Reviewer’s report

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

Version: 1 Date: 13 May 2019

Reviewer: Darren E. Stewart

Reviewer’s report:

I am glad to have had the chance to review this fascinating and potentially valuable approach to designing organ allocation systems. I had recently become familiar with the potential use of AHP for prioritizing patients on waiting lists, so the opportunity to review came at a good time. The use of multi-criteria decision-making methods clearly has the potential to help structure difficult conversations among policymakers about tradeoffs between equity and utility-related factors used to prioritize candidates on waiting lists.

The paper is extremely well referenced and has a number of strengths. For example, I found the explanation of why the three factors (location, prior living donor, transplant status) were rejected from step 2 (AHP) to be insightful. It is unclear from this paper whether the IF-AHP approach is superior than either fuzzy AHP or traditional AHP, but I appreciate the authors’ alluding to these comparisons as valuable future work.

I have the following major critical comments:

1. Table 7 nicely and effectively displays the criteria, subcriteria, 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?

2. 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. For example,
i. Medical urgency could be considered a "utility" factor prioritizing candidates with the highest waiting list mortality rates.

ii. In some allocation systems, ABO blood type identical transplants are prioritized over compatible transplants not only because of better expected post-transplant outcomes (utility), but also to improve equity (e.g., because ABO=O candidates are only biologically compatible with ABO=O kidneys).

iii. One could also argue that prioritizing younger recipients is related to utility, in terms of maximizing the value and benefit of receiving an organ (to reduce childhood developmental problems) as well as increased graft longevity (since younger patients have longer to live post-transplant).

3. 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 hyperacute rejection of the kidney based on the donor's HLA antigen profile. Completely incompatibility is not a result of the degree of tissue type-matching (e.g., ABDR mismatching of 0 up to 6 alleles). Although less desirable than low-level mismatches, six-allele mismatch transplants are not uncommon, and are only considered marginally inferior to 5 or 4 HLA mismatch transplants.

   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."

4. 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. While it is true that points are used in KAS, they serve secondary importance to the role of "classifications" in the allocation system. Patients are first ordered by classification, and points are only used to further sort patients within classifications. Examples of how Table 10 is misleading include

   a. Prior living donors receive 4 points under KAS, but the primary means by which this highly preferred group is prioritized is through their "classification," which gives them priority over most other candidates for kidneys from local (nearby) kidneys. Prior living donors, in fact, receive transplants faster than any other group in the U.S.
b. Similarly, pediatrics (age<18) receive very high priority under KAS due to an elevated local classification that is not reflected by purely evaluating the relative importance of these factors based on "points" alone.

c. Zero-ABDR mismatches are very highly prioritized under KAS, not because of points but because of highly elevated "classifications."

d. Extremely highly sensitized candidates (CPRA 98, 99, 100) receive priority in the form of elevated classifications (including broader geographic distribution of kidneys to these patients), which far exceeds the impact of their points. CPRA 86-97% patients, by contrast, are prioritized below prior living donors and pediatric candidates, due to the latter groups' elevated classifications, so to say "CPRA>85%" are the most highly prioritized patients is misleading.

e. Simultaneous liver-kidney and other multi-organ renal transplant combinations are de facto prioritized ahead of all kidney-alone candidates under KAS, so showing that SLK's are highly prioritized in ETKAS but not in US-KAS is misleading.

f. All that said, I do like Table 10, it's just that it is not possible to accurately include the US system (and possibly other countries as well) because it is not based on a system based purely on allocation points. 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. You could state that KAS awards very high priority for multi-organ renal transplant recipients; extremely highly sensitized patients (CPRA 98-100%); zero-ABDR mismatches; prior living donors; pediatric candidates; and patients with high waiting times.

5. 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?

6. 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?

7. Consistency ratios were defined but not reported.

8. 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."

9. 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.

10. 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.

11. 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.

Other comments
- 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.
- 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?
- 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?

- "Transplant status" can mean many different things in various allocation systems. A more descriptive term for this factor might be "First time vs. repeat".

- 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.

- I find the wording "decision-makers' skeptics" to be awkward and unclear.

- If "predicted survival" was based on the EPTS score, Table 7 should make that clear. Perhaps just put "(EPTS)" after Predicted survival.

- Figure 2: Under recipient age, it would seem that a fourth category "adults" is needed for completeness. Same with HLA mismatches=6.

- How was the non-membership impact factor determined in this analysis?

- 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.

- Is a reference available to cite for IRNOPT?

- The first two paragraphs on page 13 are still "Methods" but appear in "Results."

- 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.

- 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.

- Table 8:
  - Clarify which results are all vs. adult-only recipients.
  - EPTS spelling
- It would be useful to explain why a 5 point importance scale was used instead of Saaty's 9-point scale.

- 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."

- In Formula (1)
  o What if $x=m$?
  o It might be useful to have a figure showing the triangular shape of the function.
  o Spelling of otherwise

- In table 2, it doesn't seem that the third column (Reciprocal importance level) adds any value. I only found it to be confusing.

- Spell out the meaning of TIF on page 11.

- Line 223/224 is not a proper English sentence.

- Line 235: change found to find

- Line 236: change equity to equitable; change "to" to "above"

- Line 354, 356: "transported" would be a better word choice than "transmitted"

- 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.

- Line 377: Table 3 should be Table 10.

- Line 439-440: referring to patients as "priorities" is confusing.

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

No

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

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

No

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 am able to assess the statistics

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

Needs some language corrections before being published

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