HHS Public Access
Author manuscript
*Popul Health Manag*. Author manuscript; available in PMC 2022 April 11.

Published in final edited form as:
*Popul Health Manag*. 2020 April ; 23(2): 165–173. doi:10.1089/pop.2019.0076.

# How to Classify Super-Utilizers: A Methodological Review of Super-Utilizer Criteria Applied to the Utah Medicaid Population, 2016–2017

**Carl J. Grafe, PhD, MS**[1,2], **Roberta Z. Horth, PhD, MPH**[1,3,4], **Nelson Clayton, MS, MBA**[5], **Angela Dunn, MD, MPH**[4], **Navina Forsythe, PhD, MPA**[2]

[1]Division of Scientific Education and Professional Development, CDC, Atlanta, Georgia.

[2]Center for Health Data and Informatics, Utah Department of Health, Salt Lake City, Utah.

[3]United States Public Health Service, Commissioned Corps, Rockville, Maryland.

[4]Division of Disease Control and Prevention, Utah Department of Health, Salt Lake City, Utah.

[5]Division of Medicaid and Health Financing, Utah Department of Health, Salt Lake City, Utah.

## Abstract

A limited number of patients, commonly termed super-utilizers, account for the bulk of health care expenditures. Multiple criteria for identifying super-utilizers exist, but no standard methodology is available for determining which criteria should be used for a specific population. Application is often arbitrary, and poorly aligned super-utilizer criteria might result in misallocation of resources and diminished effects of interventions. This study sought to apply an innovative, data-driven approach to classify super-utilizers among Utah Medicaid beneficiaries. The authors conducted a literature review of research methods to catalogue applied super-utilizer criteria. The most commonly used criteria were applied to Utah Medicaid beneficiaries enrolled during July 1, 2016–June 30, 2017, using their previous 12 months of claims data (N = 309,921). The *k*-medoids algorithm cluster analysis was used to find groups of beneficiaries with similar characteristic based on criteria from the literature. In all, 180 super-utilizer criteria were identified in the literature, 21 of which met the inclusion criteria. When these criteria were applied to Utah Medicaid data, 5 distinct subpopulation clusters were found: non–super-utilizers (n = 163,118), beneficiaries with multiple chronic or mental health conditions (n = 68,054), beneficiaries with a single chronic health condition (n = 43,939), emergency department super-utilizers with chronic or mental health conditions (n = 7809), and beneficiaries with uncomplicated hospitalizations (n = 27,001). This

---

Address correspondence to: *Roberta Z. Horth, PhD, MPH, Division of Disease Control and Prevention, Utah Department of Health, 288 North 1460 W, Salt Lake City, UT 84116*, hxw5@cdc.gov.

Author Disclosure Statement
The authors declare that there are no conflicts of interest.
The authors are employees of the Centers for Disease Control and Prevention or the Utah Department of Health and completed this work as part of their regular job duties. No other funding was obtained for this study. The conclusions, findings, and opinions expressed by the authors do not necessarily reflect the official position of the US Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention, the Utah Department of Health, the Division of Medicaid and Health Financing, or other affiliated institutions. The Division of Medicaid and Health Financing makes no representations and extends no warranties to the accuracy of the data it provided.

Supplementary Material
Supplementary Data

study demonstrates how cluster analysis can aid in selecting characteristics from the literature that systematically differentiate super-utilizer groups from other beneficiaries. This methodology might be useful to health care systems for identifying super-utilizers within their patient populations.

### Keywords

cluster analysis; Medicaid; medical overuse; systematic review

## Introduction

THE BULK OF A HEALTH CARE system's expenditures often are directed toward a limited number of patients, commonly known as super-utilizers.[1] Estimates are that as much as 54% of US health care costs can be attributed to only 5% of patients.[1] Interventions directed toward these patients have potential to control costs,[2] increase appropriateness of care,[3] and improve health outcomes.[4] Some risk stratification tools include common groupings that address differences in utilization,[5] but these are often proprietary and may not be scalable across payers/providers using different risk stratification tools. There currently are no standard protocols for systematically determining which criteria should be used to identify super-utilization within a specific population.

In 2013, the Centers for Medicare & Medicaid Services (CMS) issued an informational bulletin identifying resources to help promote programs for super-utilizers, which it defined as "beneficiaries with complex, unaddressed health issues and a history of frequent encounters with health care providers."[1] The bulletin included the following list of super-utilizer approaches: targeting based on high observed-to-expected costs; targeting specific patterns of care; targeting very high levels of utilization; targeting based on referrals and follow-up investigation; excluding candidate clients with medical conditions associated with high but nonpreventable costs; targeting by presence of risk factors associated with high, preventable costs; and targeting by community.

Many health care systems and agencies have since used ≥1 of these approaches to implement super-utilizer programs, but no standard process has emerged to determine which criteria might best match a given population. Institutions may choose from a variety of tool-based standards or use their own definitions. However, using super-utilizer criteria that are poorly aligned with the population being served can result in misallocation of resources and diminished effects of interventions.[6,7] To be most effective, health care institutions must understand which criteria are available and most relevant to their specific patient populations. In particular, an emphasis on Medicaid agencies and how they are applying super-utilizer criteria is of interest because of CMS's role in defining super-utilizers.

This study sought to catalogue commonly used super-utilizer criteria from a literature review and to use cluster analysis to identify which combinations of super-utilizer criteria were algorithmically apparent among Medicaid beneficiaries in Utah. A broadly applicable 5-step approach was used, generalized as follows:

1. Agree on a concept of interest (eg, classification of super-utilizers) and search the literature to find studies that help define or characterize that concept;
2. Review the studies found to describe the desired criteria about that concept in more detail (eg, criteria related to super-utilizer status);
3. Find a data set of interest (eg, Utah Medicaid enrollment and claims data) and create a person-level view of that data set, then add binary (yes or no) variables for each criterion to each person's data;
4. Use an empirical clustering algorithm (eg, *k*-medoids) to find a small number of groups in the data set where the groups differ in terms of the distribution of the criteria; and
5. Interpret the cluster results, and comment on use cases, advantages, and limitations, among other factors.

## Methods

### Methodological review

On October 19, 2017, the research team queried PubMed and Google Search (Google LLC, Mountain View, CA) to identify literature containing studies or programs that had applied super-utilizer criteria to health care populations. In order to capture all available definitions that could be applied to Medicaid, the team included criteria applied to non-Medicaid populations. PubMed searches included "super-utilizers," "super utilizers," "'high need' AND 'high cost,"' "frequent fliers," and "hot spotters." These terms were taken from the team's experience and a preliminary review of the literature. Google searches included "Medicaid super utilizers"; the team only retained papers and articles in the first 40 results as this was a supplementary search meant to capture top government reports and professional papers from the gray literature. The team found additional documents from published systematic reviews identified in the PubMed search and from links, footnotes, or citations in the Google search results.

The research team included only official reports (defined as formal documents from academia, government, non-profits, or the private sector) that applied ≥1 super-utilizer criteria (as defined by CMS[1]) to original individual-level population data. Two reviewers independently reviewed document abstracts from PubMed results and title of citations from the Google search to exclude work that did not meet this inclusion criterion. Disagreements between reviewers were resolved by consensus. All remaining documents were reviewed in detail (Figure 1).

Criteria included were categorized by the CMS super-utilizer targeting approach[1] and logical subcategories (determined by the reviewers) within each approach. Specific criteria within these subcategories were tabulated for evaluation of whether they were commonly used. Combinations of criteria in the same document were separated into component parts for categorization (eg, 2 emergency department [ED] visits or 2 inpatient admissions would be tabulated under both categories). Criteria from documents containing multiple studies or multiple sets of criteria were likewise tabulated separately. Because this analysis was intended for retrospective data, the team excluded referral-based criteria, which would not be available in the data set and would be difficult to simulate, because of inconsistent application in practice.[8] Criteria that included rates over time were prorated to a 12-month period. For documents with multiple gradations of utilization, the more inclusive criterion was used (eg, some documents distinguished between high utilizers and super-utilizers). Criteria that used the term behavioral health without further explanation were categorized under both mental health and substance use disorders. Individual criteria that were applied in at least 5 documents and accounted for at least 10% of all criteria that fell in a category subset were considered commonly used. Additionally, documents were reviewed for any specific methods used to determine which criteria would be applied, and criteria developed using clearly defined data-driven methods were noted.

## Cluster analysis

The research team obtained Utah Medicaid claims data from the Division of Medicaid and Health Financing for all beneficiaries with claims in Utah during State Fiscal Year (SFY) 2017 (July 1, 2016–June 30, 2017). Cluster analysis was used to group these beneficiaries by super-utilizer criteria identified from the literature review. Two of 21 criteria that were commonly identified in the literature were excluded because Utah Medicaid lacked data on homelessness and social needs. Additionally, 2 criteria identified from the literature review were excluded because of overlap with other criteria. Jenks Natural Breaks analysis was used to identify top cost percentage break points based on methodology described in one of the literature review documents,6 which brought the total number of applied criteria for this analysis to 20. For Jenks Natural Breaks, the team used the classInt package[9] in R 3.4.2[10] (R Foundation for Statistical Computing, Vienna, Austria) with N = 3 clusters and considered the top cluster (3.1%) to be the break point of interest.

Dummy variables for each criterion were assigned based on claims data during the 12-month period before each beneficiary's last claim of SFY 2017. This included the amounts paid by Medicaid, *International Classification of Diseases, Tenth Revision* (ICD-10) codes,[11] and revenue codes. Out-of-pocket patient costs were not included in the analysis because of the study's focus on institutional costs. Chronic health conditions were assigned using Elixhauser Comorbidity Index[12] categories based on the beneficiary's ICD-10 codes. Top cost percentage variables were based on the population of beneficiaries with claims. Excluded from the study were enrollees in the Children's Health Insurance Program, refugees, juvenile justice cases, and custody medical cases because of shared stewardship with other agencies. No other subpopulations or groupings were excluded in order to achieve a largely agnostic top-level view of potentially impactable institutional costs.

The research team used the CLARA (clustering large applications) implementation of *k*-medoids from the cluster package[13] in R 3.4.2 to conduct the clustering analysis. The "Manhattan" distance[14] was used to calculate dissimilarities between super-utilizer criteria dummy variables. Eligible Utah Medicaid beneficiaries who filed no claims during SFY 2017 were excluded and treated as a separate cluster with no associated costs. The number of clusters was selected using the average silhouette method[15] for up to 30 clusters. The minimum number of clusters was confirmed visually by applying the elbow method, an approach commonly used in cluster analysis that selects the optimum number of clusters such that adding an additional cluster does not significantly reduce the objective function (ie, where the curve bends).[16] The presence or absence of each criterion among Medicaid beneficiaries (ie, the percentage of beneficiaries in a cluster who met the specified criteria) was used to characterize each cluster.

This study consisted of a methodological review of research methods used in the literature and application of review findings to claims data from the Utah Medicaid population, and entailed secondary analysis of data collected routinely by Utah Medicaid. The Institutional Review Boards at the Utah Department of Health (Approval #533) and the Centers for Disease Control and Prevention (HSR #2018–00189) determined that this study did not constitute human subjects research.

## Results

### Methodological review

Searches of PubMed and Google Search resulted in 845 documents for review (Figure 1). Of these, 136 were from PubMed searches (N = 96) and Google searches (N = 40). Of the PubMed search results, 32 were from the "frequent fliers" search, 31 from "'high need' AND 'high cost,"' 28 from "super-utilizers," and 5 from "hot spotters." Of 709 documents identified from other sources, 641 records were identified from links, footnotes, or citations from the Google search results and 68 were from systematic reviews found among the PubMed results. Of these documents, 613 were removed after deduplication, 443 were excluded, and fulltext review was conducted for the 170 remaining documents. After full review, a further 59 documents were excluded, resulting in 111 documents included in the synthesis (Supplementary Data).

The studies included produced 180 super-utilizer criteria: 112 from the academic literature, 51 (39 state and 12 federal) from government reports, and 17 from nonprofit organizations. Fifty-three criteria (29%) were specifically applied to Medicaid populations. Of the 180 super-utilizer criteria, 89% identified super-utilizers based on very high levels of utilization, most commonly ED visits or number of inpatient admission (Table 1). Half of criteria identified super-utilizers based on risk factors associated with high preventable costs, mostly chronic health conditions or behavioral health conditions (Table 1).

Altogether, 21 definitions met the minimum criteria to be considered commonly used (Table 2). Only the following categories met the inclusion criteria: targeting very high levels of utilization; targeting based on risk factors associated with high, preventable costs; and targeting based on referrals and follow-up investigation. Five subcategories were apparent within the utilization-based targeting category (number of ED visits, number of inpatient admissions, top predicted risk score, number of prescribed medications, and top cost percentage). Top predicted risk score was excluded because of substantial variation and lacking detail in the approaches used to determine the risk score among sites. Number of prescribed medications also was excluded because a time frame for determining the number of medications was available in only 3 documents. Four subcategories were apparent within the risk factor-based targeting category (chronic health conditions, number of chronic health conditions, behavioral health conditions, and social determinants of health).

Seventy-seven documents (69%) stated their super-utilizer criteria, allowing them to be categorized, but did not describe how their criteria were selected. Twenty-three documents (21%) based their criteria on previous criteria found in the literature. Four documents (4%) referred only generally to deriving their criteria from previous work the authors had completed. Only 7 documents (6%) included detailed descriptions of data-driven approaches for developing their criteria. Of those documents with detailed descriptions, 3 used the mean number of ED visits plus 2 standard deviations as a criterion,[7,17–18] 1 used the mean number of inpatient admissions plus 2 standard deviations,[19] 2 used other combinations of the mean and standard deviation of the number of visits,[20–21] and 1 used natural break points to identify the top cost percentages.[6] Certain documents acknowledged that the criteria used to identify super-utilizers were arbitrary.[22–24]

## Cluster analysis

Cluster analysis identified an optimum number of 5 distinct clusters of super-utilizer criteria among the 309,921 Utah Medicaid beneficiaries with claims during SFY 2017 (Figure 2). The average silhouette width at 5 clusters was exceeded slightly at 29 clusters, but the elbow method confirmed 5 clusters as the optimum number to use (Figure 3). These clusters included the following: Cluster 1: Beneficiaries who were not super-utilizers (<10% met any super-utilizer criteria; n = 163,118); Cluster 2: Beneficiaries with multiple chronic or mental health conditions (breakdown of chronic health conditions and mental health disorders shown in Table 3; n = 68,054); Cluster 3: Beneficiaries with a single chronic health condition (breakdown of chronic health conditions and mental health disorders shown in Table 3; n = 43,939); Cluster 4: Beneficiaries with chronic or mental health conditions who are ED super-utilizers (breakdown of chronic health conditions, mental health disorders, ED use, and costs shown in Table 3; n = 7809); and Cluster 5: Beneficiaries with ≥1 hospitalizations and no chronic or mental health conditions (breakdown of admission, chronic health conditions, and mental health disorders shown in Table 3; n = 27,001).

Only Clusters 2 and 4 had ≥4 criteria that applied to at least 10% of beneficiaries in the cluster. Cluster 1 did not exceed 10% for any criterion. Clusters 3 and 5 only exceeded 25% on 1 criterion each. Although 100% of beneficiaries in Cluster 3 had ≥1 chronic conditions, 0% had ≥2 chronic conditions, indicating that this cluster was completely composed of beneficiaries with a single chronic condition. Similarly, 100% of beneficiaries in Cluster 5 had ≥1 inpatient admissions, but only 8% had ≥2 inpatient admissions. Closer inspection (data not shown) revealed that many inpatient admissions in this cluster were for infant deliveries and other routine hospitalizations. Meanwhile, Clusters 2 and 4 had more than 10% of the cluster populations meet a majority of the super-utilizer criteria (11 and 18 criteria, respectively). These clusters included beneficiaries with complex needs and ED super-utilizers, who are more traditional super-utilizer populations.[1]

## Discussion

This study was conducted to identify super-utilizer criteria that have been used in the literature and how these might be applied to the Utah Medicaid population. Results revealed substantial variation in criteria that have been used, but with limited evidence of objective, data-driven methods being used for criteria selection. The use of cluster analysis to apply these different criteria to the Utah Medicaid population identified discrete super-utilizer criteria that were algorithmically apparent within the population of interest.

These results are important for institutional decision making. The criterion of 4 ED visits within a year has been cited frequently in the literature,[22,25–28] but this criterion often is applied arbitrarily and might not maximize in-group homogeneity and out-group heterogeneity that results in misclassification of super-utilization. For example, a health care cost-sharing program for construction workers might expect to have higher ED use on average, compared with other health plans.[29] Using a threshold of 4 ED visits for such populations might include people who are not using more care than is expected given their occupation, resulting in inefficient targeting and potentially increased costs. Patterns of health care use in the United States vary by geography,[30] sex and age,[31] and race.[32] Accordingly, super-utilizer criteria from the literature for specific populations might not remain valid across other populations or as populations change over time. Rigorous, evidence-based standards for identifying super-utilizers are essential to ensure that interventions are effective and reach intended populations.

The use of cluster analysis to identify patterns of use or risk factor prevalence in this study had multiple advantages. Most importantly, it allowed the research team to avoid applying subjective cut points to determine the appropriate population subsets to be targeted for improved care coordination and other interventions. Defining a population too narrowly might result in persons at risk for increased use being missed, and defining a population too broadly might result in wasted resources directed toward people who may not benefit sufficiently. Identifying criteria that quantitatively fit the population of interest helps prevent misspecification. The cluster analysis approach also prevented double counting of people who otherwise might fall into multiple categories, because each person in the population could be assigned to only 1 cluster. This allows overlap between criteria for the same person to be accounted for efficiently without requiring the assignment of risk categories or otherwise defining patient classes.

Certain limitations in study design and implementation were noted. Multiple terms have been used to describe super-utilizers, and this variation in terminology makes it likely that the research team did not capture all documents that could have met the inclusion criteria. The team applied super-utilizer criteria to the Utah Medicaid population using claims data, which might not fully capture the diagnoses, care, and use by individual patients.[33–34] Electronic health records or other sources of health data might capture this information more accurately,[35–38] but were not readily available for the study population. Multiple clusters identified in the analysis were substantial in size and included populations that might be primarily heterogeneous, including persons with multiple chronic conditions. Also, these clusters might not align directly with existing evidence-based interventions and further segmentation may be necessary. There was some question about whether fewer clusters might be appropriate, as it appeared in Figure 2 that Cluster 4 had considerable overlap with the other clusters as projected. However, it was clear from the data in Table 3 that Cluster 4 had distinct properties that distinguished it from the other clusters, particularly with regard to ED utilization. For the cluster analysis, the team selected the CLARA algorithm, based on the Manhattan distance between cluster objects. Other clustering algorithms or dissimilarity measures such as the Euclidean distance might have yielded different results. However, the impact of using alternate algorithms/approaches was not investigated in this exploratory analysis. Categorical dummy variables were used to represent the data because of the interest in fixed criteria obtained from the literature. Other data types could have been used and might have yielded different results.

Study results were intended to be applied to the Utah Medicaid population, and the search methodology was selected based on this functional need. The commonly used criteria collected in this study are not meant to constitute a comprehensive list, but rather to illustrate differences in criteria and establish a starting point for more data-driven approaches to identifying super-utilizers. As the review results demonstrate, determining which criteria to use has often been a subjective exercise. Considering a broader selection of potential criteria and using an analytical approach to determine which criteria are most appropriate can help remove the inherent selection biases of previous approaches. That said, nearly half of the overall population in this study could be categorized within ≥1 of the identified super-utilizer clusters. Institutions employing similar analyses will still need to make decisions about which populations to target and which approaches could have the greatest impact on achieving institutional objectives.

Ultimately, the review and analysis described has identified 21 commonly used super-utilizer criteria and multiple subsets of the Utah population that can be further segmented and targeted for intervention. Future work might include optimizing the collection of possible super-utilizer criteria by identifying and excluding criteria that might be out of date and adding new criteria that better reflect current thinking on super-utilizing or complex patients. Further research also is needed to formulate methods to optimally segment super-utilizer populations and identify the most effective interventions for the resulting subpopulations.

## Conclusion

This study identified super-utilizer criteria that have commonly been used in the literature and applied these criteria to the Utah Medicaid population. The procedures and results described demonstrate how cluster analysis can aid in selecting characteristics from the literature that systematically differentiate super-utilizer groups from other beneficiaries in a population. Other government agencies and health care entities can apply these criteria and similar data-driven approaches to identify super-utilizing people in their own populations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Other staff in the Division of Medicaid and Health Financing also provided expertise and consultation to the authors in clarifying details of the operation of Medicaid in Utah, enabling access to data, fine-tuning queries of the Utah Medicaid Data Warehouse, and improving general understanding of Medicaid organizational structure and principles. The Population Health Steering Committee at the Utah Department of Health provided valuable guidance on the overarching direction and objectives of the project, and colleagues at the Centers for Disease Control and Prevention also provided meaningful input and guidance. This study would not have been possible without the assistance and collaboration of these people.

## References

1. Mann C CMCS informational bulletin: targeting Medicaid super-utilizers to decrease costs and improve quality. Baltimore, MD: Department of Health & Human Services, 2013.
2. Health Care Transformation Task Force. Proactively identifying the high cost population—insights from the health care transformation task force. Washington, DC: Health Care Transformation Task Force, 2015.
3. Mann C CMCS informational bulletin: reducing nonurgent use of emergency departments and improving appropriate care in appropriate settings. Baltimore, MD: Department of Health & Human Services, 2014.
4. Long P, Abrams M, Milstein A, et al., eds. Effective care for high-need patients: opportunities for improving outcomes, value, and health. Washington, DC: National Academy of Medicine, 2017.
5. Chang HY, Weiner JP. An in-depth assessment of a diagnosis-based risk adjustment model based on national health insurance claims: the application of the Johns Hopkins Adjusted Clinical Group case-mix system in Taiwan. BMC Med 2010;8:7. [PubMed: 20082689]
6. Doupe MB, Palatnick W, Day S, et al. Frequent users of emergency departments: developing standard definitions and defining prominent risk factors. Ann Emerg Med 2011; 60:24–32.
7. Moe J, Bailey AL, Oland R, Levesque L, Murray H. Defining, quantifying, and characterizing adult frequent users of a suburban Canadian emergency department. CJEM 2013;15:214–226. [PubMed: 23777993]
8. Mehrotra A, Forrest CB, Lin CY. Dropping the baton: specialty referrals in the United States. Milbank Q 2011;89: 39–68. [PubMed: 21418312]
9. Bivand R classInt: choose univariate class intervals. R package version 0.1–24. Vienna, Austria: R Foundation for Statistical Computin, 2017.
10. R Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013.
11. World Health Organization. International statistical classification of disease and related health problems, Tenth revision (ICD-10). Geneva, Switzerland: World Health Organization, 1992.
12. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005;40:1130–1139.
13. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: cluster analysis basics and extensions. R package version 2.0.6. Vienna, Austria: R Foundation for Statistical Computing, 2017.
14. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Comput Appl Math 1987;20:53–65.
15. Strauss T, von Maltitz MJ. Generalising Ward's method for use with Manhattan distances. PLoS One 2017;12:e0168288. [PubMed: 28085891]
16. Fukuoka Y, Zhou M, Vittinghoff E, Haskell W, Goldberg K, Aswani A. Objectively measured baseline physical activity patterns in women in the mPED trial: cluster analysis. JMIR Public Health Surveill 2018;4:e10. [PubMed: 29391341]
17. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Characteristics of emergency department visits for super-utilizers by payer, 2014. Rockville, MD: Agency for Healthcare Research and Quality, 2006.

18. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Characteristics of hospital stays for nonelderly Medicaid super-utilizers, 2012. Rockville, MD: Agency for Healthcare Research and Quality, 2014.
19. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Characteristics of hospital stays for super-utilizers by payer, 2012. Rockville, MD: Agency for Healthcare Research and Quality, 2015.
20. Rask KJ, Williams MV, McNagny SE, Parker RM, Baker DW. Ambulatory health care use by patients in a public hospital emergency department. J Gen Intern Med 1998;13: 614–620. [PubMed: 9754517]
21. Curran GM, Sullivan G, Williams K, Han X, Allee E, Kotria KJ. The association of psychiatric comorbidity and use of the emergency department among persons with substance use disorders: an observational cohort study. BMC Emerg Med 2008;8:17. [PubMed: 19055761]
22. Markham D and Graudins A. Characteristics of frequent emergency department presenters to an Australian emergency medicine network. BMC Emerg Med 2011;11:21. [PubMed: 22171720]
23. Chan BTB and Ovens HJ. Frequent users of emergency departments: do they also use family physicians' services? Can Fam Physician 2002;48:1654–1660. [PubMed: 12449550]
24. Malone RE. Heavy users of emergency services: social interaction of a policy problem. Soc Sci Med 1995;40:469–477. [PubMed: 7725121]
25. Lucas RH, Sanford SM. An analysis of frequent users of emergency care at an urban university hospital. Ann Emerg Med 1998;32:563–568. [PubMed: 9795318]
26. Byrne M, Murpy AW, Plunkett PK, McGee HM, Murray A, Bury G. Frequent attenders to an emergency department: a study of primary health care use, medical profile, and psychosocial characteristics. Ann Emerg Med 2003;41: 309–318. [PubMed: 12605196]
27. Hunt KA, Weber EJ, Showstack JA, Colby DC, Callaham ML. Characteristics of frequent users of emergency departments. Ann Emerg Med 2006;48:1–8. [PubMed: 16781914]
28. LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med 2010;56:42–48. [PubMed: 20346540]
29. Zwerling C, Miller ER, Lynch CF, Torner J. Injuries among construction workers in rural Iowa: emergency department surveillance. J Occup Environ Med 1996;38:698–704. [PubMed: 8823661]
30. Arcury TA, Gesler WM, Preisser JS, Sherman J, Spencer J, Perin J. The effects of geography and spatial behavior on health care utilization among the residents of a rural region. Health Serv Res 2005;40:135–156. [PubMed: 15663706]
31. Cylus J, Hartman M, Washington B, Andrews K, Catlin A. Pronounced gender and age differences are evident in personal health care spending per person. Health Aff (Millwood) 2011;30:153–160. [PubMed: 21148180]
32. Chen J, Vargas-Bustamante A, Mortensen K, Ortega AN. Racial and ethnic disparities in health care access and utilization under the affordable care act. Med Care 2016;54: 140–146. [PubMed: 26595227]
33. Kharrazi H, Chi W, Chang HY, et al. Comparing population-based risk-stratification model performance using demographic, diagnosis and medication data extracted from outpatient electronic health records versus administrative claims. Med Care 2017;55:789–796. [PubMed: 28598890]
34. Kharrazi H, Weiner JP. A practical comparison between the predictive power of population-based risk stratification models using data from electronic health records versus administrative claims: setting a baseline for future EHR-derived risk stratification models. Med Care 2018;56:202–203.
35. Chang HY, Richards TM, Shermock KM, et al. Evaluating the impact of prescription fill rates on risk stratification model performance. Med Care 2017;55:1052–1060. [PubMed: 29036011]
36. Kan HJ, Kharrazi H, Leff B, et al. Defining and assessing geriatric risk factors and associated health care utilization among older adults using claims and electronic health records. Med Care 2018;56:233–239. [PubMed: 29438193]
37. Lemke KW, Gudzune KA, Kharrazi H, Weiner JP. Assessing markers from ambulatory laboratory tests for predicting high-risk patients. Am J Manag Care 2018;24:e190–e195. [PubMed: 29939509]

38. Kharrazi H, Chang HY, Heins SE, Weiner JP, Gudzune KA. Assessing the impact of body mass index information on the performance of risk adjustment models in predicting health care costs and utilization. Med Care 2018;56:1042–1050. [PubMed: 30339574]

39. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151: 264–269. [PubMed: 19622511]

Author Manuscript
Author Manuscript
Author Manuscript



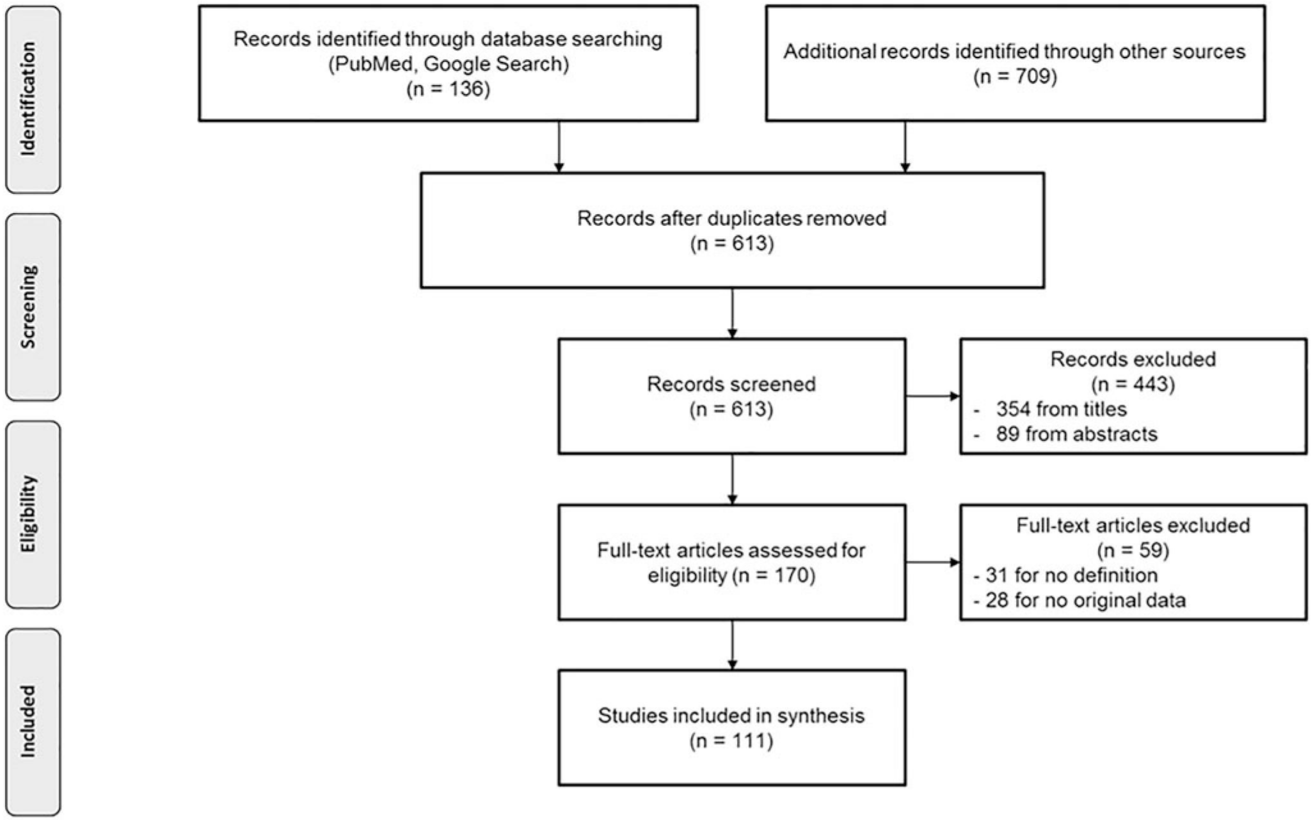


**FIG. 1.**

PRISMA[39] flow diagram of document selection for methodological review of super-utilizer criteria – 2017. Note: other sources include systematic reviews in the PubMed search results and links, footnotes, or citations in the Google Search results.

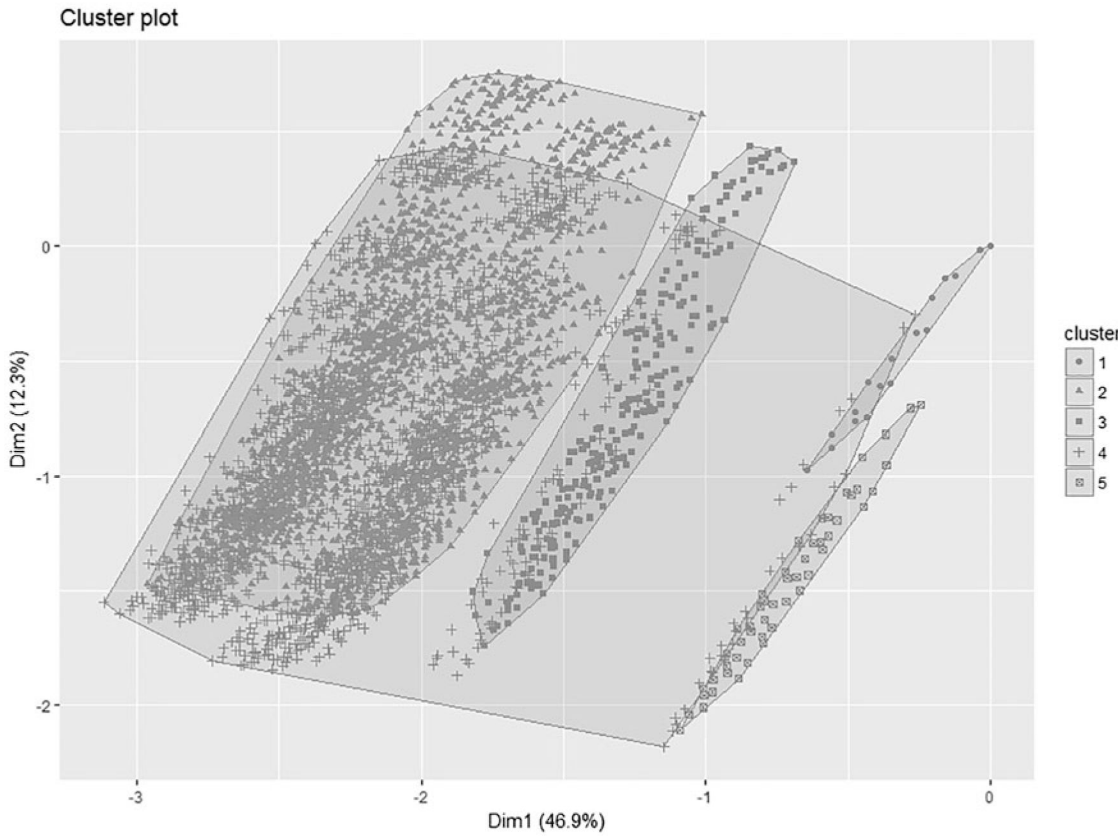


**FIG. 2.**

Clusters of super-utilizers from *k*-medoids algorithm analysis of 20 super-utilizer criteria applied to Utah Medicaid beneficiaries (N = 309,921) with claims during July 1, 2016–June 30, 2017. Note: Points represent beneficiaries and polygons represent clusters of beneficiaries meeting similar super-utilizer criteria (projected into the first 2 principal component dimensions for data visualization).

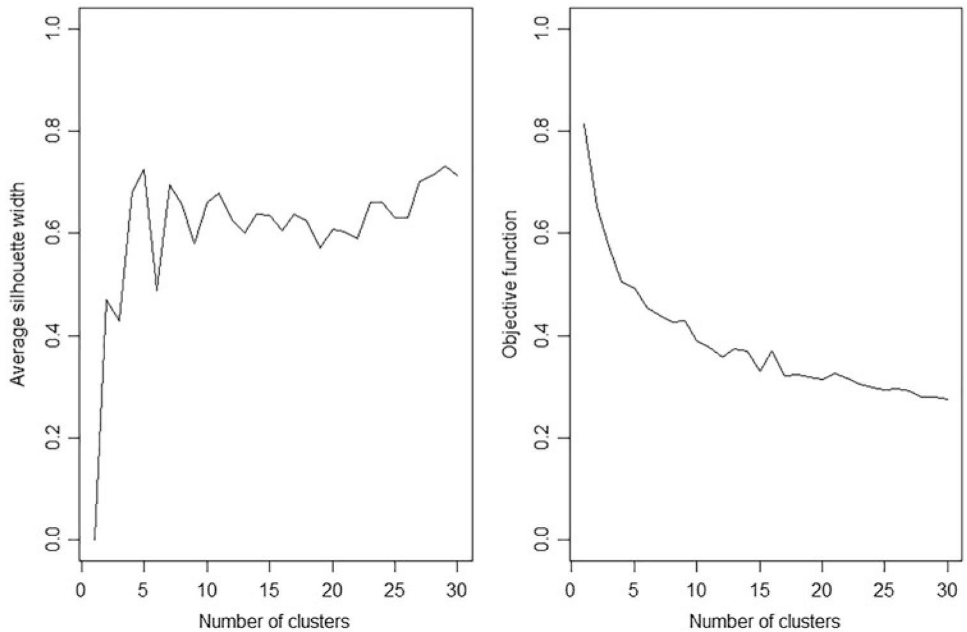


**FIG. 3.**
Number of super-utilizer criteria clusters by average silhouette width (left) and objective function (right).

**Table 1.**

General Super-Utilizer Criteria in the Literature by Centers for Medicare & Medicaid Services Targeting Approach (N = 180 Criteria)

| Targeting approach or subset | No. (%)[a] | No. (%)[b] |
|---|---|---|
| 1. Targeting very high levels of utilization | 160 (88.9) | |
| Based on number of emergency department visits | | 116 (72.5) |
| Based on number of inpatient admissions | | 66 (41.3) |
| Based on top predicted risk score | | 19 (11.9) |
| Based on number of prescribed medications | | 19 (11.9) |
| Based on top cost percentage | | 16 (10.0) |
| Based on other measures of utilization | | 6 (3.8) |
| 2. Targeting by presence of risk factors associated with high preventable costs | 90 (50.0) | |
| Based on chronic health conditions | | 54 (60.0) |
| Based on number of chronic health conditions | | 44 (48.9) |
| Based on behavioral health conditions | | 32 (35.6) |
| Based on social determinants of health | | 17 (18.9) |
| Based on other risk factors | | 22 (24.4) |
| 3. Targeting based on referrals and follow-up investigation | 19 (10.6) | |
| 4. Excluding candidate clients with medical conditions associated with high but non-preventable costs | 8 (4.4) | |
| 5. Targeting specific patterns of care | 8 (4.4) | |
| 6. Targeting based on high observed-to-expected costs | 7 (3.9) | |
| 7. Targeting by community | 5 (2.8) | |

[a] Categories are not mutually exclusive.

[b] Subset percentages are from subtotals.

**Table 2.**

Super-Utilizer Criteria in the Literature That Were Applied in At Least Five Documents and Accounted for At Least 10% of a Criteria Subset (N = 180 Criteria)

| Super-utilizer criterion | No. (% of total) |
|---|---|
| Number of emergency department visits | |
| ≥3 in one year | 11 (6.1) |
| ≥4 in one year | 27 (15.0) |
| ≥5 in one year | 11 (6.1) |
| ≥6 in one year | 16 (8.9) |
| Number of inpatient admissions | |
| ≥1 in one year | 7 (3.9) |
| ≥2 in one year | 5 (2.8) |
| ≥3 in one year | 13 (7.2) |
| ≥4 in one year | 20 (11.1) |
| Top cost percentage | |
| Top 5% | 7 (3.9) |
| Top 10% | 5 (2.8) |
| Chronic health conditions | |
| Asthma | 7 (3.9) |
| Coronary artery disease | 6 (3.3) |
| Congestive heart failure | 8 (4.4) |
| Chronic obstructive pulmonary disease | 8 (4.4) |
| Diabetes | 12 (6.7) |
| Number of chronic health conditions | |
| ≥1 | 29 (16.1) |
| ≥2 | 10 (5.6) |
| Behavioral health conditions | |
| Mental health disorders | 30 (16.7) |
| Substance use disorders | 21 (11.7) |
| Social determinants of health | |
| Homelessness | 12 (6.7) |
| Social needs | 11 (6.1) |

Categories are not mutually exclusive.

**Table 3.**

Characteristics of Five Clusters of Super-Utilizer Criteria Among Utah Medicaid Beneficiaries with Claims During July 1, 2016 to June 30, 2017 (N = 309,921)

| | **Count (%)[a] by cluster (1–5)[b]** | | | | |
|---|---|---|---|---|---|
| **Super-utilizer criterion** | **1** | **2** | **3** | **4** | **5** |
| ≥3 ED visits | *3492 (2)* | *4422 (6)* | *2049 (5)* | **7809 (100)** | *819 (3)* |
| ≥4 ED visits | *1411 (1)* | *1004 (1)* | *550 (1)* | **7346 (94)** | *365 (1)* |
| ≥5 ED visits | *594 (0)* | *0 (0)* | *0 (0)* | 5928 (76) | *168 (1)* |
| ≥6 ED visits | *295 (0)* | *0 (0)* | *0 (0)* | 4124 (53) | *89 (0)* |
| ≥1 inpatient admissions | *0 (0)* | 15,229 (22) | 6539 (15) | 3966 (51) | **27,001 (100)** |
| ≥2 inpatient admissions | *0 (0)* | 7470 (11) | *1656 (4)* | 3125 (40) | *2180 (8)* |
| ≥3 inpatient admissions | *0 (0)* | *2441 (4)* | *377 (1)* | 1850 (24) | *445 (2)* |
| ≥4 inpatient admissions | *0 (0)* | *1300 (2)* | *123 (0)* | 1455 (19) | *99 (0)* |
| Top 3.1% of cost (JNB) | *774 (0)* | *6029 (9)* | *853 (2)* | 1587 (20) | *498 (2)* |
| Top 5% of cost | *1197 (1)* | 9150 (13) | *1459 (3)* | 2596 (33) | *1094 (4)* |
| Top 10% of cost | *2518 (2)* | 16,349 (24) | *3964 (9)* | 4449 (57) | 3712 (14) |
| Asthma | *0 (0)* | 10,249 (15) | 10,937 (25) | 3154 (40) | *0 (0)* |
| Coronary artery disease | *39 (0)* | *3806 (6)* | *191 (0)* | *559 (7)* | *4 (0)* |
| Congestive heart failure | *0 (0)* | *4807 (7)* | *123 (0)* | *734 (9)* | *0 (0)* |
| COPD | *0 (0)* | 8432 (12) | 4306 (10) | 1852 (24) | *0 (0)* |
| Diabetes | *0 (0)* | 12,241 (18) | *1272 (3)* | 1450 (19) | *0 (0)* |
| ≥1 chronic conditions | *0 (0)* | **68,054 (100)** | **43,939 (100)** | **7809 (100)** | *0 (0)* |
| ≥2 chronic conditions | *0 (0)* | 53,194 (78) | *0 (0)* | 6639 (85) | *0 (0)* |
| Mental health disorders | *0 (0)* | 41,380 (61) | *0 (0)* | 3794 (49) | *0 (0)* |
| Substance use disorders | *0 (0)* | 9648 (14) | *2387 (5)* | 2577 (33) | *0 (0)* |
| Total | 163,118 | 68,054 | 43,939 | 7809 | 27,001 |

[a] Cluster percentages are from cluster totals. Note: Column percentages do not sum to 100%.

[b] Criteria representing <10% of a cluster are in *italics*; criteria representing >90% of a cluster are in **bold**.

COPD, chronic obstructive pulmonary disease; ED, emergency department; JNB, Jenks Natural Breaks.