with a given clinical factor to reliably measure the impact of the clinical factor; or (4) episode spend correlates with the presence of clinical factors that are outside the PAP’s control. We propose to address such cases using clinical exclusions and risk adjustment:

- Clinical exclusions remove specific episodes of care from performance calculations when clinical factors are unique and/or severe, make episode spend too highly variable to be reliably compared with other episodes, or if there are too few episodes with a given clinical factor to

reliably measure the effect of the clinical factor. Including these episodes in the calculation of a PAP's average episode spend would put providers at unacceptable financial risk.

- Risk adjustment accounts for different levels of clinical risk that make some episodes of care likely to be more resource-intensive and therefore more costly than others. By isolating and adjusting for clinical factors outside of the PAP's control, risk adjustment allows for comparisons across PAPs with diverse patient panels.

### 3. Guiding principles for clinical exclusions and risk adjustment

To enable correct and fair measurement of episode spend, we believe that clinical exclusions and risk adjustment should follow six guiding principles. They should be:

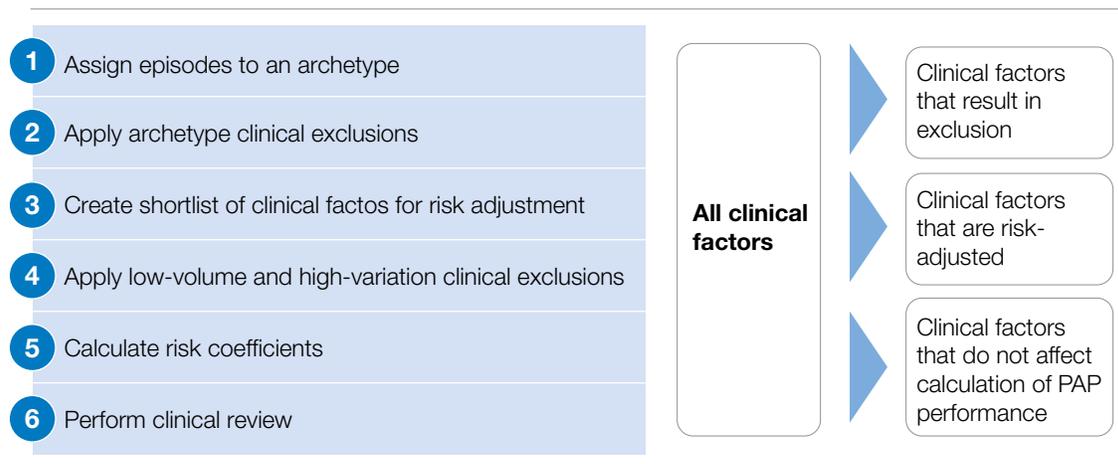
- **Reproducible:** To be practical, the clinical-exclusion and risk-adjustment process should offer a consistent methodology that works across different types of episodes, populations, and payors, even as it accounts for the unique context of the different environments.
- **Tailored by type of episode:** To be effective, clinical exclusions and risk adjustment should fairly reflect the clinical risk presented by a given patient *in the context of a specific type of episode*. For example, diabetes may have more relevant effects on the treatment and risk of a pneumonia episode than an asthma acute exacerbation episode. Each episode type requires its own set of clinical factors that result in exclusion and clinical factors that are risk adjusted.
- **Tailored by payor and population:** To be effective, clinical exclusions and risk adjustment may differ among payors and the context in which an episode-based payment model is implemented, even though the methodology to determine the clinical factors that result in exclusion and clinical factors that are risk adjusted is consistent. For example, differences in payment practices may lead to different risk coefficients for the same risk-adjusted clinical factors in the insured populations of two different payors.
- **Transparent:** To be acceptable, the clinical-exclusion and risk-adjustment process should be well documented and the documentation should be accessible to all stakeholders.
- **Statistically valid:** To be sound, the clinical-exclusion and risk-adjustment process should be based on validated statistical techniques applied in an appropriate and rigorous manner. For example, the process described in this document was developed in collaboration with experts in health economics, actuarial science, and statistics to ensure the required rigor.
- **Clinically valid:** To be meaningful, all inputs and outputs of the clinical-exclusion and risk-adjustment process should be subject to clinical review. For example, the clinical review ensures that the risk-adjusted clinical factors have logical and causal relationships with episode spend, and that undue complications of a patient's care are not considered for risk adjustment.

## 4. Clinical-exclusion and risk-adjustment process

Reflecting the principles above, we propose the following six steps to identify clinical factors that lead to exclusion of an episode or to risk adjustment of episode spend (see exhibit).

Exhibit

### Clinical-exclusion and risk-adjustment process



### 4.1 Assign episodes to an archetype

The first step of the clinical-exclusion and risk-adjustment process assigns each episode type (e.g., asthma acute exacerbation, perinatal) to an archetype. The assignment is based on the body system and type of medical condition. Developed from the highest-level categories of the Clinical Classification Software (CCS) approach to grouping diagnoses into clinically meaningful categories, the proposed archetypes are:

- Infectious and parasitic diseases
- Neoplasms
- Endocrine, nutritional, and metabolic diseases and immunity disorders
- Diseases of the blood and blood-forming organs
- Mental illness
- Diseases of the nervous system and sense organs
- Diseases of the circulatory system
- Diseases of the respiratory system
- Diseases of the digestive system
- Diseases of the genitourinary system
- Certain conditions originating in the perinatal period and complications of pregnancy, childbirth, and the puerperium
- Diseases of the skin and subcutaneous tissue
- Diseases of the musculoskeletal system and connective tissue
- Injury and poisoning

Each archetype is associated with five groups of clinical factors. The groups offer a structured approach to identify clinical factors that lead to episode exclusions (clinical exclusions) or to adjustments of episode spend (risk adjustment). Clinical factors that result in neither exclusion nor risk adjustment do not impact the measurement of PAP performance.

The five groups of clinical factors for each archetype are:

- **Group 1 – Clinical factors that always result in exclusion:** Clinical factors (e.g., HIV, ESRD) that have unique and/or severe impacts on patients within a given archetype. Episodes in which the patient has such a clinical factor are not comparable to other episodes of the same episode type, even after risk adjustment. Therefore, the clinical factors lead to exclusion of the affected episodes from the episode-based payment model (see step 2).
- **Group 2 – Clinical factors that are always considered for risk adjustment:** Clinical factors with a strong clinical rationale for a measurable impact on episode spend for episode types within an archetype. Clinical factors in this group always undergo further statistical testing (see step 3).
- **Group 3 – Clinical factors that are plausible for risk adjustment:** Clinical factors that may plausibly impact episode spend for episode types within an archetype. Clinical factors in this group are considered for risk adjustment if the presence of a given clinical factor correlates with higher spend for the episode type under consideration, but a strong clinical rationale for that correlation has not (yet) been established (see step 3).
- **Group 4 – Clinical factors that should not affect the calculation of episode spend:** Clinical factors lacking a clinically justified impact on episode spend or that are potential complications of care for episode types within an archetype. Clinical factors in this group are not considered for exclusion or risk adjustment.
- **Group 5 – Clinical factors that may be considered for risk adjustment:** Clinical factors that are not part of any of the other groups. Clinical factors in this group do not have a strong clinical relationship to the episode types within an archetype, but may affect certain episode types for specific, clinically justifiable reasons unique to the episode type. As with the clinical factors that are plausible for risk adjustment, clinical factors in this group are considered for risk adjustment if their presence correlates with higher episode spend for the given episode type, though the bar for consideration is higher (see step 3).

The assignment of clinical factors to the five groups for each archetype is based on input from a panel of clinical experts and undergoes periodic revisions to reflect clinical, academic, and empirical evidence.

The single-level Clinical Classification Software (CCS) categories for International Classification of Diseases diagnosis codes can be used as a base universe of clinical factors. The US Agency for Healthcare Research and Quality (AHRQ) developed CCS categories to group diagnoses and

procedures into clinically meaningful categories. CCS is a frequently used, validated, and transparent method of identifying clinical factors using claims data.<sup>4</sup> The clinical factors are identified based on diagnosis codes in a patient's claims data, typically over the year prior to and/or during an episode (depending on the type of clinical factor). As clinically appropriate, clinical factors may also be defined based on more granular diagnosis groups than single-level CCS categories (e.g., Body Mass Index of 30 or higher), patient age, or procedure codes in the patient's claims data (e.g., procedures for active cancer treatment). Each clinical factor is a binary flag, e.g., diabetes documented during the year before the episode yes/no, coronary artery disease documented during the year before the episode yes/no, age 25–34 at the start of the episode yes/no.

## 4.2 Apply archetype exclusions

The second step of the clinical-exclusion and risk-adjustment process identifies and excludes episodes where the patient has a clinical factor in Group 1, i.e., it always leads to exclusion for the episode type under consideration. This step removes episodes of patients with a clinical factor that does not allow fair comparisons with episodes of patients without the clinical factor.

## 4.3 Create shortlist of clinical factors for risk adjustment

The third step of the clinical-exclusion and risk-adjustment process creates a shortlist of clinical factors that are outside the control of the PAP but have a clinical and/or statistical rationale for their impact on episode spend and which are, therefore, considered for risk adjustment. The clinical factors on the shortlist are evaluated further in steps 4 and 5, below. Clinical factors are added to the shortlist in four ways:

- All clinical factors in Group 2 (Clinical factors that are always considered for risk adjustment) are added to the shortlist.
- A clinical factor in Group 3 (Clinical factors that are plausible for risk adjustment) is added to the shortlist if it correlates with episode spend for the episode type under consideration. A basic statistical evaluation with a liberal p-value threshold of, for example, 0.05 is applied to determine which plausible clinical factors correlate with episode spend.

The method uses the plausible clinical factors as independent variables and episode spend as the dependent variable. Throughout the risk adjustment process, regressions employ episode spend that is normalized for unit price variation (see section 5). A variable selection algorithm (backward step-down variable deletion using ordinary least squares [OLS] regression) is run using ten-fold cross-validation for valid episodes in the input data.<sup>5</sup> Input data typically consists of the most recent two or three full years of claims data, depending on availability and relevance.

- A clinical factor in Group 5 (Clinical factors that may be considered for risk adjustment) is added to the shortlist if it strongly correlates with episode spend for the episode type under consideration. The same basic statistical evaluation technique described above for plausible clinical factors applies; however, a more stringent p-value threshold of, for example, 0.005 or 0.01 is used as the parameter for clinical factor selection.
- Finally, clinical experts review the shortlist and may remove or include clinical factors based on clinical rationale specific to the episode type in question, e.g., factors are removed because the associated spend variation is clinically unwarranted, or factors are added because the associated spend variation is beyond the control of providers. For example, use of portable oxygen in the year before an episode may be considered as a risk factor for COPD episodes but not other episode types.

#### 4.4 Apply low-volume and high-variation clinical exclusions

The fourth step of the clinical-exclusion and risk-adjustment process identifies clinical factors on the shortlist that could not be reliably risk adjusted using statistical methods and therefore become clinical exclusions. A clinical factor on the shortlist becomes a clinical exclusion if one of the following applies:

- It is present in too few valid episodes for its impact on episode spend to be measured accurately. While there is reason to believe that all clinical factors on the shortlist may impact episode spend, if that impact cannot be accurately measured because of limited data the episode cannot be equitably included when measuring a PAP's performance. The minimum number of episodes with a given clinical factor required for accurate measurement is typically set at 10. The use of 10 as a cutoff is consistent with CMS methodology for calibrating DRGs<sup>6</sup> and should, in conjunction with other safeguards in the process (e.g., a statistically significant correlation with episode spend, acceptable co-linearity as measured by a variance inflation factor [VIF] less than 3), lead to an accurate measurement of the clinical factor's impact on episode spend.
- It is associated with high variation in episode spend and may present an unacceptable level of risk to PAPs. High variation is measured using the coefficient of variation of the re-centered residuals, which represents variation beyond what can be controlled for by a risk-adjustment model. The coefficient of variation of the re-centered residuals captures the unexplained variation in episode spend relative to the average spend of episodes with a given clinical factor. A clinical factor on the shortlist is typically considered to be associated with high variation if either of the following applies:
  - The coefficient of variation of the re-centered residuals is greater than 1 and also greater than the overall coefficient of variation of episode spend for all valid episodes.
  - The coefficient of variation of the re-centered residuals is greater than 3.

To calculate the coefficients of variation, an OLS regression is estimated on all valid episodes in the dataset used for risk adjustment, using the clinical factors on the shortlist as independent variables and episode spend as the dependent variable. Only clinical factors on the shortlist that have sufficient volume as defined above are used. The overall coefficient of variation is calculated as:

$$\text{Overall coefficient of variation} = \frac{\text{Standard deviation of episode spend for all valid episodes}}{\text{Observed mean of episode spend for all valid episodes}}$$

For each valid episode with a given clinical factor ( $CF_i$ ) the re-centered residual is calculated as:

$$\begin{aligned} \text{Re-centered residual for an episode with } CF_i = & \\ \text{Residual of episode spend for an episode with } CF_i & \\ + & \\ \text{Observed mean of episode spend for valid episodes with } CF_i & \end{aligned}$$

The mean residual for each clinical factor is exactly zero by design. To put the residual variability back in context, the residuals are “re-centered” by adding the mean spend of the clinical factor to the residuals. The coefficient of variation of the re-centered residuals is calculated as:

$$\text{Coefficient of variation of the re-centered residuals for } CF_i = \frac{\text{Standard deviation of the re-centered residuals for valid episodes with } CF_i}{\text{Observed mean of episode spend for valid episodes with } CF_i}$$

The definition of “high variation” is based on the following rationale: Risk adjustment aims to decrease the PAPs’ exposure to uncertainty that arises from clinical factors outside their control. If risk adjustment cannot reduce the uncertainty for a particular clinical factor compared with the amount of uncertainty before risk adjustment, then the clinical factor may represent an unacceptable level of relative uncertainty. We arrive at a lack of reduced uncertainty when a coefficient of variation of the re-centered residuals is higher than the overall coefficient of variation.

The definition of “high variation” is a conservative exclusion criterion, because segmenting episodes by clinical factors and controlling for other clinical factors using OLS regression should decrease the coefficient of variation of the re-centered residuals for most clinical factors relative to the overall coefficient of variation. Additionally, the “high variation” criterion is bounded by absolute cutoffs, so that clinical factors with a coefficient of variation of the

re-centered residuals that is typically considered acceptable ( $<1$ ) are not excluded and clinical factors with a coefficient of variation of the re-centered residuals that is typically considered very high ( $>3$ ) are always excluded. If the above criteria lead to a large number of exclusions for a particular episode type, the criterion for high variation may need to be adapted.

After applying the low-volume and high-variation clinical exclusions, there remains a shortlist of clinically and statistically validated clinical factors that will undergo further testing.

## 4.5 Calculate risk coefficients

The fifth step of the clinical-exclusion and risk-adjustment process estimates a risk-adjustment model that optimally predicts episode spend. The coefficients of the risk-adjustment model measure the impact of individual clinical factors on episode spend and therefore can isolate and adjust for the clinical factors outside the PAP's control.

Generating the risk-adjustment model is accomplished in two sub-steps: First, an optimally predictive subset of clinical factors is selected from the shortlist of clinical factors that resulted from steps 1 through 4. The selection of the clinical factors is based on a balance of improving in-sample prediction and reducing the potential for over-fitting random noise in the estimation sample that could lower out-of-sample prediction. Second, a risk coefficient is estimated for each clinical factor in the optimally predictive subset.

### 4.5.1 Model selection

To select the optimally predictive clinical factors, an exhaustive search of all possible risk-adjustment models using the shortlist of clinically and statistically validated clinical factors is conducted. The model selection process involves the following steps:

- For each possible combination of clinical factors on the shortlist, an OLS regression model is estimated using the clinical factors as independent variables and episode spend as the dependent variable for valid episodes in the input data.<sup>7</sup> Valid episodes do not include those excluded because of archetype, low-volume, or high-variation clinical exclusions or other exclusions as applicable. Every possible model is generated, ranging from models with a single clinically and statistically validated clinical factor on one extreme to a model including all the clinically and statistically validated clinical factors together on the other.
- From all models with a given number of clinical factors, the model with the highest adjusted R-squared (i.e., with the highest percent of variance in episode spend explained by the clinical factors) is selected. The resulting set of models is in an ordered list by the number of included clinical factors (i.e., best model with one clinical factor, best model with two clinical factors, etc.). For each of the most predictive models, the Bayesian information criterion (BIC),

Mallows's  $C_p$ , and adjusted R-squared—three model-selection parameters commonly used in the statistical literature—are calculated and the best model based on the selection criterion for each parameter is identified. The selection criteria are: the model with the lowest BIC, the model with the highest number of clinical factors where the number of clinical factors is less than one-half Mallows's  $C_p$ , and the model with the highest-adjusted R-squared. Of the three candidate models selected based on each selection criterion, the one with the median number of clinical factors is selected as the optimal risk-adjustment model.

#### 4.5.2 Estimate risk coefficients

To estimate the risk coefficients, a linear OLS and a generalized linear model (GLM) with a log link function (Poisson regression) are estimated using the clinical factors that were selected in the model selection step as the independent variables and episode spend as the dependent variable for valid episodes in the input data.<sup>9</sup>

Before the risk coefficients are finalized, the following checks are performed:

- The VIF for the clinical factors is calculated to ensure that there is no severe multi-collinearity in the data.<sup>9</sup> Any VIF higher than 3 is examined and clinical factors may be removed or transformed (e.g., clinical factors may be rolled up to a higher level of CCS categories, broken out to a lower level of CCS categories, or combined into one clinical factor) based on clinical input to reduce multi-collinearity.
- The average *predicted* episode spend for each decile of predicted spend, from highest to lowest, is compared with the average *observed* episode spend to test for systematic over- or underprediction of high- or low-spend episodes. Additionally, the average predicted episode spend for episodes with a given number of clinical factors is compared with the average observed episode spend to test for systematic over- or underprediction of high- or low-risk episodes. The regression model (OLS or GLM) that best fits the data without systematic bias is used to generate risk coefficients and the other model is discarded. If such over- or underprediction exists in both models, clinical factors may be removed or transformed based on clinical input. If over- or underprediction persists, a different regression model may need to be chosen that better conforms to the observed distribution of episode spend.
- The frequency and goodness of the prediction of the “base case” with no risk-adjusted clinical factors is tested. If too few episodes (e.g., <10%) have no risk-adjusted clinical factors or if episodes with no risk-adjusted clinical factors are systematically over- or underpredicted, clinical factors may be removed or transformed, based on clinical input, to ensure that the “base case” is well represented and well predicted.
- Highly co-morbid patients present an exception to removing or transforming risk-adjusted clinical factors. Because this group has a large number of clinical factors, the episode spend

is difficult to predict accurately using a regression model. Typically the presence of many clinical factors generates episode-spend predictions that are either too high (most likely when using a GLM) or too low (most likely when using OLS) because of the compounding effect of each clinical factor. If only highly co-morbid episodes show systematic over- or underprediction, then the clinical factors are not removed or transformed. Instead, episodes with more than a specific number of risk-adjusted clinical factors are excluded. A cutoff for the number of risk-adjusted clinical factors is typically set if there is a systematic bias in the percent difference between the predicted and the observed episode spend of more than plus or minus 7.5% and affected episodes make up less than 2% of the total valid episodes. If the exclusion of highly co-morbid episodes leads to another risk-adjusted clinical factor no longer being associated with a sufficient number of valid episodes to generate a credible estimate of its impact, then that clinical factor becomes a clinical episode exclusion.

If any clinical factors are removed or transformed at any point in the above process then the clinical-exclusion and risk-adjustment process is repeated, starting on step 4, accounting for any removed clinical factors and updating any transformed clinical factors. When repeating the steps, highly co-morbid episodes are included, because the clinical factors that make the episodes highly co-morbid may not be present in the final model.

When the model-selection step generates a set of clinical factors that passes the checks described above—without removing or transforming any clinical factors—then those clinical factors and the risk coefficients generated by the selected regression model undergo a final clinical review in step 6.

## 4.6 Perform clinical review

In the sixth step of the clinical-exclusion and risk-adjustment process, an advisory group of local clinicians and stakeholders in the episode-based payment model is asked to review the results. The group's feedback is solicited using the following questions:

- Do any of the clinical factors that result in exclusion or risk adjustment surprise you? If so, why?
- In particular, are there any clinical factors that result in exclusions or risk adjustment that you believe should *not* influence treatment pathway, cost, or outcomes?
- Are the top three clinical factors most likely to impact spend for this particular episode type but outside the control of the PAP on the lists? If not, which clinical factors should be added? In what way do they impact episode spend?

If the feedback from the advisory group raises concerns about any of the clinical factors that result in exclusion or risk adjustment for the episode type, updates may be made as appropriate:

- Clinical factors that appear to be causally unrelated to episode spend may be removed as clinical exclusions or from risk adjustment.
- Clinical factors that appear likely to be causally related to clinically justified episode spend may be added to the shortlist for risk adjustment.
- Clinical factors that appear likely to have unique and/or severe impacts may become clinical exclusions.

Based on the clinical review, if any updates are made, the clinical-exclusion and risk-adjustment process is re-run from step 4 using the additional episode exclusion(s) and the updated shortlist of clinical factors. The clinical factors for risk adjustment and the risk coefficients that result from the re-run are the final risk-adjustment output.

## 5. Statistical and technical Q&A

Below are questions and answers that clarify statistical and technical considerations for the clinical-exclusion and risk-adjustment process.

■ **How does the clinical-exclusion and risk-adjustment process address episode types that may have only a limited number of occurrences in the data?**

Given a limited number of episodes, the main concern is to balance the desire to include all of a potentially large set of clinical factors that may affect episode spend with the need for each clinical factor to be sufficiently independent and have sufficient volume to reliably estimate its impact on episode spend. While episodes may be affected by many clinical factors, only a subset of the clinical factors occur frequently enough to reliably measure their effect on episode spend. Therefore, a process was chosen that first defines a large set of clinical factors that may affect an episode of a given type and then narrows down the large set of clinical factors to those that are clinically validated, statistically reliable, and practically measurable.

■ **How does the clinical-exclusion and risk-adjustment process balance the risk of including unimportant clinical factors with the risk of not including important ones?**

The process uses pre-selected clinical factors from archetypal lists and repeated clinical input, to ensure that every clinical factor has a strong clinical justification for its use as a clinical exclusion or in risk adjustment. Requiring any clinical factor to have a strong prior probability of meaningful impact on episode spend reduces the likelihood that it is included in risk adjustment by chance. To minimize the risk of not including important clinical factors, the group of clinical factors that may be considered for risk adjustment (Group 5) is examined using statistical techniques that reveal any strong relationships in the data. Clinical evaluation and review are again used to reduce the risk of including clinical factors that occur only by chance.

■ **How does the clinical-exclusion and risk-adjustment process address unit price differences?**

The process should capture the impact of clinical factors apart from any impact of unit price differences and avoid capturing clinical factors that correlate with the use of high- or low-cost facilities but do not impact the actual care pathway for the patient. Since inpatient spend typically constitutes a large share of episode spend and may show considerable unit price variation, the episode spend used for risk adjustment is calculated using normalized inpatient spend. Depending on the payment method for inpatient hospitals, normalization may be accomplished using normalized DRG base rates or normalized per diems.

■ **How does the clinical-exclusion and risk-adjustment process address episodes with very high episode spend?**

The effect of outlier episodes with very high episode spend should be limited to ensure that the estimates are not unduly biased by a small number of anomalous episodes. Therefore, for the purpose of risk adjustment, spend of outlier episodes is capped (winsorized) at 3 standard deviations above the mean episode spend of valid episodes. Episodes with very high episode spend are capped instead of excluded in order to generate more accurate and equitable predictions.

**Bowen Garrett** (bgarrett@urban.org) is an economist and senior fellow in the Health Policy Center at the Urban Institute. **Adi Kumar** (adi\_kumar@mckinsey.com) is an associate principal in McKinsey's Washington, DC, office, where **Friederike Haass** (friederike\_haass@mckinsey.com) and **Josh Haselkorn** (josh\_haselkorn@mckinsey.com) are consultants with McKinsey's Healthcare Value Analytics Solution. **Tom Latkovic** (tom\_latkovic@mckinsey.com) is a director in McKinsey's Cleveland office. **Daniel Tsai** (daniel\_tsai@mckinsey.com) is a partner in McKinsey's Boston office. **Tim Ward** (tim\_ward@mckinsey.com) is a partner in McKinsey's Southern California office.

McKinsey's Healthcare Value Analytics Solution helps clients measure, encourage, and deliver high-value healthcare. HCVA works in lockstep with management consulting colleagues to design, implement, and evaluate healthcare value strategies. HCVA's dedicated team of healthcare, analytics, big data, and software experts collaborates to craft end-to-end answers—from diagnostics to design to implementation and performance tracking—to serve clients' needs.

## 6. References

1. Michael Bailit and Margaret Houy, “Key Payer and Provider Operational Steps to Successfully Implement Bundled Payments.” HCl3 Issue Brief, May 2014. See also: “Bundled payments: Moving from pilots to programs.” *Health Affairs* (blog), June 2, 2014.
2. Tom Latkovic, “The trillion dollar prize: Using outcomes-based payment to address the US healthcare financing crisis.” *McKinsey Quarterly*, March 2013, mckinsey.com.
3. Adi Kumar, Tom Latkovic, and Daniel Tsai, “It’s time for episode-based health care spending,” *Harvard Business Review* (blog), October 2013.
4. Healthcare Cost and Utilization Project (HCUP) Clinical Classifications Software (CCS) for ICD-9-CM: <http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>
5. Implemented using the statistical package “RMS” – <http://cran.r-project.org/web/packages/rms/rms.pdf>
6. For example, Federal Register Vol. 75, No. 157, Page 50136
7. Implemented using the statistical package “LEAPS” – <http://cran.r-project.org/web/packages/leaps/leaps.pdf>
8. Implemented using the “lm” and “glm” functions in R – <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/lm.html> [lm] and <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html> [glm]
9. Implemented using the statistical package “HH” – <http://cran.r-project.org/web/packages/HH/HH.pdf>

Contact for distribution: Melanie Bierweiler

Phone: +1 (727) 561-5650

Email: [Melanie\\_Bierweiler@mckinsey.com](mailto:Melanie_Bierweiler@mckinsey.com)

Copyright © 2014 McKinsey & Company

CONFIDENTIAL AND PROPRIETARY

Any use of this material without specific permission of McKinsey & Company is strictly prohibited.

[www.mckinsey.com/client\\_service/healthcare\\_systems\\_and\\_services](http://www.mckinsey.com/client_service/healthcare_systems_and_services)