A Fuzzy Multi-criteria Decision Making Approach for Analyzing the Risks and Benefits of Opening Data

Ahmad Luthfi, Zeenat Rehena, Marijn Janssen, Joep Crompvoets

A Fuzzy Multi-Criteria Decision Making Approach for Analyzing the Risks and Benefits of Opening Data

Ahmad Luthfi[1,2 *][0000-0001-5416-1529], Zeenat Rehena[1,3 [0000-0002-0689-1903]]
Marijn Janssen[1][0000-0001-6211-8790], Joep Crompvoets[4][0000-0003-1077-597X]

[1]Delft University of Technology
Faculty of Technology, Policy and Management
Jaffalaan 5, 2628 BX Delft, The Netherlands
{a.luthfi, z.rehena, m.f.w.h.a.janssen}@tudelft.nl
[2]Universitas Islam Indonesia, Yogyakarta, Indonesia
ahmad.luthfi@uii.ac.id
[3]Aliah University, Kolkata, India
zeenatrehena@aliah.ac.in
[4]Katholieke Universiteit Leuven, Leuven, Belgium
joep.crompvoets@kuleuven.be

Abstract. Governments are releasing their data to the public to accomplish benefits like the creation of transparency, accountability, citizen engagement and to enable business innovation. At the same time, decision-makers are reluctant to open their data due to some potential risks like misuse, sensitivity, ownership, and inaccuracy of the data. The goal of the study presented in this paper is to develop a Fuzzy Multi-Criteria Decision Making (FMCDM) approach to analyze the risks and benefits to determine the decision to open a dataset. FMCDM is chosen due to its capability to measure and weight the relative importance of the criteria. FMCDM need the weighting of criteria as input. For this Fuzzy Analytical Hierarchy Process (FAHP) is utilized by collecting input from experts' knowledge and expertise. The scores for each criterion are summed up to rank the importance of the alternatives. Four main criteria are used, e.g. data sensitivity and data ownership representing risks criteria, and data availability and data trustworthy as benefits criteria. For each criterion, there were two sub-criteria identified. Four types of decisions to open data can be made: completely open, maintain suppression, provide limited access, and remain closed. A health patient record dataset is used to illustrate the approach. In further research, we recommend to develop automated approaches that take a dataset as an input and can provide an advice.

Keywords: Multi-criteria Decision Making, Fuzzy, Analytic Hierarchy Process, Open Data, Risks, Benefits

1 Introduction

The motivation to open data by governments and private organizations have increased extensively over the last few years. The creation of transparency and accountability,

to sustain citizen engagement and to enable business innovation are the main drivers to open more data [1-4]. The disclosure of data is expected to improve decision-making initiatives by both government and society [3, 5]. Furthermore, the opening of data by organizations can improve an organization's reputation by showing that they are an open institution [6].

However, although divers types of datasets have already been opened [7-9], in reality many datasets are still not opened [10]. There are several reasons why the data providers reluctance for opening datasets, including: (1) barriers of implementing the systems [11, 12]; (2) risks like inaccuracy, misuse, sensitivity, and inconsistency of the data [3, 10, 12-17]; and (3) inappropriate interpretation of the data resulting in an inadequate comprehend of the data [3]. Moreover, mistakes in interpreting data or misuse of data can jeopardize the reputation of data providers [11]. This result in many datasets to remain closed, whereas this might not be necessary.

The goal presented in this study is to develop a Fuzzy Multi-Criteria Decision Making (FMCDM) approach to analyze the risks and benefits and to determine the best alternative decision for a given dataset. The use of Fuzzy set theory in this research is to manage decision-making problem of alternative selection of a dataset status. These alternatives are developed by establishing and incorporating the FMCDM based on Fuzzy Analytic Hierarchy Process (FAHP) [18, 19]. The main function of the Fuzzy logic is to capture the expertise of open experts and to express it with computational approach [20-22]. A Fuzzy theory is based on the intuitive reasoning by considering the human subjectivity and incorrectness, which are common in the natural language [23]. The natural language is an intricate structure both in the human communication and the way how the human being thinks [23, 24].

Fuzzy theory is used in this paper to provide a mathematical strength for the emulation of the higher cognitive function from the human thought and perception associated with weights of the risks and benefit of opening data. The main function of the FMCDM is to assess the alternative selection with respect to predetermined criteria for a single decision making [25]. The appropriateness of the alternative compares to the criteria and the priority weights of each criterion can be analyzed and computed using linguistic matrix values reflected by the fuzzy [20, 26]. FAHP, furthermore, is used to determine the preference weightings of criteria by collecting expert's judgment [18, 27]. The scores for each criterion are summed up to rank the importance of the alternatives [28, 29].

This FAHP technique used in this study consists of the six following steps [18, 19, 27], namely: (1) select experts team; (2) determine the evaluation criteria and construct the hierarchy, including alternatives; (3) construct pairwise comparison matrix and evaluate the relative importance of the criteria; (4) transform the linguistic terms into triangular fuzzy number; (5) calculate the Fuzzy weights matrix, and check the consistency of the pairwise comparison matrix; and (6) select the best alternative. A dataset of health patient records is used in the illustration part to show how the risk and benefit multiple criteria can be analyzed by employing the FMCDM approach. The four possible decisions are completely open, maintain suppression, provide limited access, or remain closed. These are the alternatives for the FMCDM and decisions on these alternatives will be analyzed based on the four main criteria, namely

data sensitivity and data ownership for the risk criteria, while data availability and data trustworthiness are the criteria for the benefit. Data sensitivity and ownership are selected as input because of these criteria can represent some privacy violation issues containing in health patient records dataset. For example, in the case of data sensitivity, by releasing the actual value of name, date of birth, place of birth, home address, or insurance provider of a patient, it might be potentially misused by the unauthorized users. In addition, data availability and data trustworthiness are chosen criteria due to they can reflect the benefit of transparency and accountability in opening data. Each of the criteria has sub-criteria to further refine the risks and benefits. In Section 3.3, we will explain the sub-criteria definition and relationship in more detail.

This paper is consists of six sections. In Section 1 the rationale of this research is presented, Section 2 contains the related work of decision-making to open data. In Section 3 the approaches are described, including proposed flow process, alternatives, and criteria selection for FMCDM which is based on FAHP method. Section 4 provides the illustration and results. Section 5 some findings of the study are provided. Finally, the paper will be concluded in Section 6.

2   Related Work

In order to present the current approaches of decision analysis in the domain of open data, we reviewed literature which is summarized in Table 1. We found three limited works about decision-making analysis for opening data. Existing work uses the following methods: (1) trade-offs method to weigh the values and risks of open data by conducting interview sections with exclusive groups like civil servants and archivists, (2) decision-support framework to develop a prototype based on the open data ecosystem for specific groups like business and private organization, and (3) an iterative method using Bayesian-belief Networks to weigh the risks and benefits of opening data.

Table 1. The previous methods of Decision support for opening data

	Name of approach	Research overview and boundary
1	Trade-offs method [10]	• A decision-making model offers respectfully capturing trade-offs, and in this method also providing guidance for weighing the potential values and risks of opening data. • No specific algorithm to weigh the risks and benefits of open data. • The trade-offs model can only be used for decision-making with Boolean expression (open or closed) decision.
2	Decision Support Framework [30]	• A developed prototype is based on the concept of open data ecosystems.

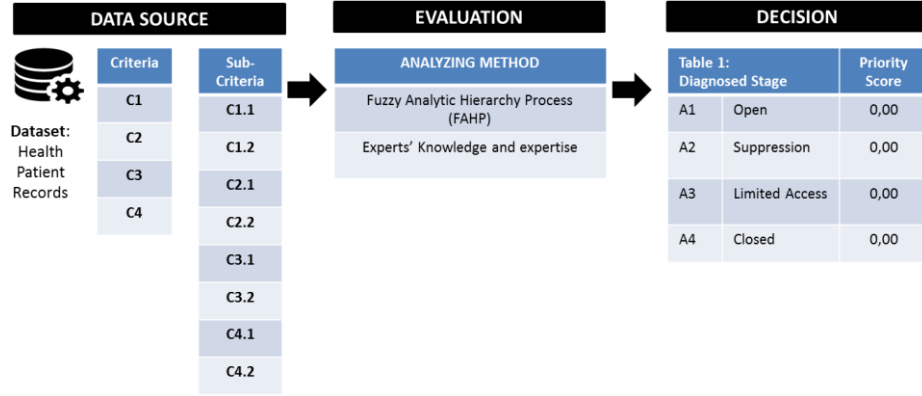|		• The proposed model is specifically for business and private organizations. • There was no evaluation and assessment model available
3	Iterative Model of Decision Support for opening data [31]	• An Iterative model decision support is based on the Bayesian-belief Network analysis. • The outcomes can be used to mitigate the risks and still gain benefits of opening data by taking action removing privacy-sensitive data from a dataset. • There are two alternative decisions available (Open or Closed).

Yet, none of these related works utilized an FMCDM approach in a sense to measure and determine the best alternative for deciding a single status of a dataset. Some possible advantages the use of the FMCDM approach compare to three other methods are: (1) the capability to consider the human subjectivity and incorrectness from the common natural language [32]; (2) provides assessment of the alternatives selection with respect to predetermined criteria for a single decision making [25]; and (3) its simplicity characteristic to evaluate multiple conflicting in decision-making as one of the most popular problems handled by researchers in the literature [25, 32].

3 Decision-making Approach

In this section, we aim to describe the decision-making approach for analyzing risks and benefits of open data. Four subsections are described, namely flow process of the proposed method, alternatives, selection of criteria, and FAHP technique.

3.1 Flow Process of Proposed Method

To describe how the FMCDM approach works, we use a flow of decision-making process having three main phases, namely data source, evaluation, and decision. The entire process starts with the selection of the dataset from the data source to create the input for the evaluation phase. The input data are processed next in the evaluation phase. The output of the evaluation namely decision stage is a suggestion to make a decision. The latter is done by showing the rank of decision priority (decision), as shown in Figure 1.

Fig. 1. The flow process of the approach

The flow process is based on the data source, evaluation of input data (data source) and decision. Figure 1 illustrates the staging of analyzing the risks and benefits of opening data, and it can be narrated as follows:
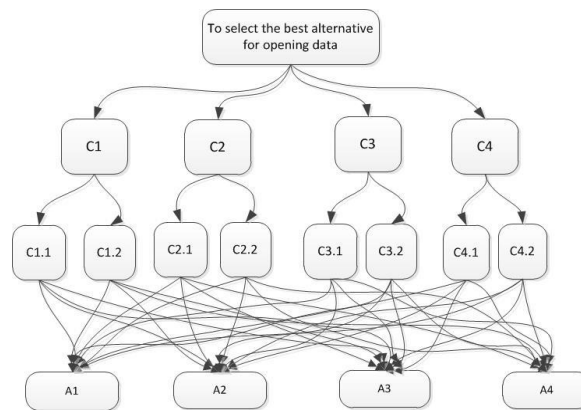
- **Data Source**: First, we need to select the type of a dataset. For example, in this study, we have chosen health patient records and table 1: diagnosed stage (see Figure 3) as the object to be analyzed. To define the criteria and sub-criteria, an extensive literature review related to the risks and benefits of opening data has done in Section 2. In this study, we designed four criteria and eight sub-criteria of the risks and benefits as the input data.
- **Evaluation**: In the second stage, we used FMCDM to assess the alternatives based on criteria defined in the data source elicitation phase and the criteria uses linguistic matrix values reflected by the Fuzzy. FMCDM works on Fuzzy AHP technique has an essential role to measure the relative importance of defined criteria for dealing with decision-making problem. To quantify the relative importance of the risks and benefits, we picked up the knowledge from the experts' judgment. There are two main steps to conduct an evaluation process by the experts in AHP, as follows [27, 33]: To begin with, experts should rank the criteria in a descending or ascending order of their significance. Then, determining the most important criteria and compare it with others. For example, an expert ranked that data sensitivity (C1) is higher or essentially important than data ownership (C2). Second, experts will determine the criteria weights by transforming pairwise comparison matrix into a triangular fuzzy number, as can be seen in Figure 5.
- **Decision**: Finally, the outcome of this flow process is to get the final weights of the best alternative as the priority of a decision.

6

3.2 Alternatives

The following four alternatives of opening data in this paper are: opening the dataset (A1), maintaining a dataset suppression (A2), providing limited access (A3), or keeping the dataset closed (A4). First, the alternative "open the dataset" is defined as publishing the dataset presents a low risk to an individual or organization identity, and/or the potential benefits of the dataset substantially outweigh the potential risks. Second, the alternative "maintaining suppression" is specified as removing a data field and/or an individual record into particular groups or generate unique characteristics to avoid the personal identity. In this alternative, data that might create significant risks are not opened in the actual form, as the potential benefits do not outweigh the possibility of the risks. Third, the alternative "limited access" defines that only a certain group will be given access to the data. The level of openness is limited. Often those who will gain access have to sign a document that outlines the rules of access. The reason for this is releasing the dataset will create a moderate risk, or potential benefits of the dataset do not outweigh the potential privacy risks. Fourth, the alternative "keeping the dataset closed", it means that by publishing the dataset generates a very high risk to an individual or organization and significantly outweigh the potential benefits.

3.3 Selection of Criteria

Figure 2 represents the hierarchy of the four criteria, eight sub-criteria, and four alternatives. The four criteria C1, C2, C3 and C4 define data sensitivity, data ownership, data availability, and data trustworthy respectively. The data sensitivity (C1) composes of two sub-criteria: individual life-threatening (C1.1) and data identifiable (C1.2). Individual life-threatening (C1.1), can be defined as a potential risk to an individual or personal life because of the possibility to recognize the sensitive value of the dataset. Data identifiable (C1.2) is specified as the potential leak of the personal, organizational, business or even government data identifiable e.g. by combining some attributes of the field.

Fig. 2. The hierarchy of criteria and alternatives

The second criterion is data ownership (C2) which consists of two sub-criteria namely metadata scanning (C2.1) and fake or misleading (C2.2). Metadata scanning (C2.1) can be represented to figure out the property and structure of the dataset. Fake or misleading (C2.2) is defined by a user to potentially change and modify the dataset and affect an unreliable and wrong decision. Data availability (C3) is the third criterion and it has two sub-criteria namely data manageability (C3.1) and data recoverability (C3.2). Data manageability (C3.1) is specified as the chance to manage the availability and accessibility of the dataset. Data recoverability (C3.2) is indicated by delivering a dataset and it can have a highly positive impact on recovering the availability of the data. The fourth criterion is data trustworthiness (C4) which consists of two sub-criteria like data traceability (C4.1) and data authenticity (C4.2). Data traceability (C4.1) can make the possibility to trace the source of the dataset. Data authenticity (C4.2) is defined as the potentially affected to recognize the authentication of the data.

3.4   Fuzzy AHP Technique

The AHP process is a quantitative method that deals with the multi-attribute, multi-criteria, multi-period problem hierarchically [34]. Only with AHP, it is not possible to overcome the deficiency of the fuzziness during decision making [35]. Hence, in this study, the Fuzzy AHP which is the extension of the conventional AHP method by integrating fuzzy comparison ratios is used for multi-criteria analysis [18, 27, 34, 36]. It uses the triangular fuzzy number of fuzzy set theory directly into the pairwise comparison matrix of the AHP. The geometric mean method is used to generate fuzzy weights and performance scores [37]. The steps of the Fuzzy AHP can be summarized as follows:

- Step 1. Select experts. The quality of the evaluation process depends on experts' knowledge and experience. Hence the selection of experts is crucial.
- Step 2. Determine the evaluation criteria and construct the hierarchy including alternatives.
- Step 3: Construct pairwise comparison matrix and evaluate the relative importance of the criteria. The experts are expected to provide their judgment on the basis of their knowledge.

    For any expert the comparison matrix is given by eq. (1) as:

a) $\tilde{C}_k = \begin{bmatrix} 1 & \tilde{c}12 \cdots & \tilde{c}1n \\ \vdots & \ddots & \vdots \\ \tilde{c}n1 & \tilde{c}n1 \cdots & 1 \end{bmatrix}$ (1)

    where n is the number of criteria, $\tilde{C}_k$ is a pairwise comparison matrix belongs to k^{th} expert for k=1, 2.. k.

b)   Arithmetic mean is used to aggregate experts' opinion as given in eq. (2).

$\tilde{C} = \frac{1}{k}\left(\frac{1}{c} + \frac{2}{c} + \cdots + \frac{k}{c}\right)$ (2)

- Step 4: Transform the linguistic terms into triangular fuzzy numbers. The following linguistic terms provided in Table 2 are utilized for the evaluation procedure.
- Step 5: Calculate the fuzzy weight matrix using eq. (3) and eq. (4).

$$\tilde{r}_i = (\tilde{c}_{i1} \otimes \tilde{c}_{i2} \otimes ... \otimes \tilde{c}_{in})^{\frac{1}{n}} \tag{3}$$
$$\tilde{w}_i = \tilde{r}_i \otimes (\tilde{r}_1 + \tilde{r}_2 + \cdots + \tilde{r}_n)^{-1} \tag{4}$$

where $\tilde{r}_i$ is the geometric mean of fuzzy comparison value and $\tilde{w}_i$ is the fuzzy weight of the ith criteria.

- Step 6: Apply normalization procedure as eq. (5)

$$w_i = \frac{\tilde{w}_i}{\sum_{j=1}^{n} \tilde{w}_j} \tag{5}$$

4 Illustration of FMCDM

In this section, we will illustrate the FMCDM using a health patient records dataset with the help of Fuzzy AHP technique. The reason for selecting this dataset is that it contains the typical both benefits and risks. The variety of benefits from the selected dataset, include the data availability of the hospital medical records by providing accurate, up-to-date, and enable quick access by the users to the patient records. However, from the side of the risks, by releasing the patient health records attributes, it might also encounter endangers like the name_of_patient, date_of_birth, and place_of_birth that result in the identification of individuals in a privacy violation.

4.1 Data Source: Health Patient Records Dataset

In the scenario of the illustration part, we designed that the government proposes a Department of Health to release a dataset of medical records of patient to the public that can enable individual or organization to access and see the current trend of a disease [38, 39]. By doing so, for instance, the government is able to generate a location map related to the disease landscape for some regions or specific attributes. However, if the government decides to open the dataset and actual values immediately, there are some potential privacy issues of the patients containing in the dataset that might be very harmful like misuse, inaccuracy, and identifiable of the data [39-41]. Figure 3 shows the dataset structure of the health patient records that will be analyzed using FMCDM in this study.
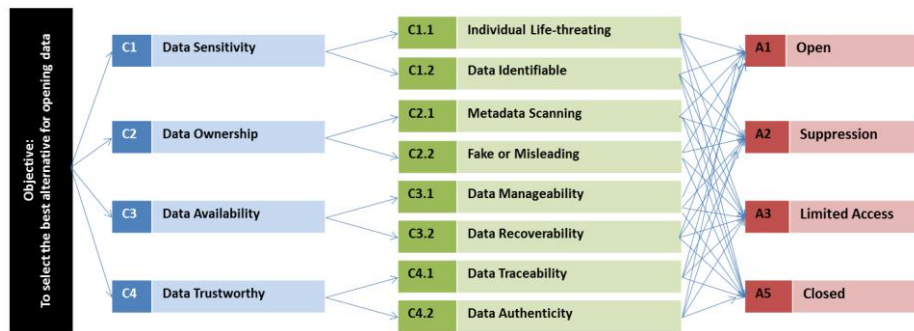
Fig. 3. Raw table of Health Patient Records (adapted from: [31, 42])

For the illustration of this work, we designed to analyze the Table 1 namely Diagnosed Stage which is containing six attributes/fields: Name_of_patient, Date_of_birth, Place_of_Birth, Gender, Race, Insurance, Stage, and TNM_staging.

4.2   Evaluation: Analyzing the Dataset

The following steps are the scenarios of FMCDM. Figure 4 shows the hierarchy of criteria and alternatives are used in the illustration of FMCDM.

- Step 1. Establish an expert team. We picked up the knowledge as well as expertise from some experts. The selected experts were interviewed based on the three consideration rationales, namely: (1) Domain knowledge, where the importance of educational background of the experts in this field ought to accommodate various specializations with partial overlap to confirm completeness of the data and available information [43, 44]; (2) Functional knowledge, where the experts chose are capable in the scope of the existing problems and the requirements of the process as well as solution proposed [43, 45]; and (3) Best practice, where the interviewee's expertise and their own insight have to be outstanding to warrant the quality as well as the validity of information sources [44].
- Step 2. Determine the evaluation criteria and construct the hierarchy including alternatives.

Fig. 4. Hierarchy of criteria and alternatives for the illustration

- Step 3. Construct pairwise comparison matrix and evaluate the relative importance of criteria. The experts are asked to provide their consideration based on their knowledge and expertise. For simplicity, in this illustration a pairwise comparison matrix for expert one is given in Figure 5. Before the experts started to quantify the criteria, we expected to construct a Fuzzy evaluation linguistic scale for the weights as presented in Table 2.

Table 2. The Fuzzy linguistic scales (adapted from: [18])

Fuzzy Number	Linguistic Scales	Scale of Fuzzy Number
1	Equal Important (EI)	(1,1,3)
3	Weakly Important (WI)	(1,3,5)
5	Essentially Important (SI)	(3,5,7)
7	Very Strongly Important (VI)	(5,7,9)
9	Absolutely Important (AI)	(7,9,9)

- Step 4: Transform the linguistic terms into triangular fuzzy numbers. The linguistic terms provided in Table 2 are utilized for the evaluation procedure.
- Step 5: Calculate the fuzzy weight matrix using Eq. (3) and Eq. (4). The final weights of the alternatives are calculated using Eq. (3), (4), and (5). The linguistic terms provided in Table 2 are utilized for the evaluation and fuzzy operational laws are used for the calculation[18, 27]. Illustrative examples for weights of sub-criteria C11 and C12 are given as follows:

Calculating sub-criteria: Linguistic terms for the pairwise comparison, we are getting from Figure 5 and the corresponding fuzzy numbers are getting from the Table 2. For example, pairwise comparison of (C1.1 C1.2) is "Equal Important" and the fuzzy number of this linguistic term is (1,1,3).

$$\tilde{r}_{c11} = (\tilde{c}_{c11c11} \otimes \tilde{c}_{c11c12})^{\frac{1}{2}}$$
$$\tilde{r}_{c11} = ((1,1,1) \otimes (3,5,7))^{\frac{1}{2}}$$
$$\tilde{r}_{c11} = (1.73, 2.23, 2.64)$$
$$\tilde{r}_{c12} = (\tilde{c}_{c12c11} \otimes \tilde{c}_{c12c12})^{\frac{1}{2}}$$
$$\tilde{r}_{c12} = ((1/(3,5,7)) \otimes (1,1,1))^{\frac{1}{2}}$$
$$\tilde{r}_{c12} = (0.37, 0.44, 0.57)$$

Calculating weights: For calculating weights, we are using eq. 4. In the previous step, we are getting the value of $\tilde{r}_{c1.1}$ and $\tilde{r}_{c1.2}$ and putting these values in the following equation.

$$\widetilde{w}_{c1.1} = (0.36, 0.5, 1.10)$$
$$\widetilde{w}_{c1.2} = \tilde{r}_{c1.2} \otimes (\tilde{r}_{c1.1} + \tilde{r}_{c1.2})^{-1}$$
$$\widetilde{w}_{c1.2} = (0.57, 1, 1) \otimes [(1,1,1.73) + (0.57, 1, 1)]^{-1}$$
$$\widetilde{w}_{c1.2} = (0.2, 0.5, 0.63)$$

Criteria	C1	C2	C3	C4
C1	1	SI	WI	WI
C2	1 / SI	1	WI	EI
C3	1 / WI	1 / WI	1	EI
C4	1 / WI	1/ EI	1 / EI	1

C1	C11	C12
C11	1	SI
C12	1/ SI	1

C2	C21	C22
C21	1	EI
C22	1/ EI	1

C3	C31	C32
C31	1	SI
C32	1 / SI	1

C4	C41	C42
C41	1	VI
C42	1 / WI	1

C11	A1	A2	A3	A4
A1	1	WI	WI	WI
A2	1 / WI	1	WI	WI
A3	1 / WI	1 / WI	1	WI
A4	1/ WI	1 / WI	1 / WI	1

C12	A1	A2	A3	A4
A1	1	WI	WI	WI
A2	1 / WI	1	WI	WI
A3	1 / WI	1 / WI	1	WI
A4	1 / WI	1 / WI	1 / WI	1

C21	A1	A2	A3	A4
A1	1	WI	WI	WI
A2	1 / WI	1	SI	SI
A3	1 / WI	1 / SI	1	WI
A4	1 / WI	1 / SI	1 / WI	1

C22	A1	A2	A3	A4
A1	1	WI	WI	WI
A2	1 / WI	1	SI	SI
A3	1 / WI	1 / SI	1	WI
A4	1 / WI	1 / SI	1 / WI	1

C31	A1	A2	A3	A4
A1	1	SI	SI	VI
A2	1 / SI	1	EI	SI
A3	1 / SI	1 / EI	1	SI
A4	1 / VI	1 / SI	1 / SI	1

C32	A1	A2	A3	A4
A1	1	SI	SI	VI
A2	1 / SI	1	EI	VI
A3	1 / SI	1 / EI	1	VI
A4	1 / VI	1 / VI	1 / VI	1

C41	A1	A2	A3	A4
A1	1	SI	SI	VI
A2	1 / SI	1	EI	SI
A3	1 / SI	1 / EI	1	VI
A4	1 / VI	1 / SI	1 / VI	1

C42	A1	A2	A3	A4
A1	1	SI	VI	VI
A2	1 / SI	1	EI	VI
A3	1 / VI	1 / EI	1	VI
A4	1 / VI	1 / VI	1 / VI	1

Fig. 5. The pairwise comparison matrices of criteria and alternatives

- Step 6: Apply normalization procedure.

Normalized weight values: To find the normalized weights of C1.1 and C1.2 we used eq. 5.

$$w_{c1.1} = \frac{\widetilde{w}_{c1.1}}{\sum_{j=1}^{2} \widetilde{w}_{1j}} = \frac{L_{c1.1} + M_{c1.1} + U_{c1.1}}{\widetilde{w}_{c1.1} + \widetilde{w}_{c1.2}}$$

$$w_{c1.1} = \frac{(0.36 + 0.5 + 1.10)}{(0.36 + 0.5 + 1.10 + 0.2 + 0.5 + 0.63)} = 0.59$$

$$w_{c1.2} = \frac{\widetilde{w}_{c1.2}}{\sum_{j=1}^{2} \widetilde{w}_{1j}} = \frac{L_{c1.2} + M_{c1.2} + U_{c1.2}}{\widetilde{w}_{c1.1} + \widetilde{w}_{c1.2}}$$

$$w_{c12} = \frac{(0.2 + 0.5 + 0.63)}{(0.36 + 0.5 + 1.10 + 0.2 + 0.5 + 0.63)} = 0.40$$

The similar calculation approach is applied for all pairwise comparisons. The final weights of the alternatives are provided in Table 3. An illustrative example for W_{A1} is given as follows:

$$W_{A1} = C1 \times C11 \times A1 + C1 \times C12 \times A1 + \cdots + C4 \times C41 \times A1 + C4 \times C42 \\ \times A1$$

$$W_{A1} = 0.53 \times 0.59 \times 0.39 + 0.53 \times 0.40 \times 0.41 + \cdots + 0.07 \times 0.59 \times 0.44 \\ + 0.07 \times 0.40 \times 0.35$$

$$W_{A1} = 0.34$$

Table 3. Final weights of the criteria and alternatives

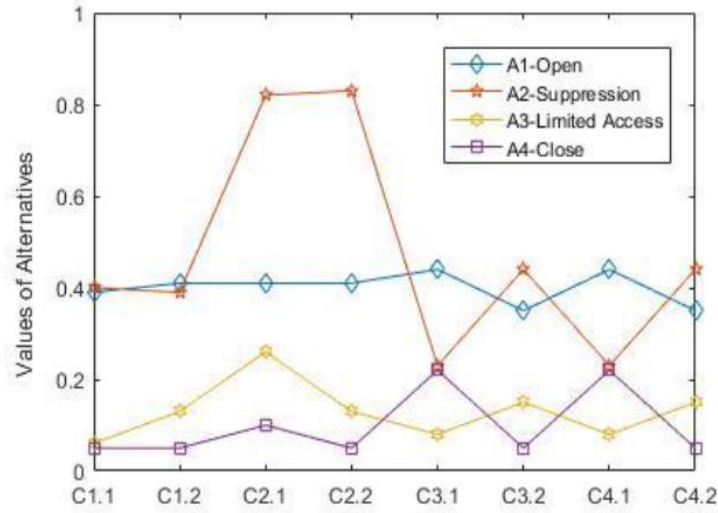|	C1		C2		C3		C4		
	0.53		0.25		0.13		0.07		
	C1.1	C1.2	C2.1	C2.2	C.31	C3.2	C4.1	C4.2	
	0.59	0.40	0.82	0.17	0.59	0.40	0.59	0.40	Weight
A1	0.39	0.41	0.41	0.41	0.44	0.35	0.44	0.35	0.34
A2	0.40	0.39	0.82	0.83	0.23	0.44	0.23	0.44	0.43
A3	0.06	0.13	0.26	0.13	0.08	0.15	0.08	0.15	0.08
A4	0.05	0.05	0.10	0.05	0.22	0.05	0.22	0.05	0.06

4.3 Decision: Recommendations

According to the Table 3, the highest priority of the decision for the Table 1 (Diagnose Stage) of Health Patient Records is A2 (0.42), following by A1 (0.34), and A3 (0.08), while in the last ranking of decision recommendation is A4 (0.06). Based on the analyzing and computing process, in this case we recommended that the Table 1 (Diagnose Stage) should be maintaining suppression as the highest priority recommendation in this illustration.

5 Findings

In order to present the recommendations based on the final results of the analyzing process using FMCDM, we designed a graphical view to support the decision-makers to decide to release their dataset. Figure 6 shows how the Fuzzy AHP could help the

decision-makers with the better understanding of the comparison score for each alternative.

Fig. 6. Ranking of Decision recommendations

Furthermore, to design the action plan of the maintaining suppression, some possible procedures could be taken into account as follows: (1) removing a data field or an individual attributes into particular group of the data and replace it into unique characteristics; (2) obscuring a data field by making substitution precise data values with ranges to minimize the provision of the personal identity; and (3) Aggregating a data field by summarizing the data across the amounts of the data and visualizing the data value into statistics form like graphics or charts.

6 Conclusion

In this paper, we presented the results of a study by utilizing Fuzzy AHP to analyze the risks and benefits of opening data for determining the best alternative in the health patient records dataset. A set of criteria and a variety of sub-criteria were designed and identified base on the literature review and experts' judgment. Some advantages the use of the FMCDM approach compare to other three methods as follows: (1) the capability to transform the human subjectivity and incorrectness from the common natural language to weights the complex problems, and (2) provides assessment method of the selected alternatives to rank a single decision making. However, a disadvantage found while using this approach is because the fuzzy is a ruled-based system, hence it needs to get enough rules to be accurate and expressively. The contribution resulted from this paper is to provide a decision-making model to analyze the potential risks and benefits of opening data. A given dataset is evaluated by taking action like measuring and weighing the relative importance of the multiple criteria.

Thus, the approach might contribute decision makers to decide to open a dataset. In the further research, we recommend refining this approach by adding more datasets in which and advice for (not) opening data can be generated without human involvement.

References

1. Ali-Eldin, A.M.T., A. Zuiderwijk, and M. Janssen. *Opening More Data: A New Privacy Scoring Model of Open Data*. in *Seventh International Symposium on Business Modelling and Software Design (BMSD 2017)*. 2017. Barcelona, Spain: SCITEPRESS - Science and Technology Publication, Lda.
2. Lourenço, R.P., *An analysis of open government portals: A perspective of transparency for accountability.* Government Information Quarterly, 2015. **32**(3): p. 323-332.
3. Zuiderwijk, A. and M. Janssen, *Open Data Policies, Their Implementation and Impact: A Framework for Comparison.* Government Information Quarterly, 2013. **31**(1).
4. Luthfi, A., M. Janssen, and J. Crompvoets. *A Causal Explanatory Model of Bayesian-belief Networks for Analysing the Risks of Opening Data*. in *8th International Symposium, BMSD 2018*. 2018. Vienna, Austria: Springer International Publishing AG.
5. Ubaldi, B., *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives.* OECD Working Papers on Public Governance, 2013. **22**: p. 60.
6. Janssen, M., Y. Charalabidis, and A. Zuiderwijk, *Benefits, Adoption Barriers and Myths of Open Data and Open Government.* Information System Management, 2012. **29**(4): p. 258-268.
7. Grimmelikhujsen, S.G. and A.J. Meijer, *Effects of Transparency on the Perceived Trustworthiness of a Government Organization: Evidence from an Online Experiment.* Journal of Public Administration Research and Theory, 2014. **24**(1): p. 137-157.
8. Kulk, S. and B.V. Loenen, *Brave New Open Data World?* International Journal of Spatial Data Infrasructure Research, 2012. **7**: p. 196-206.
9. Meijer, A. and M. Thaens, *Public information strategies: Making government information available to citizens.* Information Polity, 2009. **14**(1-2): p. 31-45.
10. Zuiderwijk, A. and M. Janssen, *Towards decision support for disclosing data: Closed or open data?* Information Polity, 2015. **20**(2-3): p. 103-107.
11. Barry, E. and F. Bannister, *Barriers to Open Data Release: A View from the Top.* Information Polity, 2014. **19**(1-2): p. 129-152.
12. Martin, S., et al., *Risk Analysis to Overcome Barriers to Open Data.* Electronic Journal of e-Government 2013. **11**(1): p. 348-359.

13. Barnickel, N., et al., *Berlin Open Data Strategy*, in *Concept, pilot system and recommendations for action*. 2012, Organisational, legal and technical aspects of Open Data in Berlin.
14. Conradie, P. and S. Choenni, *On the barriers for local government releasing open data.* Government Information Quarterly, 2014: p. S10-S17.
15. Uhlir, P.F. *The Socioeconomic Effects of Public Sector Information on Digital Networks: Toward a Better Understanding of Different Access and Reuse Policies*. in *National Research Council*. 2009. Washington DC.
16. Walter, S., *Heterogeneous Database Integration in Biomedicine.* Journal of Biomedical Informatics, 2001. **34**(4): p. 285-298.
17. Zuiderwijk, A., M. Janssen, and C. David, *Innovation with open data: Essential elements of open data ecosystems.* Information Polity, 2014. **19**(2-3): p. 17-33.
18. Hsieh, T.-Y., S.-T. Lu, and G.-H. Tzeng, *Fuzzy MCDM Approach for Planning and Design Tenders Selection in Public Office Buildings.* International Journal of Project Management, 2004. **22**: p. 573-584.
19. Rezaei, P., et al., *Application of Fuzzy Multi-Criteria Decision Making Analysis for Evaluating and Selecting the Best Location for Construction of Underground Dam.* Acta Polytechnica Hungarica, 2013. **10**(7): p. 187-205.
20. Zadeh, L.A., *The Concept of Linguistic Variable and Its Application to Approximate Reasoning.* Information Sciences, 1975. **8**(3): p. 199-249.
21. Fuller, R., *Fuzzy Logic and Neural Nets in Intelligent Systems*, in *Information System Day*, C.C. Ed., Editor. 1999: Turku Centre for Computer Science. p. 74-94.
22. Gupta, M.M., *Fuzzy Logic and Neural Systems.* International Series in Intelligent Technologies, 1995: p. 225-244.
23. Werro, N., *Fuzzy Classification of Online Customers.* Fuzzy Management Methods, 2015.
24. Novák, K., *An Introduction to Fuzzy Logic Applications in Intelligent System.* Kluwer Academic, 1992: p. 185-200.
25. Kahraman, C., S.C. Onar, and B. Oztaysi, *Fuzzy Multicriteria Decision-Making: A Literature Review.* International Journal of Computational Intelligence System, 2015. **8**(4): p. 637-666.
26. Chen, S.-J. and C.-L. Hwang, *Fuzzy Multiple Attribute Decision Making*, in *Lecturer Notes in Economics and Mathematical Systems*. 1992.
27. Hancerliogullari, G., K.H. Oymen, and E. Koksalmis, *The use of multi-criteria decision making models in evaluating anesthesia method options in circumcision surgery.* BMC Medical Informatics and Decision Making, 2017. **17**(14).
28. Lin, C. and C.H. Twu, *Fuzzy MCDM for Evaluating Fashion Trend Alternatives.* International Journal of Clothing Science and Technology, 2012. **24**(2/3): p. 141-153.
29. Sloane, E., M. Liberatore, and R. Nydick, *Medical Decision Support Using the Analytic Hierarchy Process.* Journal of Health Informatics Management, 2011. **16**(4): p. 38-43.

16

30. Buda, A., et al., *Decision support framework for opening business data*, in *Department of Engineering Systems and Services*. 2015, Delft University of Technology: Delft.

31. Luthfi, A. and M. Janssen, *A Conceptual Model of Decision-making Support for Opening Data*, in *7th International Conference, E-Democracy 2017*. 2017, Springer CCIS 792: Athens, Greece. p. 95-105.

32. Mohsen, D., et al., *A Combined Fuzzy MCDM Approach for Identifiying the Suitable Lands for Urban Development: An Example from Bandar ABBS, Iran.* Journal of Urban and Environmental Engineering, 2014. **8**(1): p. 11-27.

33. Podvezko, V., *Application of AHP Technique.* Journal of Business Economics and Management, 2011. **10**(2): p. 181-189.

34. Saaty, T., *The Analytic Hierarchy Process Planning, Priority Setting, Resource Allocation.* New York: McGraw, 1980.

35. Kuo, M., S, G.S. Liang, and W.C. Huang, *Extension of the multicriteria analysis with pairwise comparison under a fuzzy environment.* Journal of Science Direct, 2006. **43**: p. 268-285.

36. Isselhardt, D.J. and J. Cappuci, *The Analytic Hierarchy Process in Medical Decision Making A Tutorial.* Med Decis Mak, 1989. **1**: p. 40-50.

37. Sehra, S.K., Y.S. Brar, and N. Kaur, *Multi Criteria Decision Making Approach for Selecting Effort Estimation Model.* International Journal of Computer Application, 2012. **39**(1): p. 10-17.

38. Kostkova, P., et al., *Who Owns the Data? Open Data for Healthcare.* Frontiers of Public Health, 2016. **4**(7): p. 1-6.

39. Bøttcher, S.G. and C. Dethlefsen. *Learning Bayesian Networks with R*. 2003. Vienna University of Technology: Dept. of Mathematical Science, Aalborg University Denmark.

40. Spooner, L.M. and K.A. Pesaturo, *Chapter 1: The Medical Record*. 2013.

41. Ozair, F.F., et al., *Ethical issues in electronic health records: A general overview.* Perspective in Clinical Research, 2015. **6**(2): p. 73-76.

42. Abernethy, A., *Real World Evidence: Opportunities and Challenges*. 2016, Flatiron Health.

43. Teicher, M., *Interviewing Subject Matter Experts*, in *International Cost Estimating and Analysis Association (ICEAA)*. 2015.

44. Herland, K., H. Hämmäinen, and P. Kekolahti, *Information Security Risks Assessment of Smartphones Using Bayesian Networks.* Journal of Cyber Security, 2016. **4**: p. 65-85.

45. Damian, D., et al. *The Role of Domain Knowledge and Cross-Functional Communication in Socio-Technical Coordination*. in *35th International Conference on Software Engineering (ICSE) 2013*. 2013. San Fransisco, CA: International Conference on Software Engineering.