The study titled "Expenditure Variations Analysis Using Residuals for Identifying High Health Care Utilizers" aims to examine the variations of expenditures associated to two medical diagnosis i.e., hypertension and chronic kidney disease. Overall, the study is significant because it attempts to address the hot topic of over-consumption of public health care services and related spending. This appears as a timely article with the merits of highlighting an alternative method with promising features and outcomes that can benefit to the Texas Medicaid program and possibly beyond.

The novel method used seems appropriate and is well described. For instance, the tree-based model aims to better capture N-way interactions between independent variables, especially to account for comorbidities that may greatly influence patients' needs in health care services.

The results reported on comparative outputs from a linear regression and a tree-based model. Authors looked at the profiles of high utilizers that were identified by the models. Their findings tend to support their research hypothesis regarding patients with similar conditions such as demographics and comorbidities who may have drastically different expenses - which in turn suggests the existence of unexplained health care utilization. Then, they investigated whether excessive utilization of care services would persist through time. Their findings confirmed the existence of a correlation structure, which actually showed that unexplained variance did not occur at random from year to year. The temporal consistency finding thus supports a second hypothesis where preventing health care events, which would possibly result from over-utilization, may not be as challenging as expected. In addition, authors tested their findings on preventable health care utilization and expenditure using a comparison with outputs generated by a well-validated software. Their concluding words are strong and reveal a promising use of their novel approach to analyze variations in Medicaid health care expenditures, as well as to better identify high utilizers.

The conclusion is very concise but still drawn and supported by the data. However, several aspects of the study can be raised and I think authors should consider to take care of. The following:
First, the results are very interesting but little or even not discussed. A first paragraph briefly highlights the added value of their method, which is good but the following paragraphs focus on the limitations and the next steps of their research project. Instead, it may be helpful to provide more context and discuss the findings with respect to available literature in this field. Also, the reasons why high utilizers spend more than other people should be further investigated. In addition, the authors should focus greater attention on the potential reasons why certain limitations exist and how they may be overcome. In sum, the manuscript lacks of a proper discussion both on the method and the findings.

Second, the particular issue of comorbidity has been raised by the authors and highlighted as crucial to take into account. In contrast, comorbidity seems poorly documented or discussed in the paper.

Third, while the conclusion efficiently summarizes the study contributions, it lacks of lessons learned for improved and effective policy and practice, especially regarding the group of patients that are potentially "over-utilizing" the health system. Are there specific conclusions to improve the supply or the use of health care services for hypertension (including the result of "the major source of variation was found within one ICD code - ICD-9-CM 4019 of unspecified essential hypertension")? Similarly, what can be learned for the Medicaid management of chronic kidney diseases, for which ICD-9-CM 5856 End-stage renal disease was reported with the main driver of expenditure? Authors should also consider to further detail the discussion and conclusion in order to formulate concrete recommendations regarding both inpatient and EDs.

Finally, I have a few other observations:

a. The method reports on a nonzero expenditures group that has been isolated for analysis. What percent of zero expenditures was found?

b. The various charts are not readable while printed.

c. Be more specific in the title as the paper focused on the two medical diagnosis of hypertension and chronic kidney diseases. Also consider to add a geographical scope in the title i.e., Texas, USA.

d. Page 14 in section Residuals and potentially preventable readmission: line 2 the section identifier is missing.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes
Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal