A New Concordance Coefficients-Based Approach to Compare Improved FMECA Methods

Andrés A. Zúñiga, João F.P. Fernandes, Paulo J.C. Branco

***Abstract*—**This paper introduces the use of Cohen's kappa concordance coefficient to compare different methods for improving FMECA analysis. Typically, qualitative comparisons of resulting rankings and the balance between risk factors are employed in such comparisons. However, the application of Cohen's kappa in the FMECA context is limited, despite its broad use in medical and social sciences. The proposed approach assesses the agreement between various methodologies (Risk Isosurface function, VIKOR, ITWH, Type-I and Type-II Fuzzy Inference System) when applied to the same problem, using an FMECA ranking as the reference. A blood transfusion case study with eleven widely-used failure modes is analyzed. The results demonstrate that Type-II fuzzy inference systems achieve the highest agreement with the reference ranking, possibly due to their inclusion of uncertainty as an additional parameter. This statistical approach effectively compares different FMECA methods, replacing the traditional qualitative comparison between rankings.

***Keywords*—** FMECA, Risk assessment, Type-II fuzzy inference systems, Concordance measurement, Cohen’s kappa.

I. Introduction

Failure Modes, Effects and Criticality Analysis is a qualitative risk assessment method designed to identify potential failure modes, their causes, and systems performance effects [1]. The objective of FMECA is to identify the possible ways a failure can occur, how often it occurs, how severe the failure affects the system performance, and what should be the preventive measures to avoid the failure.

The classical FMECA analysis is based on three factors, called risk factors, to characterize each failure mode [1]: the Severity (SEV) that characterize qualitatively the effect of the failure mode, the Frequency of Occurrence (OCC) that characterize how likely is it the failure mode to occur, and the Detectability (DET) that characterize how detectable is the failure mode before to occur. Each risk factor is classified in specific risk categories represented by a numerical scale, it can be a 1 to 10 scale as used in [1], or a 1 to 5 scale as in [2].

Each failure mode is assessed through a risk priority number (RPN); in general terms, the RPN results from the composition between SEV, OCC and DET as in (1), being the product the generally adopted approach.

$$RPN = (SEV) \circ (OCC) \circ (DET) \tag{1}$$

Although FMECA is a very popular qualitative method for failure analysis, computation of the RPN has some disadvantages [3]–[5]. They are:

1) The RPN computation does not consider any difference degree between the three risk factors OCC, SEV, and DET (i.e., no weight averaging each risk factor).
2) Although a higher RPN is usually associated with a more critical failure modes, this is not always true [6], [7], and.
3) The scales for the three risk factors are generally considered arbitrarily and may not accurately represent the risk characteristics in specific problems.

To deal with the FMECA shortcomings, in the past years were proposed some approaches based on computational intelligence and decision-making methods. In [3], the authors present one of the firsts applications of Fuzzy Inference Systems (FIS) to improve the FMECA analysis; results shown that proposed FIS allows to overcome some FMECA issues like imprecise information related to the risk factors. In [6], the authors conducted a literature review about FMECA methods published between 1998 and 2018; the review shows that grey theory and FIS are the most used methods in the last decades.

In [7] it is shown the application of Multi-Criteria Decision-Making (MCDM) methods and uncertainty theory to model the vagueness related to FMECA processes; this book includes a broad review of academic works. In [8] authors present a combination of fuzzy rules base and grey relation theory to improve the FMECA analysis conducted for an ocean going fish vessel; the proposed method includes the use of linguistic terms and allows to assign weights to each risk factor. In [4] the authors proposed the application of interval 2-tuple hybrid weighted distance (ITHWD) in an FMECA analysis conducted in a blood transfusion problem; the results proved that proposed approach is an useful way to prioritizing the failure modes in the presence of uncertainty and incomplete information.

Reference [9] shows the application of type-2 fuzzy-based FMECA in the risk assessment of manufacturing facility in the automotive industry. When consider triangular membership

Andrés A. Zúñiga, João F.P. Fernandes, Paulo J.C. Branco, IDMEC/Instituto Superior Técnico, University of Lisbon, Portugal

functions to represent the risk factors, the suggested approach offers additional flexibility to the experts in making judgments and provides better modeling of uncertainty.

Anes et al. [10] show an FMECA approach based on two mathematical functions: the first one deals with cases where the order of importance of risk variables is sufficient to prioritize failure modes; the second functions are an extension of the first one and allow taking into account the relative weight of each variable; when applied to the blood transfusion problem proposed in [4], the results evidence this method as promissory to improve the failure modes prioritization.

Commonly, the efficacy of new FMECA approaches is evaluated qualitatively by comparing one-to-one the rankings obtained. When the number of failure modes is small this approach can be adequate, otherwise, for a high number of failure modes this qualitative analysis becomes unpractical. The results of FMECA can be considered as a single ordinal ranking and, therefore, the concordance measurement is a suitable approach to improve the qualitative FMECA comparison.

The measure of concordance is a well-known problem in biological and social sciences.

The application of concordance measurements in the FMECA context is still limited. In [11] the author includes the application of Kendall's coefficient to determine the agreement between human experts in medical risk analysis context. In [12] authors show the application of FMECA's web-based three-round Delphi technique for the risk assessment related to transition from paper to digital based record in radiotherapy department; the authors propose the Kendall's coefficient to establish the consensus between the FMECA's risk factor. In both papers, the concordance was measured between human experts conducting the classical FMECA's and not between FMECA methods.

II. PRELIMINARIES

A. The measure of rank agreement

Let a collection of n objects classified by a particular characteristic, and let m a finite number of judges or evaluators who rank the n objects according to their appreciation of the objects' characteristics. It is important to know the degree of agreement between the evaluators' decisions. This problem was defined by Kendall and Smith as known *the problem of m ranking* [13].

Agreement, also known as concordance, reproducibility [14], or interrater reliability [15], is a concept closely related to, but fundamentally different from correlation [14]–[17]. The existence of agreement implies correlation, but the reciprocal may not be true [18]. The agreement focuses on the degree of concordance in the opinion between individuals regarding the same attribute or characteristic [14]; in contrast, correlation is usually applied to represent the association between two or more variables that do not necessarily measure the same attribute.

This paper considers the application of Cohen's Kappa to measure the concordance between pairs of raters.

B. The measure of rank agreement

Cohen's coefficient, usually known as Cohen's kappa and denoted by κ, is a statistic useful for inter-rater or intra-rater reliability measures [19], [20]. Cohen's kappa compares the proportion of objects in which the raters agreed and the proportion of objects for which disagreement is expected [19]. Originally, the coefficient κ was proposed to measure the agreement between two raters but it can be extended for more than two [20].

Let N objects, $n = 1, 2, \cdots, N$, classified independently into k categories by two separated and independent raters, called A and B, as shown in Table I. Here, as an example, Object 1 was rated as Category 5 by Rater A and Category 3 by Rater B. The categories can represent an intrinsic characteristic of the classified objects or a single ordinal ranking from 1 to k.

TABLE I
EXAMPLE OF *N* OBJECTS RANKED BY TWO RATERS

Objects	Rater A	Rater B
Object 1	Category 5	Category 3
Object 2	Category 2	Category k
⋮	⋮	⋮
Object n	Category k	Category 5
⋮	⋮	⋮
Object N	Category 1	Category 1

Let p_{ij} be the proportion of objects that rater A classified in the category i, $i = 1, 2, \cdots, k$, and rater B classified in the category j, $j = 1, 2, \cdots, k$, respectively. Table II shows the proportion of classified objects.

TABLE II
THE PROPORTION OF CLASSIFIED OBJECTS FOR EACH CATEGORY

		Rater B						
	Categories	1	2	…	j	…	k	Total
	1	p_{11}	p_{12}	…	p_{1j}	…	p_{1k}	p_{1+}
	2	p_{21}	p_{22}	…	p_{2j}	…	p_{2k}	p_{2+}
	⋮	⋮	⋮		⋮		⋮	⋮
Rater A	i	p_{i1}	p_{i2}	…	p_{ij}	…	p_{ik}	p_{i+}
	⋮	⋮	⋮		⋮		⋮	⋮
	k	p_{k1}	p_{k2}	…	p_{kj}	…	p_{kk}	p_{k+}
	Total	p_{+1}	p_{+2}	…	p_{+j}	…	p_{+k}	1

The proportions p_{i+} and p_{+j}, where the symbol + represents summation over the index, are the frequencies or marginal probabilities for an object to be assigned into category i for rater A and category j for rater B:

$$p_{i+} = \sum_{j=1}^{k} p_{ij} \tag{2}$$

$$p_{+j} = \sum_{i=1}^{k} p_{ij} \tag{3}$$

Where $\sum_{i=1}^{k} p_{i+} = 1$ and $\sum_{j=1}^{k} p_{+j} = 1$. Let p_0 be the observed proportion of agreement between raters [19] and expressed by (4):

$$p_0 = \sum_{i=1}^{k} p_{ii} \tag{4}$$

The observed proportion of agreement does not take into account the agreement obtained only by chance (this means not really "agreeing" at all) [21]. Therefore, the expected proportion of agreement obtained by chance, denoted by p_e, is based on the probability that rater A assigns the objects in the category i overall and rater B assigns the objects in the same category overall, that is for all $i = j$:

$$p_e = \sum_{i=1}^{k} \left(p_{i+} \cdot p_{+i} \right) \tag{5}$$

Then, Cohen's κ coefficient can be defined as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{6}$$

The lower and upper limits for κ are -1 and 1, respectively, but usually falls between 0 and 1 [21]. When the observed agreement is greater than the agreement expected by chance, κ takes positive values. When the observed agreement is less than the agreement expected by chance, κ takes negative values [19]. $\kappa = 1$ occurs when (and only when) there is a perfect agreement between raters. $\kappa = 0$ indicates that the observed agreement is no better than that expected by chance as if the raters had simply *guessed* every rating [21].

The value of κ can be interpreted using labels assigned for different ranges, as proposed in [16] and shown in Table III.

In some circumstances, the κ coefficient produces unexpected results; this problem has been referred in literature as the *kappa paradoxes* [15]. These paradoxes are related the use of marginal probabilities to compute the proportion p_e. As indicated in book [15], the application of weights to the original κ coefficient overcomes the paradoxes.

TABLE III
LABELS TO INTERPRET κ IN TERMS OF THE STRENGTH OF AGREEMENT

κ Range	Strength of agreement
$\kappa < 0.00$	Poor agreement
$0.00 < \kappa \leq 0.20$	Slight agreement
$0.20 < \kappa \leq 0.40$	Fair agreement
$0.40 < \kappa \leq 0.60$	Moderate agreement
$0.60 < \kappa \leq 0.80$	Substantial agreement
$0.80 < \kappa \leq 1.00$	Almost perfect agreement

C. Cohen's weighted kappa

Let w_{ij} the weight for agreement assigned to the $i^{th} - j^{th}$ cell of Table II. The weighted kappa coefficient κ_w can be defined by (7) [22]:

$$\kappa_w = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}\, p_{ij} - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}\, p_{i+}\, p_{+j}}{1 - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}\, p_{i+}\, p_{+j}} \tag{7}$$

Considering (7), the unweighted kappa is a special case of weighted kappa where weights are equal to 1 [22], [23]. The weights can be assigned using any judgment procedure or consensus of a committee of experts [23].

This work considers the quadratic weighting scheme, defined by (8) as the squared difference between categories i and j [15]:

$$w_{ij}^{(2)} = 1 - \left(\frac{i-j}{n-1} \right)^2 \tag{8}$$

D. Cohen's weighted kappa test of significance

Let H_0 be the null hypothesis stated as *raters' agreement is no better than agreement expected by chance* and let H_1 be the alternative hypothesis stated as *raters' agreement is better than agreement expected by chance*. The probability distribution of κ_w can be approximated by the Normal distribution [24] and the estimated variance for the null hypothesis is (9) [25]:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} \left(p_{i+}\, p_{+j} \left[w_{ij} \left(\bar{w}_{i+} + \bar{w}_{+j} \right)^2 \right] \right) - p_e^2}{n\left(1 - p_e\right)^2} \tag{9}$$

Where $\bar{w}_{i+} = \sum_{j=1}^{k} w_{ij}\, p_{+j}$ represents the weighted average of the weights in the i^{th} row and $\bar{w}_{+j} = \sum_{i=1}^{k} w_{ij}\, p_{i+}$ represents the weighted average of the weights in the j^{th} column [25].

Assuming that $\kappa_w / \hat{\sigma}$ follows a normal distribution, it is possible to test the hypothesis of agreement expected by chance by reference to the standard normal distribution [22]. The test statistics is thus defined by (10):

$$z = \frac{\kappa_w}{\hat{\sigma}} \tag{10}$$

For a one-sided alternative, the null hypothesis H_0 is rejected if $|z| \geq z_\alpha$, where z_α is the value that leaves α in the upper tail of the standard normal distribution. This work considers the level of significance as $\alpha = 0.05$ $\left(z_\alpha = 1.645 \right)$ [26].

E. The measurement of rank agreement in the FMECA context

The Cohen's kappa was selected to compare different improved FMECA methods under the following assumptions:

1) The failure modes represent the *n* classified objects;
2) The FMECA methods represent the *m* independent raters;

3) The ordinal failure modes' ranking represents the *k* categories;
4) One FMECA methods was selected as the reference one.

III. FMECA Case Study

The selected FMECA case study corresponds to a risk assessment in the blood transfusion process analyzed using the classical FMECA in [27] and subsequently analyzed using fuzzy-based and MCDM-based FMECA approaches in [4], [7], [10], [28]. Eleven failure modes with RPN higher than 80 were selected for further analysis.

This case study was selected by the following reasons:

1) It was already used for benchmarking in some references;
2) Because it has only 11 failure modes, comparison between different FMECA methods becomes more intuitive.

Table IV shows the FMECA analysis for the case study including the ranking obtained using the classical RPN [27].

TABLE IV
FMECA Table for the Case Study

Failure mode	Failure mode	SEV	OCC	DET	RPN	RANK
FM1	Insufficient and/or incorrect clinical information on request form	7	6	3	126	5
FM2	Blood plasma abuse	6	6	5	180	4
FM3	Insufficient preoperative assessment of the blood product requirement	7	5	7	245	1
FM4	Blood group verification incomplete	7	5	3	105	8
FM5	Delivery of blood sample and/or request form delayed	5	3	6	90	9
FM6	Incorrect blood components issued	10	1	8	80	10
FM7	Quality checks not performed on blood products	8	2	5	80	10
FM8	Preparation time before infusion >30 min	8	6	5	240	2
FM9	Transfusion cannot be completed within the appropriate time	7	4	4	112	6
FM10	Blood transfusion reaction occurs during the transfusion process	8	4	7	224	3
FM11	Bags of blood products are improperly disposed of bags	7	4	4	112	6

The method denoted as RPI(SC4) [10] was selected as the reference ranking to be used to measure the concordance between it and other FMECA methods. The reasons to justify this selection is based on the conclusions shown in [10]:

1) The selected FMECA, RPI(SC4), does not require additional previous knowledge about the problem, and;
2) The failure modes prioritization agrees with the expectation made for the risk scenario.

IV. Fuzzy-based FMECA Methods

In addition to the FMECAs listed in

Table V, this paper includes the application of Type-I and Type-II Fuzzy Inference Systems (Type I-FIS and Type-II FIS). The following membership functions were considered for the Type-I and Type-II Fuzzy Inference System: triangular (trimf), trapezoidal (trapmf), gaussian (gaussmf), generalized bell (gbellmf).

TABLE V
Different Rankings for FMECA Improvement Methods

Failure mode	RPN Rank	Fuzzy VIKOR	ITHWD	RPI(SC_4)	RPI(SC_5)
FM1	5	4	4	4	5
FM2	4	7	6	5	7
FM3	1	2	1	2	4
FM4	8	8	10	7	9
FM5	9	11	11	11	11
FM6	10	1	3	6	3
FM7	10	6	9	9	6
FM8	2	5	5	1	1
FM9	6	10	7	8	8
FM10	3	3	2	3	2
FM11	6	9	8	10	10

We defined eight fuzzy configurations for the Type-I FIS, denoted follow as Type-I FIS, T1-FIS 01 to T-FIS 08 and as shown in Table VI.

TABLE VI
Configurations for the FMECA based on Type-I Fuzzy Inference System

Config	Symmetry	MFSEV	MFOCC	MFDET	MFRPN
T1-FIS 01	symm	trimf	trimf	trimf	trimf
T1-FIS 02	symm	trapmf	trapmf	trapmf	trapmf
T1-FIS 03	symm	gaussmf	gaussmf	gaussmf	gaussmf
T1-FIS 04	symm	gbellmf	gbellmf	gbellmf	gbellmf
T1-FIS 05	asymm	trimf	trimf	trimf	trimf
T1-FIS 06	asymm	trapmf	trapmf	trapmf	trapmf
T1-FIS 07	asymm	gaussmf	gaussmf	gaussmf	gaussmf
T1-FIS 08	asymm	gbellmf	gbellmf	gbellmf	gbellmf

The term *symm* means that the used membership functions were all symmetrical, *asymm* means that the used membership functions were all asymmetrical; the prefix MF represents *membership function* for severity (SEV), occurrence (OCC), detection (DET), and RPN number.

For Type-II FIS, an exhaustive combination of four different types of membership functions for the severity, occurrence, detection, and RPN, were considered. The total combination of these set of parameters results in 41472 Type-II FIS configurations, denoted as T2-FIS 01 to T2-FIS 41472.

V. Results

Table VII shows the quadratic weighted concordance coefficient κ_{w-quad}, the value of the test statistics *z*, the strength of agreement and the result of the hypothesis test for the FMECA methods RPI(SC5), Fuzzy VIKOR and ITHWD, when compared with the reference ranking RPI(SC4). As shown, quadratic weighted kappa takes values between 0.727 and 0.855, revealing the scenario RPI(SC5) as those one achieving better concordance of 0.855, which can be considered as an almost perfect agreement.

Using now coefficient κ_{w-quad} ,

Table VIII shows the results for the FMECA based on the eight Type1-FIS proposed configurations.

TABLE VII

QUADRATIC WEIGHTED KAPPA κ_{w-quad} BETWEEN REFERENCE RANKING RPI(SC4) AND RPI(SC5), VIKOR, AND ITHWD

	RPI(SC$_5$)	Fuzzy VIKOR	ITHWD
κ_{w-quad}	0.855	0.727	0.809
Strength of agreement	Perfect	Substantial	Substantial
z	2.834	2.412	2.683
H_0 test	Reject	Reject	Reject

Results show that best agreement coefficient equal to 0.8 corresponding to the configuration T1-FIS-08 (all membership functions type gbell and asymmetrical). Two Type1-FIS configurations (T1-FIS 03, T1-FIS 04) achieved the worst value for the agreement coefficient, 0.536, which can be considered as moderate agreement.

At last, the quadratic weighted kappa is used in the FMECA based the 41472 Type-2 FIS proposed configurations. Table IX shows the results achieved for the eight best scenarios. Concordance coefficient achieves its highest value, 0.973, which can be considered as an almost perfect concordance, with the null hypothesis H_0 rejected in all scenarios.

TABLE VIII

QUADRATIC WEIGHTED KAPPA κ_{w-quad} BETWEEN REFERENCE RANKING RPI(SC4) AND TYPE-I FUZZY INFERENCE SYSTEM

	T1-FIS 01	T1-FIS 02	T1-FIS 03	T1-FIS 04	T1-FIS 05	T1-FIS 06	T1-FIS 07	T1-FIS 08
κ_{w-quad}	0.70	0.60	0.536	0.536	0.764	0.682	0.736	0.80
Strength of agreement	Substantial	Substantial	Moderate	Moderate	Substantial	Substantial	Substantial	Substantial
z	2.322	1.990	1.799	1.779	2.533	2.261	2.442	2.653
H_0 test	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

TABLE IX

QUADRATIC WEIGHTED KAPPA κ_{w-quad} BETWEEN REFERENCE RANKING RPI(SC4) AND TYPE-II FUZZY INFERENCE SYSTEM

	T2-FIS 25361	T2-FIS 29969	T2-FIS 30033	T2-FIS 34580	T2-FIS 35089	T2-FIS 35153	T2-FIS 38673	T2-FIS 38675
κ_{w-quad}	0.973	0.973	0.973	0.973	0.973	0.973	0.973	0.973
Strength of agreement	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect
z	3.226	3.226	3.226	3.226	3.226	3.226	3.226	3.226
H_0 test	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

In the FMECA context, the relationship between categories is not always linear and is difficult to establish; this relationship should determine the weighting scheme that will be used for the calculation of κ_w, by this reason the quadratic weighting schemes was considered.

A more in-depth study is needed to quantify the influence of the weighting scheme on the Cohen's kappa. Table X shows the ranking for the reference FMECA RPI(Sc4), the RPI(Sc5), ITHWD, T1-FIS 05, FWGM 08, and T2-FIS 38675, and their corresponding κ_{w-quad} ; the rankings were ordered from highest to lowest kappa.

Because the FMECA case study has only a few failure modes, it is possible to identify the concordances and discordances between the five FMECA methods. The ranking for failure modes FM1, FM2, FM5, FM10 and FM11 are the same for the reference RPI(SC$_4$) and T2-FIS 38675; both models agree 5 times and disagree 6 times. Comparing the base case with RPI(SC5), the rankings agree 4 times and disagree 7 times. For ITHWD and T1-FIS 05, the rankings agree 3 times and disagree 8 times.

Notice, that the number of agreements and disagreements can indicate the level of concordance between two raters in a simple way, however, it does not provide an effective metric to measure it; the Cohen's coefficient deals with this issue and also gives a concordance level based on the coincidences between ratings and the agreement that occurs by chance.

TABLE X

DIFFERENT RANKINGS FOR FMECA IMPROVEMENT METHODS

Failure mode	RPI(SC$_4$)	T2-FIS 38675	RPI(SC$_5$)	ITHWD	T1-FIS 05
FM1	4	4	5	4	8
FM2	5	5	7	6	5
FM3	2	1	4	1	3
FM4	7	6	9	10	9
FM5	11	11	11	11	11
FM6	6	7	3	3	2
FM7	9	8	6	9	10
FM8	1	2	1	5	1
FM9	8	9	8	7	6
FM10	3	3	2	2	4
FM11	10	10	10	8	7
κ_{w-quad}	Reference	0.973	0.855	0.809	0.764

When compared with the reference raking RPI(Sc4), the approach T2-FIS 38675 has perfect agreement in 5 failure modes (FM1, FM2, FM5, FM10 and FM11), the approach RPI(Sc5) has perfect agreement in 4 failure modes (FM5, FM8, FM9 and FM11), and the approach ITHWD has perfect agreement in 3 failure modes (FM1, FM5 and FM7).

VI. Conclusion and Future Work

This paper introduces an approach based on the Cohen's kappa concordance coefficient to compare different methods used in the FMECA context. A simple and further analyzed case study was selected to conduct the comparisons. FMECA approaches based on Type-I Fuzzy and Type-II Fuzzy Inference System were developed and applied to the test case. From the results and its previous discussion, one pulls out four critical conclusions:

1) The comparison between different FMECA methods based on the qualitative comparison between ranking and balance between the three risk factors can be impractical for more extensive problems.
2) The proposed approach aims to contribute to the quantitative comparison between methods used to improve the prioritization of failure modes regarding a reference ranking.
3) The results shown that the Cohen's κ coefficient gives a quantitative level for the agreement between two different rankings in the FMECA analysis context.
4) The ranking based on Type-II Fuzzy Inference System's achieves the best agreement regarding the reference FMECA method. This occurs due to the uncertainty being considered now as an additional parameter in the fuzzy inference process.
5) The selection of the weighting scheme is another essential aspect to take into account in the proposed approach; since the relationship between categories in FMECA's risk factors is not linear, results show that quadratic weighting scheme allows obtaining a better strength of agreement.
6) The reference FMECA's ranking identification is a critical aspect for the success of the proposed approach.

References

[1] International Electrotechnical Comission, *IEC 60812:2006 - Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA)*. 2018.

[2] X. Li, M. He, and H. Wang, "Application of failure mode and effect analysis in managing catheter-related blood stream infection in intensive care unit," *Med. (United States)*, vol. 96, no. 51, 2017, doi: 10.1097/MD.0000000000009339. https://doi.org/10.1097/MD.0000000000009339

[3] J. B. Bowles and C. E. Peláez, "Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis," *Reliab. Eng. Syst. Saf.*, vol. 50, no. 2, pp. 203–213, 1995, doi: 10.1016/0951-8320(95)00068-D. https://doi.org/10.1016/0951-8320(95)00068-D

[4] H. C. Liu, J. X. You, and X. Y. You, "Evaluating the risk of healthcare failure modes using interval 2-tuple hybrid weighted distance measure," *Comput. Ind. Eng.*, vol. 78, pp. 249–258, 2014, doi: 10.1016/j.cie.2014.07.018. https://doi.org/10.1016/j.cie.2014.07.018

[5] A. A. Zúñiga, A. Baleia, J. Fernandes, and P. J. da Costa Branco, "Classical failure modes and effects analysis in the context of smart grid cyber-physical systems," *Energies*, vol. 13, no. 5, pp. 1–26, 2020, doi: 10.3390/en13051215. https://doi.org/10.3390/en13051215

[6] J. Huang, J. X. You, H. C. Liu, and M. S. Song, "Failure mode and effect analysis improvement: A systematic literature review and future research agenda," *Reliab. Eng. Syst. Saf.*, vol. 199, no. February, p. 106885, 2020, doi: 10.1016/j.ress.2020.106885. https://doi.org/10.1016/j.ress.2020.106885

[7] H. C. Liu, *FMEA using uncertainty theories and MCDM methods*. Singapore: Springer Science+Business Media Singapore, 2016. https://doi.org/10.1007/978-981-10-1466-6

[8] A. Pillay and J. Wang, "Modified failure mode and effects analysis using approximate reasoning," *Reliab. Eng. Syst. Saf.*, vol. 79, no. 1, pp. 69–85, 2003, doi: 10.1016/S0951-8320(02)00179-5. https://doi.org/10.1016/S0951-8320(02)00179-5

[9] E. Bozdag, U. Asan, A. Soyer, and S. Serdarasan, "Risk prioritization in failure mode and effects analysis using interval type-2 fuzzy sets," *Expert Syst. Appl.*, vol. 42, no. 8, pp. 4000–4015, 2015, doi: 10.1016/j.eswa.2015.01.015. https://doi.org/10.1016/j.eswa.2015.01.015

[10] V. Anes, E. Henriques, M. Freitas, and L. Reis, "A new risk prioritization model for failure mode and effects analysis," *Qual. Reliab. Eng. Int.*, vol. 34, no. 4, pp. 516–528, 2018, doi: 10.1002/qre.2269. https://doi.org/10.1002/qre.2269

[11] P. C. Okwesili Jr., "Risk Assessmnt Using Paired Comparison Expert Elicitation for Ranking of Compounding Outsourcing Facilities," The George Washington University, 2016. https://doi.org/10.1109/EMR.2016.2530646

[12] H. Frewen, E. Brown, M. Jenkins, and A. O'Donovan, "Failure mode and effects analysis in a paperless radiotherapy department," *J. Med. Imaging Radiat. Oncol.*, vol. 62, no. 5, pp. 707–715, 2018, doi: 10.1111/1754-9485.12762. https://doi.org/10.1111/1754-9485.12762

[13] M. G. Kendall and B. B. Smith, "The Problem of m Rankings," *Ann. Math. Stat.*, vol. 10, no. 3, pp. 275–287, 1939, doi: 10.1214/aoms/1177732186. https://doi.org/10.1214/aoms/1177732186

[14] J. Liu, W. Tang, G. Chen, Y. Lu, C. Feng, and X. M. Tu, "Correlation and agreement: overview and clarification of competing concepts and measures," *Shanghai Arch. Psychiatry*, vol. 28, no. 2, pp. 115–120, 2016, doi: 10.11919/j.issn.1002-0829.216045.

[15] K. L. Gwet, *Handbook of Inter-Rater Reliability*, 4th ed. Gaithersburg: Advanced Analytics, LLC, 2014.

[16] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, p. 159, 1977, doi: 10.2307/2529310. https://doi.org/10.2307/2529310

[17] T. R. Vetter and P. Schober, "Agreement analysis: What he said, she said versus you said," *Anesth. Analg.*, vol. 126, no. 6, pp. 2123–2128, 2018, doi: 10.1213/ANE.0000000000002924. https://doi.org/10.1213/ANE.0000000000002924

[18] J. Teles, "Concordance coefficients to measure the agreement among several sets of ranks," *J. Appl. Stat.*, vol. 39, no. 8, pp. 1749–1764, 2012, doi: 10.1080/02664763.2012.681460. https://doi.org/10.1080/02664763.2012.681460

[19] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104. https://doi.org/10.1177/001316446002000104

[20] M. L. McHugh, "Lessons in biostatistics interrater reliability: the kappa statistic," *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, 2012, [Online]. Available: https://hrcak.srce.hr/89395. https://doi.org/10.11613/BM.2012.031 https://doi.org/10.1093/ptj/85.3.257

[21] J. Sim and C. C. Wright, "The kappa statistic in reliability studies: Use, interpretation, and sample size requirements," *Phys. Ther.*, vol. 85, no. 3, pp. 257–268, 2005, doi: 10.1093/ptj/85.3.257. https://doi.org/10.1093/ptj/85.3.257

[22] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha, "Beyond kappa: A review of interrater agreement measures," *Can. J. Stat.*, vol. 27, no. 1, pp. 3–23, 1999, doi: 10.2307/3315487. https://doi.org/10.2307/3315487

[23] J. Cohen, "Weighted kappa: nominal scale agreement with provision for scaled disagreement of partial credit," *Psychol. Bull.*, vol. 70, no. 4, pp. 213–220, 1968, doi: https://doi.org/10.1037/h0026256.

https://doi.org/10.1037/h0026256

[24] A. Ben-David, "Comparison of classification accuracy using Cohen's Weighted Kappa," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 825–832, 2008, doi: 10.1016/j.eswa.2006.10.022.
https://doi.org/10.1016/j.eswa.2006.10.022

[25] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa," *Psychol. Bull.*, vol. 72, no. 5, pp. 323–327, 1969, doi: 10.1037/h0028106.
https://doi.org/10.1037/h0028106

[26] S. M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Sixth Edit. London: Academic Press, 2021.
https://doi.org/10.1016/B978-0-12-824346-6.00010-7

[27] Y. Lu, F. Teng, J. Zhou, A. Wen, and Y. Bi, "Failure mode and effect analysis in blood transfusion: A proactive tool to reduce risks," *Transfusion*, vol. 53, no. 12, pp. 3080–3087, 2013, doi: 10.1111/trf.12174.
https://doi.org/10.1111/trf.12174

[28] H.-C. Liu, *Improved FMEA Methods for Proactive Healthcare Risk Analysis*. Singapore: Springer, 2019.
https://doi.org/10.1007/978-981-13-6366-5

[29] H. C. Liu, L. Liu, N. Liu, and L. X. Mao, "Risk evaluation in failure mode and effects analysis with extended VIKOR method under fuzzy environment," *Expert Syst. Appl.*, vol. 39, no. 17, pp. 12926–12934, 2012, doi: 10.1016/j.eswa.2012.05.031.
https://doi.org/10.1016/j.eswa.2012.05.031