Author’s response to reviews

Title: Identification and weighting of kidney allocation criteria: a novel multi-expert fuzzy method

Authors:
Nasrin Taherkhani (n.taherkhani@modares.ac.ir)
Mohammad Mehdi Sepehri (mehdi.sepehri@gmail.com;mehdi.sepehri@modares.ac.ir)
Shadi Shafaghi (shafaghishadi@sbmu.ac.ir)
Toktam Khatibi (toktam.khatibi@modares.ac.ir)

Version: 2 Date: 25 Jun 2019

Author’s response to reviews:

Manuscript Number: MIDM-D-18-00478R1

Title: Identification and weighting of kidney allocation criteria: a novel multi-expert fuzzy method

Dear Dr. Cassady-Cain,

We are grateful to have been given the opportunity to revise our manuscript entitled "Identification and weighting of kidney allocation criteria: a novel multi-expert fuzzy method" (MIDM-D-18-00478R1) for BMC Medical Informatics and Decision Making. We have made every attempt to ensure that the reviewers' comments have been adequately addressed in the revised manuscript. We believe these revisions have resulted in a significantly improved manuscript. Our responses to the reviewers’ comments are listed below.

Thank you again for consideration of our revised manuscript.

Yours Sincerely,

Mohammad Mehdi Sepehri

AUTHORS’ RESPONSES TO THE REVIEWER 1’S COMMENTS

Reviewer’s Comment: It is an interesting application of policy development algorithms.
Authors’ Response: We would like to thank you for your precious time and consideration. We appreciate the positive feedback from the reviewer.

AUTHORS’ RESPONSES TO THE REVIEWER 2’S COMMENTS

We thank you for the time and energy expended on our behalf. Your constructive feedback provided us a great opportunity to improve our manuscript. Below, you will find our responses to your comments.

1. Reviewer’s Comment: Table 7 nicely and effectively displays the criteria, sub criteria, and level weights derived from the IF-AHP process. However,

   a. It is unclear how the lowest level weights (e.g., 0.54, 0.29, 0.16 for recipient age) were derived. The supplementary information included the questionnaire that was used for the AHP, showing high level and sub-criteria pairwise comparisons. Were pairwise comparisons also included in the questionnaire to derive these lowest-level weights?

Authors’ Response: Thank you for pointing this out. Yes, pairwise comparisons were used to drive the lowest level weights the same as other levels. Given that there were many questions, only a part of the AHP questionnaire was placed in the supplementary file.

We have completed the questionnaire and uploaded revised version as additional file2 (questionnaire2-revised).

2. Reviewer’s Comment: In Figure 2:

   a. Why are no sub-criteria levels shown for medical urgency, PRA, and waiting time? How can patients be ranked based on differential PRA (or waiting time) values without having level-specific weights for PRA of 98% versus 0%, for example (or waiting time of 5 years versus 2 years)?

   b. It may be worth noting or discussing that several of the 8 sub-criteria do not fit neatly under either the Utility or Equity top-level criteria.

Authors’ Response:

   a. The reason of not considering sub-criteria levels for medical urgency, PRA, and waiting time is as follows:

Medical urgency: In some kidney allocation models such as "ETKAS", several levels for medical urgency are considered [1]. But in some models such as the kidney allocation models in Australia [2] and Iran, the medical urgency is defined as a two-state variable (whether a patient is an urgency or not). As defined at page 11, lines 284-285, we considered the medical urgency as a
two-state variable in this study. Therefore, emergency patients will get point of this criterion and others will not.

To address the mentioned comment, we added explanations in the revised manuscript (page 11, lines 285-289).

PRA: For this factor, the same as medical urgency, only two states were considered (Patients with PRA > 80% and the remainder). The AHP questionnaire also refers to this value for PRA. The value of 80% for PRA is based on a number of kidney allocation models [2] and opinion of the experts who contributed in this research.

To address the mentioned comment, we added explanations in the revised manuscript (page 11, lines 270-272).

Waiting time: Calculated weight for waiting time in AHP method is for one year. Patients will receive points in proportion to their waiting time. For example, a patient who has been waiting for 2 years will get points twice (point = calculated weight in AHP method * 2) as a patient who has been waiting for 1 year (point = calculated weight in AHP method).

To address the mentioned comment, we added explanations in the revised manuscript (page 19, bottom of Table 7).

b. Thank you for raising this important point. As you suggested, we added discussions and explanations about why several sub-criteria do not fit neatly under either the utility or equity top-level criteria (page 13, lines 330-339).

3. Reviewer’s Comment: The description of HLA matching on page 9 is misleading

a. The paper states that “six HLA mismatches mean complete incompatibility.” This is not true. While it is true that a patient can be incompatible with a donated kidney due to HLA, it is because the patient is “sensitized” by having pre-formed HLA antibodies expected to result in hyper acute rejection of the kidney based on the donor’s HLA antigen profile.

i. Hence, the AHP probably should have included “six” among the levels of the HLA mismatch factor, and 6-HLA mismatches would presumably have the lowest weight.

b. The statement “Each person has two of the above three antigens” may confuse readers. Consider, “Each person has up to two different alleles associated with each of the above three antigens.”

Authors’ Response:

a. Thank you for raising this important point. We scanned the concept of HLA matching in organ transplant after you mentioned this point, and the corrections have been made (page 11, lines 264-265).
i. Given the correct point that you stated: "6-HLA mismatches would presumably have the lowest weight.", patients with 6-HLA mismatches should get the lowest point from this factor. In this model, we have not consider point for these patients, therefore it not necessary to add this level in the AHP structure. This does not have any effect on ranking among patients.

b. As you suggested, the correction has been made (page 10, line 263).

4. Reviewer’s Comment: The US kidney allocation system (KAS) is not primarily based on a system of points that, when added together, leads to a prioritized ordering of candidates on the waiting list. Consequently, the comparisons shown in Table 10 are highly misleading.

I suggest removing US-KAS from this table and adding a footnote that the US KAS was not included since it is not based strictly on a points system, but rather is a classification-driven system with points playing a secondary role.

Authors’ Response: We appreciate your suggestion. The correction has been made (page 25, Table 10).

5. Reviewer’s Comment: The long paragraph on Page 4 is confusing, raising these questions

a. It took me several reads to understand that “existing method” = “allocator.” The paragraph could perhaps be reworded for clarity.

b. Is the current Iranian “allocator” the same algorithm used in each of the 15 regions?

c. What is the “observation and human review” step in the current allocation system?

d. What is meant by “location” – within-region distances?

e. How are “location” and “age difference” combined? How are they used together with waiting time?

f. How can it be that existing method “does not pay much attention to criteria that affect graft survival” when “age difference” is one of the two key factors used in the existing system?

Authors’ Response: Thank you for this observation. We added some statements to answer the questions.

a. We reworded the paragraph for clarify (page 4).
b. Yes, the algorithm that used in each of the 15 regions is the same (page 4, line 82).

c. After the successive filters on the waiting list, the algorithm suggests several candidates. Finally, an expert person selects the recipients among the proposed candidate (page 4, lines 99-100).

d. Location is defined based on the distance between the patient's place of residence and the TC. In Iran, the patient's place of residence may be in one region while the TC is in another one. Instead of transporting the kidneys, the patients are transported from their place of residence to the city where they are to be transplanted there (page 4, lines 89-90).

e. “Location” and “age difference” are not combined. After the waiting list is filtered based on the blood type and medical urgency, it will be sorted based on waiting time. Then an expert determines high priorities considering bellow factors:

1. Can the patient reach the TC at the right time?

2. Is the age difference between the patient and the donor appropriate? (page 4, lines 89-92)

   f. Age difference is not the key factors for allocation in the existing system in Iran. Only in the final step, the expert will check that the age difference is not outside the specified range (less than 15 years) (page 4, lines 95-97).

6. Reviewer’s Comment: In step 1 (Delphi), what is the rationale for rejecting factors simply because their weights are below the average of all factor weights?

   a. Even if a factor is not as important as others, wouldn’t it be better for an allocation system to account for it, even if with a relatively low weight?

   b. Was the practicality of conducting an AHP with many versus fewer factors a consideration for limiting the number of factors in step 1?

Authors’ Response:

   a. The goal of step 1 is to identify essential factors using Fuzzy Delphi Method (FDM). Delphi technique is an expert opinion survey method. In this study, we used Delphi technique to reject or accept the factors as important factors in the kidney allocation. The method we used was adapted from Bouzon, 2016 [3].

   b. No, the objective of step 1 is to identify important factors and eliminate unimportant factors. However, limiting the number of factors will make it easier to conduct the step 2 (AHP).
7. Reviewer’s Comment: Consistency ratios were defined but not reported.

Authors’ Response: Consistency ratios were reported (page 18, bottom of Table 5,6). A 2×2 pairwise comparison matrix is always consistent and there is no need to calculate the consistency for this small example [4] (page 17, Table 4).

8. Reviewer’s Comment: The “Model Validation” section of the paper is really more of a “Model Evaluation.” The simulations, comparison of patient rankings, and sensitivity analyses do not, in my opinion, “validate” the IF-AHP model but rather just evaluate it in different ways. I recommend changing the section to “Model Evaluation.”

Authors’ Response: We agree with you, therefore we have changed "Model Validation" to "Model Evaluation" (Page 19, line 451-452; page 24, line 545).

9. Reviewer’s Comment: The way the model was evaluated using data from Oct – Dec 2017 is unclear. It just says “We ran the proposed model…”

a. Was it a stochastic simulation? How were waiting times estimated?

b. Was the analysis performed assuming one national system, or separately within each of the 15 Iranian regions?

c. Table 8 shows that the average waiting time is projected to decline 26% (1.7 to 1.26 years) with the proposed allocation system. Clearly under a new system, the waiting time for some groups of patients will increase, and others will decrease. But how do you explain such a large drop in projected average waiting time just by reordering the list? Typically, such large waiting time reductions are only possible with a large increase in the number of available donors (or a reduction in demand for organs).

d. Medical urgency was found to be the most important variable in the AHP, and hence in the proposed allocation system. Yet the average waiting time actually increased slightly for medically urgent patients in the new system, whereas in the old system waiting times were approximately half of average. Can the authors explain this counterintuitive result?

e. Can the authors show projected waiting times (or at least the distribution of transplants) by blood type and PRA to help further assess the impact of this proposed allocation system on equity?

f. The results “20 of 22” and “18 of 22” are unclear. I believe these are number of pediatric transplants out of pediatric patients waiting, but this could be made more clear.

Authors’ Response:
a. To evaluate the proposed model, we used the real dataset rather than stochastic simulation. Therefore, it was not necessary to estimate waiting time. Waiting time was calculated based on the available dataset and the allocation time.

\[
\text{Waiting time} = (\text{registration time}) - (\text{allocation time})
\]

The proposed model was implemented using the kidney transplantation dataset in Tehran (one of the 15 regions that has longest waiting list) from October 2017 to December 2017. This dataset includes information about both the patients registered on the kidney waiting list (484 registered patients) and deceased donors (124 donors). The proposed model was run for each donated kidney and the chosen patients of each run were recorded. Chosen patients of existing system were available. Results of the existing system and the proposed one have been compared. We added explanations in the manuscript for more clarity (page 20, lines 455-457).

b. The analysis performed within one region. Tehran, with the longest waiting list and the highest number of TC and OPU, has been selected as case study. We added explanations in the manuscript for more clarity (page 19-20, line 453-454).

c. The point you mentioned is exactly true: "Typically such large waiting time reductions are only possible with a large increase in the number of available donors (or a reduction in demand for organs)." This reduction in projected average waiting time for us was also unexpected. Maybe this reduction is caused because the limitation of the identical blood type for allocation has been considered in the existing model, while it has been eliminated in the proposed model.

d. It is true that medical urgency is the most important factor in the proposed model, but the allocation is based on the total score of all factors. As described in page 4 of the paper, in existing system in Iran, the waiting list is filtered based on the blood type and medical urgency. In cases where an emergency patient is on the list, he/she will be selected for allocation, and non-emergency patients will be ignored. Therefore, it is logical that the existing system has a shorter waiting time for emergency patients.

e. Thank you for your comment. We added some of candidate factors (e.g. blood type, age, waiting time) to Table 9 (page 22, Table 9).

f. Yes, these are number of pediatric transplants out of pediatric patients waiting. We added explanation in Table 8 for more clarity (page 20, Table 8).

10. Reviewer’s Comment: Table 9 is very useful, but consider improving it in these ways:

a. Add ranks (1-30) for the proposed vs. current models to the table, to allow more direct rank-ordering comparisons.
b. Add candidate factors (e.g. PRA, medical urgency, waiting time). Consider removing Mean-diff columns to save space.

c. Replace “/” with “.” throughout the table.

Authors’ Response:

   a. Yes, you are right. Ranking of the current system were not added to the table, because only the six first priorities in the existing system are ranked for each donated kidney (which is referenced in the text: page 4, line 89) and the remainders are not ranked. So this information was not available to be added the table.

   b. As you suggested, the correction has been made (page 22, Table 9).

   c. As you suggested, the correction has been made (page 22, Table 9).

11. Reviewer’s Comment: Figure 3 is very confusing.

   a. Firstly, the use of “region” is confusing in the context of organ allocation and distribution. Consider changing to “quadrant” in the figure and associated text as well.

   b. Connecting the points with lines is unnecessarily confusing.

   c. Consider labeling points with their ranks (1-30, in order of AHP-derived points), instead of their (meaningless) patient ID’s.

   d. For direct comparison, consider adding a second figure showing patient ranks under the current system.

   e. Considering narrowing the axis ranges so that variations among the data points is more evident

Authors’ Response: Because we added a figure for Triangular Fuzzy Number (TFN), the number of figures has been changed.

   a. Thank you for your comment. As you suggested, the correction has been made (Figure 4; page 22, line 504-506).

   b. The correction has been made (Figure 4).

   c. The correction has been made (Figure 4).
d. As mentioned before, only six first priorities are ranked for each donated kidney and the rest of the patients are not ranked in the current system. So this information was not available to be presented a new figure.

e. We narrowed the axis range as you recommend, but this change made the points too close together which made their separation very difficult. So we did not change the axis ranges.

Other comments:

Reviewer’s Comment: In the abstract and conclusion, it would be more appropriate to state that the proposed model “has the potential to improve” allocation outcomes, as opposed to a definitive statement that the model will perform better than the current system.

Authors' Response: We appreciate your suggestion. The correction has been made (page 2, lines 39-40; page 24, line 546).

Reviewer’s Comment: The background states that fuzzy approaches are “better” when “uncertainty is high.” The authors could elaborate more on this type of uncertainty, i.e., how other researchers would recognize that they have a setting of “high uncertainty.” Is it whenever a linguistic scale is used?

Authors’ Response: We agree with your comment and, therefore, added some statements to elaborate more on this type of uncertainty (page 7, lines 171-177).

Reviewer’s Comment: The paper talks about “linguistic variables” and a “linguistic scale.” Table 1 uses both of these terms, in fact. I believe the authors are using both terms to refer to the same thing, a “linguistic scale” (e.g., Absolute importance, Very strong importance, etc.). I found the mention of “linguistic variables” to be confusing. Is it the same as a linguistic scale?

Authors’ Response: Thank you for this observation. Yes, linguistic variable is the same as a linguistic scale. We changed "linguistic variable" to "linguistic scale" throughout the manuscript (page 6, line 142; page 8, line 216; page 9, line 220, 221, 231, Table 1).

Reviewer’s Comment: “Transplant status” can mean many different things in various allocation systems. A more descriptive term for this factor might be “First time vs. repeat”.

Authors’ Response: As you suggested, we added the more descriptive for this factor (page 12, line 294).
Reviewer’s Comment: In Table 2, is the reason there are two fuzzy sets for each importance level, instead of just one, because of the use of the intuitionistic approach, as explained in the appendix?

If so, it would be helpful to make this more clear in the main body of the paper.

Authors’ Response: Yes. That’s right. We added some statements to explain the intuitionistic fuzzy set (page 14, lines 347-354).

Reviewer’s Comment: I find the wording “decision-makers’ skeptics” to be awkward and unclear.

Authors’ Response: We changed "decision-makers' skeptics" to "hesitancy of decision makers" (page 2, line 45).

Reviewer’s Comment: If “predicted survival” was based on the EPTS score, Table 7 should make that clear. Perhaps just put “(EPTS)” after Predicted survival.

Authors’ Response: Predicted survival is not necessarily considered based on EPTS. Different methods for predicting survival may be used in different countries. In this study, we used the EPTS because there was no factor to indicate the outcomes of the transplant in Iran.

Reviewer’s Comment: Figure 2: Under recipient age, it would seem that a fourth category “adults” is needed for completeness. Same with HLA mismatches=6.

Authors’ Response: Similar to the explanation given for comment 3 (HLA mismatches=6), in this model, we have not considered point for adult patients. Only patients < 18 years will get points.

Reviewer’s Comment: How was the non-membership impact factor determined in this analysis?

Authors’ Response: Value of the non-membership impact factor is determined by decision makers according to the type of problem. In this research, according to expert opinion, we set its value to 30.

Reviewer’s Comment: Throughout the paper, it would be better to use the phrase “Blood type compatibility” (identical, compatible) instead of “ABO blood type,” so as to avoid potential confusion with another way to use blood type in kidney allocation, namely directly prioritizing candidates based on their blood types to improve equity, since certain blood types often have longer waiting times than others.
Authors’ Response: Following your useful comment, we changed the "ABO blood type" to the "Blood type compatibility", throughout the paper (page 10, line 255; page 17, Table 3; page 19, Table 7).

Reviewer’s Comment: Is a reference available to cite for IRNOPT?

Authors’ Response: Thank you for this observation. We added a reference to cite the IRNOPT (page 3, line 79).

Reviewer’s Comment: The first two paragraphs on page 13 are still “Methods” but appear in “Results.”

Authors’ Response: We agree with you. Therefore, the two paragraphs were transferred to Section "Methods" (page 7-8, lines 185-210).

Reviewer’s Comment: It seems incomplete to cite the Bersimas paper’s use of “linear regression” to determine priority weights without mentioning that the thrust of their approach relies upon formal mathematical optimization.

Authors’ Response: We added some statements to complete this citation (page 5, lines 132-133).

Reviewer’s Comment: Page 6 suggests that traditional AHP is definitively “not suitable” for evaluating a linguistic scale. Is this too strong of a statement? The Conclusions suggest that it is worthwhile future research comparing traditional vs. fuzzy vs. IF-AHP methods in this setting.

Authors’ Response: As you suggested, we modified this statement (page 6, lines 159-160).

Reviewer’s Comment: Table 8:

- Clarify which results are all vs. adult-only recipients.
- EPTS spelling

Authors’ Response: The correction has been made (page 20, Table 8)

Reviewer’s Comment: It would be useful to explain why a 5-point importance scale was used instead of Saaty’s 9-point scale.

Authors’ Response: As you suggested, we added some statements to explain why the 5-point scale was used (page 13-14, lines 344-346).
Reviewer’s Comment: As a non-expert in fuzzy methods, I was not immediately clear what a “crisp” number is. It might be worth parenthetically defining this term at its first mention for the novice reader. Same with “fuzzified.”

Authors’ Response: We truly appreciate your comment. We added some statements to explain the crisp number and fuzzified (page 6, lines 140-142; page 9, line 235).

Reviewer’s Comment: In Formula (1)

- What if x=m?
- It might be useful to have a figure showing the triangular shape of the function.
- Spelling of otherwise

Authors’ Response: The correction has been made. (Page 9, line 226; page 9, Figure 2 is added)

Reviewer’s Comment: In Table 2, it doesn’t seem that the third column (Reciprocal importance level) adds any value. I only found it to be confusing.

Authors’ Response: Thank you for your useful comment. We deleted the third column in Table 2 (page 14, Table 2).

Reviewer’s Comment: Spell out the meaning of TIF on page 11.

Authors’ Response: The correction has been made (page 13, line 344).

Reviewer’s Comment: Line 223/224 is not a proper English sentence.

Authors’ Response: We rewrote these sentences (page 10, lines 256-258).

Reviewer’s Comment: Line 235: change found to find.

Authors’ Response: The correction has been made (page 11, line 269).

Reviewer’s Comment: Line 236: change equity to equitable; change “to” to “above”

Authors’ Response: As you suggested, the correction has been made (page 11, line 270).
Reviewer’s Comment: Line 354, 356: “transported” would be a better word choice than “transmitted”

Authors’ Response: The correction has been made (page 16, lines 401, 402, 404).

Reviewer’s Comment: Lines 362-364 suggest that predicted survival (EPTS) was rejected as being too unreliably calculated, yet it appears Table 3 and the rest of the analysis as not having been rejected. This was confusing.

Authors’ Response: "Predicted survival" was not rejected. Regarding that many interfering factors are needed to predict the graft survival, and practically these factors are less well recorded and followed in Iran; therefore, we would have difficulty in calculating this factor. To address the mentioned comment, we added explanations in the revised manuscript (page 16, line 408).

Reviewer’s Comment: Line 377: Table 3 should be Table 10.

Authors’ Response: As you suggested, the correction has been made (page 17, line 425).

Reviewer’s Comment: Line 439-440: referring to patients as “priorities” is confusing.

Authors’ Response: The correction has been made (page 22, lines 503,504).

References


