Received: 29 October 2020 | Revised: 1 February 2021 | Accepted: 17 February 2021

DOI: 10.1002/int.22407

RESEARCH ARTICLE

WILEY

Nonparametric fuzzy hypothesis testing for quantiles applied to clinical characteristics of COVID-19

Nataliya Chukhrova | **Arne Johannssen**

Faculty of Business Administration, University of Hamburg, Hamburg, Germany

Correspondence
Nataliya Chukhrova, Faculty of Business Administration, University of Hamburg, 20146 Hamburg, Germany.
Email: nataliya.chukhrova@uni-hamburg.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 409030527

Abstract

The sign test is one of the most popular nonparametric tests for location problems and allows testing for any quantile of a population. However, the common sign test has serious drawbacks such as loss of information by considering solely signs of observations but not their magnitudes, various problems related to handling of ties in the data, and the lack of embedding uncertainty regarding the fraction of underlying quantile. To address these issues, we present an extended sign test based on fuzzy categories and fuzzy formulated hypotheses that improves the generality, versatility, and practicability of the common sign test. This generalized test procedure is neat in theory and practice and avoids disadvantages that are often associated with fuzzy tests (e.g., a considerably higher complexity of the underlying model, a fuzzy test decision, and a possibilistic instead of a probabilistic interpretation of test results). In addition, we perform a comprehensive case study on COVID-19 in HIV-infected individuals with a focus on human body temperature and related measurement problems. The results of the study clearly indicate that fuzzy categories and fuzzy hypotheses improve the performance of the sign test.

KEYWORDS

fuzzy categories, fuzzy statistics, median test, sign test, statistical inference, temperature measurement

1 | INTRODUCTION

Statistical inference aims at deducing properties and learning characteristics of an underlying population with some form of sampling. This is done by selecting a statistical model for the data generating process which is based on a set of assumptions regarding data generation and observed data. In general, three types of model assumptions for statistical inference are distinguished: parametric, semiparametric, and nonparametric. Among these three types, parametric tests meet the highest requirements by demanding that the underlying data generating process is fully captured by a specific probability distribution defined on a finite number of unknown parameters. In contrast, nonparametric tests impose only minimal requirements, and semiparametric tests typically imply a number of assumptions lying between the parametric and the nonparametric case. It is crucial for valid statistical inference to make appropriate assumptions about the data generating process (see e.g., Grzegorzewski[1]), which is why specific assumptions have to be formulated very carefully and nonparametric test methods are preferable in case of any doubt about the underlying population.

There are some popular nonparametric tests such as sign test, Wilcoxon signed-rank test, Mann–Whitney U-test (Wilcoxon rank sum test), Kruskal–Wallis test, and Kolmogorov–Smirnov test, just to name a few. In this paper, we focus on improving the sign test for practical applications by embedding concepts of fuzzy statistics. The sign test is a test for location problems like any quantile of a population that provides a series of benefits for practical test problems. First, its usage is bounded to almost no assumptions about the population distribution, which is why the sign test provides a very general applicability. Since quantiles such as the median or quartiles are used for testing on location parameters, the sign test is based on a more robust measure of central tendency compared to common parametric tests, for example, for the mean. It is also applicable to paired-sample data like pre- and post-treatment observations for each of involved subjects. Moreover, the sign test is an exact (binomial) test, operates also with small sample sizes, and is intuitive and simple in application.

On the other hand, there are a few drawbacks of the common sign test:

1. It lacks statistical power due to utilization of information solely regarding the signs of the differences between observations and the hypothesized quantile (and ignoring e.g., the magnitude/order of these differences).
2. There are difficulties in the appropriate handling of null differences (ties), either via reduction of the sample size (elimination of ties) or classification to both categories (negative/positive signs) in an equal magnitude (i.e., 1 or 0.5) or classification to the underrepresented category.
3. It has a poor performance for small levels of significance in combination with very small sample sizes.

In addition, the assumptions of crisp data and crisp hypotheses (on which the sign test is also founded) are often inaccurate as most observations are more or less imprecise (see Shafiq et al.[2]) or there is uncertainty regarding a hypothesized quantile value. For these reasons, some authors have considered approaches of fuzzy statistics to deal appropriately with fuzziness in data and hypotheses formulation. In the literature of fuzzy hypothesis testing, there are a few publications considering the sign test in fuzzy environments (see Chukhrova and Johannssen,[3] for a systematic review):

- Fuzzy/interval-valued data caused by the imprecision of observations (see Grzegorzewski,[4,5] Grzegorzewski and Spiewak,[6-8] Hesamian and Chachi,[9] Hesamian and Taheri,[10] Kahraman

et al.,[11] Momeni and Sadeghpour-Gildeh,[12] and Shams and Hesamian[13]), that is, fuzzy data as perception of a crisp but unobservable random variable (so called epistemic perspective, see Couso and Dubois[14]), or fuzzy set-/interval-valued random variables (so called ontic perspective) (see Grzegorzewski and Spiewak[6,7]).
- Fuzzy/interval-valued hypotheses caused by fuzzy quantiles like the fuzzy median (see Grzegorzewski and Spiewak[6,7]) or imprecision of linguistic statements on quantiles (see Hesamian and Chachi,[9] Hesamian and Taheri,[10] Momeni and Sadeghpour-Gildeh,[12] and Shams and Hesamian[13]).

To the best of our knowledge, there is no sign test proposed in the literature to date where fuzzy instead of crisp categories are considered and/or where fuzziness of the hypotheses is caused by imprecision of linguistic statements on fractions of underlying quantiles. However, the embedding of fuzzy categories may help to overcome shortcomings 1.–2. of the classical sign test stated above on the one hand. On the other hand, fuzzy formulation of hypotheses regarding fractions would be more natural and simple for the practitioner/researcher compared to common formulation regarding quantiles, since the underlying test is of binomial type. In addition, the standardization of modeling membership functions for popular quantiles like the median is also possible, for example, for a better embedding into automated processes like big data knowledge extraction. Thus, the paper pursues a twofold goal: the sign test will be extended by fuzzy categories and fuzzy hypotheses, to improve its generality, versatility, and practicability.

To achieve the first goal, the crisp categories "negative signs (differences)" and "positive signs (differences)" can be reformulated to fuzzy ones like "rather negative signs" and "rather positive signs." Such classifications can be adequately described by fuzzy sets with monotonically decreasing and increasing membership functions, respectively. In this way, a suitable consideration of ties as well as of magnitude of border values regarding a hypothesized quantile value is feasible by means of corresponding membership degrees and can be easily implemented into the test statistic. In addition, there is also the possibility to consider the order of observations beyond the border area via full or no membership (1 vs. 0) to the respective category in compliance with the classical approach. Further, following Dixon and Mood,[15] the classification of ties to both (crisp or fuzzy) categories in an equal magnitude of 0.5 can be proposed. But contrary to the classical approach, where the real-valued test statistic is either rounded up or down without consideration of the impact of this transformation on the sample size, the implementation of the untransformed real-valued test statistic in the definition of the p-value can be suggested, in particular via using gamma functions instead of binomial coefficients.

It should be mentioned that an implementation of fuzzy categories in the framework of fuzzy hypothesis testing has already been proposed by De Garibay,[16] Grzegorzewski and Szymanowski,[17] Taheri and Hesamian,[18] and Wu and Chang.[19] They utilize real-valued coding to model memberships of crisp variables to fuzzy categories. The test statistic of the test under consideration (no sign test has been examined) is then calculated based on the membership degrees to specific categories of these crisp variables (i.e., $[0, 1]$ instead of $\{0, 1\}$).

As for achieving the second goal, the methodology proposed by Chukhrova and Johannssen[20] can be transferred and adapted to the sign test. In particular, this methodology provides the following benefits in the framework of a fuzzy sign test:

- consideration of both whole hypotheses as well as the indifference zone in the test procedure;
- specification of crisp and fuzzy areas in hypothesis formulation;

- generalization of the indifference zone by its gradual consideration;
- formulation of the generalized p-value using supremum of weighted conditional probabilities; and
- definition of weight function under consideration of membership functions of both hypotheses.

In addition, using the proposed methodology, the practitioner benefits from a crisp test decision, which corresponds to a classical test decision “reject/do not reject H_0,” supported by interpretations of the generalized p-value in the classical probabilistic sense. In contrast, the majority of fuzzy tests provides key measures of possibilistic nature (like fuzzy test statistic, fuzzy critical value, fuzzy p-value, fuzzy confidence interval or fuzzy critical region) and a fuzzy test decision “reject/do not reject H_0 to a particular degree.” Such statements often lead to indifference regarding the final test decision and can be overcome only by additional sampling or defuzzification of test results (which demands in turn specific defuzzification approaches). Last but not least, the proposed formulation of fuzzy hypotheses requires only specification of the indifference zone as well as of both (mostly complementary) membership functions. Such modeling avoids a considerably increased complexity of the underlying model that often comes along with fuzzy hypothesis testing, due to the necessity of implementing plenty of additional assumptions regarding formulation of fuzzy data, fuzzy random variables, and fuzzy parameters.

To emphasize the benefits of the proposed generalized sign test in practical applications, we present a comprehensive case study based on a real data set on COVID-19 in HIV-infected individuals. In particular, we analyze the human body temperature level and present methods for handling problems regarding its categorical classification and thus appropriate interpretation of observed results by utilizing fuzzy categories. The generalized sign test is then performed with crisp and fuzzy hypotheses, and the results are compared and interpreted.

The paper is organized as follows. Section 2 introduces some necessary preliminaries on fuzzy sets. In Section 3 and 4, the one-tailed test of significance for quantiles with crisp and fuzzy hypotheses is discussed, respectively. Section 5 proposes adequate ways of modeling fuzzy categories. In Section 6, we perform an extensive case study with regard to COVID-19 in HIV-infected individuals. Finally, in Section 7 the paper concludes with an overview of study results.

2 | PRELIMINARIES ON FUZZY SETS

Considering the classical (crisp) set theory, sets are defined as collections of elements $u \in \mathcal{U}$, where each u either belongs to or does not belong to a crisp set $A \subseteq \mathcal{U}$. Thus, a crisp set A is described by an indicator function $m_A : \mathcal{U} \to \{0, 1\}$ with

$$m_A(u) = \begin{cases} 1 & \text{if } u \in A \\ 0 & \text{if } u \notin A \end{cases}$$

While crisp sets allow only for differentiating between membership (1) and nonmembership (0) of single elements u to a set A, fuzzy sets enable various degrees of membership by generalizing indicator functions $m_A : \mathcal{U} \to \{0, 1\}$ to membership functions $m_A : \mathcal{U} \to [0, 1]$. A fuzzy set A in $\mathcal{U}$ is then given by a set of ordered pairs

$$A := \{(u;\, m_A(u)) | u \in \mathcal{U}\}$$

A fuzzy set A is referred to as

- normal, if there exists an $u \in \mathcal{U}$ such that $\text{hgt}(A) = \sup_{u \in \mathcal{U}} m_A(u) = 1$,
- subnormal, if $0 < \text{hgt}(A) < 1$ for all $u \in \mathcal{U}$,
- convex, if $m_A(\lambda u_1 + (1 - \lambda)u_2) \geq \min(m_A(u_1), m_A(u_2))$ for all $u_1, u_2 \in \mathcal{U}$ and $\lambda \in [0, 1]$,

where hgt denotes the height of a fuzzy set A.

As we are generally referring to a nonempty (crisp) universal set $\mathcal{U}$, there may be elements of A having the degree of membership zero (see e.g., Zimmermann[21]). However, elements with a nonzero degree of membership are mostly of primary interest. This leads us to the so called γ-cut of A, which is denoted by $A_\gamma = \{u \in \mathcal{U} | m_A(u) \geq \gamma\}$ and is defined as the (crisp) set of elements belonging at least up to the degree $\gamma \in (0, 1]$ to A. There are some common γ-cuts of A like the support (supp), also called the topological closure of A, and the core (ncl):

$$\begin{aligned} \text{supp}(A) = A_0 &= \{u \in \mathcal{U} | m_A(u) > 0\} \\ \text{ncl}(A) = A_1 &= \{u \in \mathcal{U} | m_A(u) = 1\} \end{aligned}$$

The (crisp) set of all fuzzy subsets of $\mathcal{U}$ is denoted by $\mathcal{F}(\mathcal{U})$.

Given two sets $A, B \in \mathcal{F}(\mathcal{U})$ with $m_A(u) \leq m_B(u)$ for all $u \in \mathcal{U}$, then A is a fuzzy subset of B ($A \subseteq B$). If there is at least one $u \in \mathcal{U}$ with $m_A(u) < m_B(u)$, then A is a proper fuzzy subset of B ($A \subset B$).

In addition, we define a fuzzy subset ${}_\gamma A$ of the fuzzy set A, whose support corresponds to the γ-cut of A, that is,

$${}_\gamma A := \{(u; \min\{m_{A_\gamma}(u), m_A(u)\}) | u \in \mathcal{U}\}$$

Since the membership function is the crucial part of fuzzy sets, operations with fuzzy sets are defined by means of their membership functions. In this paper, we make use of basic set-theoretic operations on fuzzy sets like complement, intersection and union defined as follows (see e.g., Chukhrova and Johannssen[22, 23]):

$$\begin{aligned} m_{\overline{A}}(u) &= 1 - m_A(u) && \text{(complement of a normal fuzzy set } A \in \mathcal{F}(U)) \\ m_{A \cap B}(u) &= \min\{m_A(u), m_B(u)\} && \text{(intersection of fuzzy sets } A, B \in \mathcal{F}(U)) \\ m_{A \cup B}(u) &= \max\{m_A(u), m_B(u)\} && \text{(union of fuzzy sets } A, B \in \mathcal{F}(U)) \end{aligned}$$

3 | ONE-TAILED TEST OF SIGNIFICANCE FOR QUANTILES WITH CRISP HYPOTHESES

In the following, we discuss the classical one-tailed sign test of significance, considering the category "negative signs" and "positive signs" for the formulation of the test statistic ($S^{(-)}$ and $S^{(+)}$) in Sections 3.1 and 3.2, respectively. This separation is necessary since we propose to use the p-value for "left-tailed" events for decision making in the framework of the right- and left-tailed sign test, based on the realization of the test statistic $S^{(-)}$ and $S^{(+)}$, respectively. This proposal simplifies the calculation and is in line with the classical approach.

3.1 | The case of the category "negative signs"

To make statements about a true unknown quantile k_q (such as median, quartiles, etc.), with $k_q \in \mathbb{R}$ and fraction $q \in (0, 1)$, of a population of interest (whose cumulative distribution function is assumed to be continuous and strictly increasing in vicinity of k_q), we propose a test procedure, which unifies the formulation of a respective test problem and its solution in the following four steps:

1. Formulation of the hypotheses;
2. Specification of the sample size and the significance level;
3. Random sampling and determination of the test statistic;
4. Test decision regarding H_0 on the basis of the p-value.

According to the test procedure described above, in the first step hypotheses $H_0^{'}$ and $H_1^{'}$ (with an empty indifference zone I' defined as a set between null and alternative hypothesis) have to be formulated for one-tailed test problems preliminary over the real numbers as complementary statements on k_q (e.g., the median) with a hypothesized value k (see Table 1, row 1–2, and Figure 1). As for noncomplementary hypotheses, this case is less relevant in the framework of a test of significance and is therefore not considered in this paper.

As well known, a quantile k_q is defined by the constraint

$$\mathbb{P}(X_i \leq k_q) = q, \tag{1}$$

TABLE 1 Subsets of the parameter space and their indicator functions (using category "negative signs")

	Preliminary left-tailed test	Preliminary right-tailed test
Preliminary hypothesis $H_0^{'}$	$k_q \geq k$	$k_q \leq k$
Preliminary hypothesis $H_1^{'}$	$k_q < k$	$k_q > k$
	Right-tailed test	Left-tailed test
Hypothesis H_0	$q \leq q_0$	$q \geq q_0$
Hypothesis H_1	$q > q_1$	$q < q_1$
With	$0 < q_0 = q_1 < 1$	$0 < q_1 = q_0 < 1$
Set H_0	$\{q \in \Theta \| q \leq q_0\}$	$\{q \in \Theta \| q \geq q_0\}$
Set H_1	$\{q \in \Theta \| q > q_1\}$	$\{q \in \Theta \| q < q_1\}$
Set I	$\{q \in \Theta \| q_0 < q \leq q_1\} = \varnothing$	$\{q \in \Theta \| q_1 \leq q < q_0\} = \varnothing$
$m_{H_0}(q) =$	$\begin{cases} 1 & \text{if } 0 < q \leq q_0 \\ 0 & \text{if } q_0 < q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < q < q_0 \\ 1 & \text{if } q_0 \leq q < 1 \end{cases}$
$m_{H_1}(q) =$	$\begin{cases} 0 & \text{if } 0 < q \leq q_1 \\ 1 & \text{if } q_1 < q < 1 \end{cases}$	$\begin{cases} 1 & \text{if } 0 < q < q_1 \\ 0 & \text{if } q_1 \leq q < 1 \end{cases}$
$m_I(q) =$	$\begin{cases} 0 & \text{if } 0 < q \leq q_0 \\ 1 & \text{if } q_0 < q \leq q_1 \\ 0 & \text{if } q_1 < q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < q < q_1 \\ 1 & \text{if } q_1 \leq q < q_0 \\ 0 & \text{if } q_0 \leq q < 1 \end{cases}$

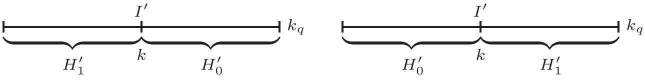

FIGURE 1 Representation of preliminary hypotheses $H_0^{'}, H_1^{'}$ for the left- and right-tailed case

where X_i is the random variable of interest. Thus, both preliminary hypotheses can be alternatively reformulated to complementary statements H_0 and H_1 on the fraction q with parameter space $\Theta := (0, 1)$ and a hypothesized fraction value, say q_0 (e.g., $q_0 = 0.5$ regarding a test for the median) (see Table 1, row 3–5, and Figure 2). Note that there is a transition from a left- to a right-tailed test problem and vice versa due to the reformulation of hypotheses in compliance with constraint (1). Further, it should be mentioned that in this paper the additional notation q_1 instead of q_0 for an edge element of H_1 is introduced for generalization reasons in accordance with the theory stated in the next section.

In terms of classical test theory the subsets of the parameter space Θ are defined here as crisp disjoint sets H_0, H_1, I $(H_0 \cup H_1 \cup I = \Theta)$ with corresponding indicator functions $m_{H_0}(q), m_{H_1}(q), m_I(q)$ (see Table 1, row 6–11). In particular, the sets H_0, H_1, I are specified by the hypotheses H_0 and H_1 as well as by the indifference zone I. The sets H_0 and H_1 are nonempty, however, the set I is empty.

Assuming for the moment that the true value of the quantile k_q of a population is known, then the values of the indicator functions $m_{H_0}(q)$ and $m_{H_1}(q)$ could be determined. Therefore, the null hypothesis would be rejected for $m_{H_0}(q) < m_{H_1}(q)$ (this applies to $q \in H_1$) and would not be rejected for $m_{H_0}(q) > m_{H_1}(q)$ (this applies to $q \in H_0$). In both cases these decisions are correct and consistent with reality.

In general, however, the true value of the fraction q is unknown. Consequently, the statistical decision regarding the null hypothesis is to make using a random sample of size n, $n \in \mathbb{N}$. Therefore, in the framework of a level-α-test (test of significance), which controls the type I error, the appropriate α-level, $\alpha \in (0, 1)$, is to specify in the second step. It should be mentioned that although the sample size is user-defined here, it should not be set arbitrarily low, otherwise cost savings are only possible at the expense of increasing probabilities for the type II error. These probabilities are the largest in the case of complementary hypotheses as considered above. On the other hand, a very low sample size can lead to test situations, where the null hypothesis can hardly be rejected. For instance, at the 5% level of significance, $n \geq 6$ is necessary before any conclusion can be drawn (see Dixon and Mood[15]).

In the third step, random sampling of size n is to conduct. In particular, the n random variables X_i, $i = 1, \ldots, n$, are stochastically independent and drawn from a continuous distribution, whose quantile k_q is uniquely defined by constraint (1). Considering exemplary a test for the median, that is, $\mathbb{P}(X_i \leq k_q) = 0.5$, under H_0 half and half of the sample observations should lie below and above the hypothesized value k on average, respectively. Further, a disjunctive coding by 0/1 of a continuous variable X_i provides its convenient treatment as a categorical variable with indicator functions $m_{C^{(-)}}(X_i) \in \{0, 1\}$ and $m_{C^{(+)}}(X_i) \in \{0, 1\}$ regarding

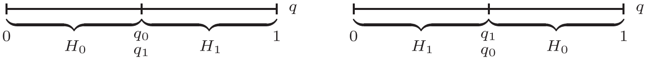

FIGURE 2 Representation of respective reformulated hypotheses H_0, H_1 for the right- and left-tailed case

both possible categories: "negative signs" ($C^{(-)}$) and "positive signs" ($C^{(+)}$), which are essential to construct the test statistic in the framework of the sign test.

Considering the membership of X_i to category $C^{(-)}$ as success, that is, $\mathbb{P}(m_{C^{(-)}}(X_i) = 1) = q$ and $\mathbb{P}(m_{C^{(-)}}(X_i) = 0) = 1 - q$, the random variable $S^{(-)} = \sum_{i=1}^{n} m_{C^{(-)}}(X_i)$ (in the following: the test statistic) is defined as the number of negative signs (i.e., the number of X_i, $i = 1, \ldots, n$, where $X_i < k$ holds) in the sample and follows the binomial distribution with probability mass function

$$\mathbb{P}(S^{(-)} = s^{(-)}) = \binom{n}{s^{(-)}} q^{s^{(-)}} (1-q)^{n-s^{(-)}}.$$

That is why the sign test actually corresponds to the exact binomial test with power function $G_{n,c^{(-)}}(q)$, where $c^{(-)}$ is the critical (rejection) value with $0 \leq c^{(-)} \leq n-1$, $c^{(-)} \in \mathbb{N}_0$ for the right-tailed case and $1 \leq c^{(-)} \leq n$, $c^{(-)} \in \mathbb{N}$ for the left-tailed case. The power function $G_{n,c^{(-)}}(q)$ quantifies probabilities for a correct test decision (rejection of H_0) for $q \in H_1$ and probabilities for a false test decision (rejection of H_0) for $q \in H_0$. Since the power function for a right-tailed test, $G_{n,c^{(-)}}(q) = \sum_{m=c^{(-)}+1}^{n} \binom{n}{m} q^m (1-q)^{n-m}$, is monotonically increasing, and the power function for a left-tailed test, $G_{n,c^{(-)}}(q) = \sum_{m=0}^{c^{(-)}-1} \binom{n}{m} q^m (1-q)^{n-m}$, is monotonically decreasing, the respective argument value of the type I error criterion (defined as the supremum of probabilities for false rejection of H_0) is equal to the edge element q_0 of the set H_0, that is,

$$E_1(n, c^{(-)}) = \sup_{q \in H_0} G_{n,c^{(-)}}(q) = G_{n,c^{(-)}}(q_0) \leq \alpha$$

As for a test decision, it is to achieve in the fourth step by comparing the α-level to the p-value. The p-value is defined as the conditional probability that the observed sampling results or more extreme (in terms of H_0) sampling results occur, when the null hypothesis is true. Applying this definition to the binomial test, we obtain

$$\begin{aligned}
p_r(n, s^{(-)}) = \; & \mathbb{P}_{q \in H_0}(S^{(-)} \geq s^{(-)}) = \sum_{m=s^{(-)}}^{n} \binom{n}{m} q^m (1-q)^{n-m} \\
= \; & \sum_{m=s^{(-)}}^{n} \binom{n}{m} q_0^m (1-q_0)^{n-m} \qquad \text{(right-tailed test)} \\
p_l(n, s^{(-)}) = \; & \mathbb{P}_{q \in H_0}(S^{(-)} \leq s^{(-)}) = \sum_{m=0}^{s^{(-)}} \binom{n}{m} q^m (1-q)^{n-m} \\
= \; & \sum_{m=0}^{s^{(-)}} \binom{n}{m} q_0^m (1-q_0)^{n-m} \qquad \text{(left-tailed test)}
\end{aligned}$$

If the quantile of interest is defined as median, the definition of the p-value simplifies to

$$p_r(n, s^{(-)}) = \mathbb{P}_{q \in H_0}(S^{(-)} \geq s^{(-)}) = 0.5^n \sum_{m=s^{(-)}}^{n} \binom{n}{m} \qquad \text{(right-tailed test)}$$

$$p_l(n, s^{(-)}) = \mathbb{P}_{q \in H_0}(S^{(-)} \leq s^{(-)}) = 0.5^n \sum_{m=0}^{s^{(-)}} \binom{n}{m} \qquad \text{(left-tailed test)}$$

The null hypothesis is to reject, if the p-value is lower than or equal to the given α-level, otherwise H_0 can not be rejected.

It should be noted that if the hypothesized quantile value is chosen close to the true value, the common sign test generally does not result in a rejection of H_0 since there is on average no strong evidence against the validity of H_0. Thus, to ensure a significant test result, the required sample size is to specify the larger, the closer is the hypothesized quantile value to the true value (see also Dixon and Mood[15]). However, embedding of fuzzy hypotheses and fuzzy categories may lead to improved test results and may rather enable a rejection of H_0 when the hypothesized quantile value is chosen very close to the true one in combination with a comparatively small sample size. This is due to (1) the inclusion of the hypothesized percentage deviation regarding q within reformulated hypotheses by means of the specification of q_0, q_1 on the one hand (fuzzy hypotheses) and (2) embedding of the magnitude of observations in the range (x_l, x_r) around the hypothesized quantile value k into the test statistic on the other hand (fuzzy categories). Consequently, we discuss modeling of fuzzy hypotheses and fuzzy categories in Sections 4 and 5, respectively.

3.2 | Outline for the case of the category "positive signs"

Assume again that the n random variables X_i, $i = 1, \ldots, n$, are stochastically independent and are drawn from a continuous distribution, whose quantile k_q is uniquely defined by constraint (1). Considering the membership of X_i ($m_{C^{(+)}}(X_i) \in \{0, 1\}$) to the category "positive signs" ($C^{(+)}$) as success, that is, $\mathbb{P}(m_{C^{(+)}}(X_i) = 1) = 1 - q$ and $\mathbb{P}(m_{C^{(+)}}(X_i) = 0) = q$, the random variable $S^{(+)} = \sum_{i=1}^{n} m_{C^{(+)}}(X_i)$ (in the following: the test statistic) is defined as the number of positive signs (i.e., the number of X_i, $i = 1, \ldots, n$, where $X_i > k$ holds) in the sample and follows the binomial distribution with probability mass function

$$\mathbb{P}(S^{(+)} = s^{(+)}) = \binom{n}{s^{(+)}}(1-q)^{s^{(+)}} q^{n-s^{(+)}}$$

This alternative test statistic leads in turn to the exact binomial test with the following adjustments in

1. Hypotheses formulation:

   While the preliminary hypotheses $H_0^{'}$ and $H_1^{'}$ remain unchanged (see Table 1, rows 1–2), the reformulated hypotheses H_0 and H_1 with respective indicator functions (see Table 1, rows 3–11) should be alternatively remodeled to statements on $1 - q$ (instead of statements on q) in compliance with the remodeled test statistic (see Table 2, rows 3–11).
2. Power function:

$$G_{n,c^{(+)}}(1-q) = \sum_{m=0}^{c^{(+)}-1} \binom{n}{m}(1-q)^m q^{n-m} \quad \text{with } 1 \le c^{(+)} \le n, c^{(+)} \in \mathbb{N} \quad (\text{left-tailed test})$$

$$G_{n,c^{(+)}}(1-q) = \sum_{m=c^{(+)}+1}^{n} \binom{n}{m}(1-q)^m q^{n-m} \quad \text{with } 0 \le c^{(+)} \le n-1, c^{(+)} \in \mathbb{N}_0 \quad (\text{right-tailed test})$$

where $c^{(+)}$ is the respective critical (rejection) value.

TABLE 2 Subsets of the parameter space and their indicator functions (using category "positive signs")

	Preliminary left-tailed test	**Preliminary right-tailed test**
Hypothesis H_0'	$k_q \geq k$	$k_q \leq k$
Hypothesis H_1'	$k_q < k$	$k_q > k$
	Left-tailed test	Right-tailed test
Hypothesis H_0	$1-q \geq 1-q_0$	$1-q \leq 1-q_0$
Hypothesis H_1	$1-q < 1-q_1$	$1-q > 1-q_1$
With	$0 < q_0 = q_1 < 1$	$0 < q_1 = q_0 < 1$
Set H_0	$\{(1-q) \in \Theta \| 1-q \geq 1-q_0\}$	$\{(1-q) \in \Theta \| 1-q \leq 1-q_0\}$
Set H_1	$\{(1-q) \in \Theta \| 1-q < 1-q_1\}$	$\{(1-q) \in \Theta \| 1-q > 1-q_1\}$
SSet I	$\{(1-q) \in \Theta \| 1-q_1 \leq 1-q < 1-q_0\} = \varnothing$	$\{(1-q) \in \Theta \| 1-q_0 < 1-q \leq 1-q_1\} = \varnothing$
$m_{H_0}(1-q) =$	$\begin{cases} 0 & \text{if } 0 < 1-q < 1-q_0 \\ 1 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 1 & \text{if } 0 < 1-q \leq 1-q_0 \\ 0 & \text{if } 1-q_0 < 1-q < 1 \end{cases}$
$m_{H_1}(1-q) =$	$\begin{cases} 1 & \text{if } 0 < 1-q < 1-q_1 \\ 0 & \text{if } 1-q_1 \leq 1-q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_1 \\ 1 & \text{if } 1-q_1 < 1-q < 1 \end{cases}$
$m_I(1-q) =$	$\begin{cases} 0 & \text{if } 0 < 1-q < 1-q_1 \\ 1 & \text{if } 1-q_1 \leq 1-q < 1-q_0 \\ 0 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_0 \\ 1 & \text{if } 1-q_0 < 1-q \leq 1-q_1 \\ 0 & \text{if } 1-q_1 < 1-q < 1 \end{cases}$

3. Type I error criterion:

$$E_1(n, c^{(+)}) = \sup_{1-q \in H_0} G_{n,c^{(+)}}(1-q) = G_{n,c^{(+)}}(1-q_0) \leq \alpha$$

4. p-value:

$$\begin{aligned} p_l(n, s^{(+)}) &= \mathbb{P}_{1-q \in H_0}(S^{(+)} \leq s^{(+)}) = \sum_{m=0}^{s^{(+)}} \binom{n}{m} (1-q)^m q^{n-m} \\ &= \sum_{m=0}^{s^{(+)}} \binom{n}{m} (1-q_0)^m q_0^{n-m} \qquad \text{(left-tailed test)} \\ p_r(n, s^{(+)}) &= \mathbb{P}_{1-q \in H_0}(S^{(+)} \geq s^{(+)}) = \sum_{m=s^{(+)}}^{n} \binom{n}{m} (1-q)^m q^{n-m} \\ &= \sum_{m=s^{(+)}}^{n} \binom{n}{m} (1-q_0)^m q_0^{n-m} \qquad \text{(right-tailed test)} \end{aligned}$$

If the quantile of interest is defined as median, the definition of the p-value simplifies to

$$p_l(n, s^{(+)}) = \mathbb{P}_{1-q \in H_0}(S^{(+)} \leq s^{(+)}) = 0.5^n \sum_{m=0}^{s^{(+)}} \binom{n}{m} \qquad \text{(left-tailed test)}$$

$$p_r(n, s^{(+)}) = \mathbb{P}_{1-q \in H_0}(S^{(+)} \geq s^{(+)}) = 0.5^n \sum_{m=s^{(+)}}^{n} \binom{n}{m} \qquad \text{(right-tailed test)}$$

Summarized, when considering both test statistics $S^{(-)}$ and $S^{(+)}$ based on categories "negative signs" ($C^{(-)}$) and "positive signs" ($C^{(+)}$), we obtain a consistent representation of the p-value for preliminary formulated left- and right-tailed test, respectively, as follows

$$\mathbb{P}_{q\in H_0}(S^{(-)} \geq s^{(-)}) = \mathbb{P}_{1-q\in H_0}(S^{(+)} \leq s^{(+)})$$
$$\mathbb{P}_{q\in H_0}(S^{(-)} \leq s^{(-)}) = \mathbb{P}_{1-q\in H_0}(S^{(+)} \geq s^{(+)})$$

since it holds $s^{(-)} + s^{(+)} = n$. In addition, the calculation effort can be reduced to the determination of the "left-tailed" p-value for both types of preliminary formulated one-tailed hypotheses

$$p_l(n, s^{(+)}) = \mathbb{P}_{1-q\in H_0}(S^{(+)} \leq s^{(+)}) \qquad \text{(preliminary left-tailed test)} \tag{2}$$

$$p_l(n, s^{(-)}) = \mathbb{P}_{q\in H_0}(S^{(-)} \leq s^{(-)}) \qquad \text{(preliminary right-tailed test)} \tag{3}$$

In the following, we use this representation for definitions and calculations in Sections 4–6.

4 | ONE-TAILED TEST OF SIGNIFICANCE FOR QUANTILES WITH FUZZY HYPOTHESES

4.1 | The case of the category "negative signs"

Given that a crisp hypothesis is a special case of a fuzzy hypothesis, the conventional test problem described above is now to generalize using the basic approach of Chukhrova and Johannssen[20] and to solve in four steps presented in Section 3.

In the first step, we have to formulate adequate fuzzy hypotheses. Since we assume that fuzziness of the hypotheses is caused by imprecision of linguistic statements on fractions of underlying quantiles, we consider fuzzification of hypotheses in this paper only regarding reformulated hypotheses as statements on an unknown fraction q, not regarding preliminary hypotheses as statements on an unknown quantile k_q. The preliminary hypotheses $H_0^{'}$ and $H_1^{'}$ are still crisp defined as in Section 3 (see Table 3, rows 1–2). Fuzzifying solely the reformulated hypotheses H_0 and H_1 is quite natural due to the fact that the sign test is a special case of the binomial test and there is in the first line an uncertainty regarding the magnitude of the fraction q, defined uniquely by a crisp hypothesized quantile value k. Considering exemplary a test for the median, such an uncertainty regarding the 50%-threshold can be appropriately modeled in an interval-valued way instead of a point-valued specification, especially with the help of a symmetric deviation (e.g., 2.5%, 5%, or 10%) from the 50%-point. An interval-valued specification, obtained by hypothesized values q_0 and q_1, constitutes the indifference zone and simultaneously the fuzzy areas of the reformulated hypotheses.

The particular formulation of fuzzy hypotheses as fuzzy statements on the fraction q of interest with parameter space $\Theta := (0, 1)$ (see Table 3, rows 3–5, and Figure 3) implies the definition of fuzzy sets H_0, H_1, I (with $H_0 \cup H_1 \cup I = \Theta$) as well as modeling of membership functions $m_{H_0}(q), m_{H_1}(q), m_I(q)$ for $q \in \Theta$ (instead of indicator functions). Table 3 (rows 6–11) gives an overview of these sets and their membership functions.

TABLE 3 Fuzzy subsets of the parameter space and their membership functions (using category "negative signs")

	Preliminary left-tailed test	Preliminary right-tailed test
Preliminary hypothesis H_0'	$k_q \geq k$	$k_q \leq k$
Preliminary hypothesis H_1'	$k_q < k$	$k_q > k$
	Right-tailed test	Left-tailed test
Hypothesis H_0	$q \leq q_0, q \boxed{\leq} q_1$	$q \boxed{\geq} q_1, q \geq q_0$
Hypothesis H_1	$q \boxed{>} q_0, q > q_1$	$q < q_1, q \boxed{<} q_0$
With	$0 < q_0 \leq q_1 < 1$	$0 < q_1 \leq q_0 < 1$
Set H_0	$\{(q; m_{H_0}(q))\|q \in \Theta, m_{H_0}(q) \in [0,1]\}$ $\text{ncl}(H_0) = \{q \in \Theta\|q \leq q_0\}$	$\{(q; m_{H_0}(q))\|q \in \Theta, m_{H_0}(q) \in [0,1]\}$ $\text{ncl}(H_0) = \{q \in \Theta\|q \geq q_0\}$
Set H_1	$\{(q; m_{H_1}(q))\|q \in \Theta, m_{H_1}(q) \in [0,1]\}$ $\text{ncl}(H_1) = \{q \in \Theta\|q > q_1\}$	$\{(q; m_{H_1}(q))\|q \in \Theta, m_{H_1}(q) \in [0,1]\}$ $\text{ncl}(H_1) = \{q \in \Theta\|q < q_1\}$
Set I	$\{(q; m_I(q))\|q \in \Theta, m_I(q) \in \{0,1\}\}$	$\{(q; m_I(q))\|q \in \Theta, m_I(q) \in \{0,1\}\}$
$m_{H_0}(q) =$	$\begin{cases} 1 & \text{if } 0 < q \leq q_0 \\ u_0(q) & \text{if } q_0 < q < q_1 \\ m_{H_0^C}(q) & \text{if } q = q_1 \\ 0 & \text{if } q_1 < q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < q < q_1 \\ m_{H_0^C}(q) & \text{if } q = q_1 \\ u_0(q) & \text{if } q_1 < q < q_0 \\ 1 & \text{if } q_0 \leq q < 1 \end{cases}$
$m_{H_1}(q) =$	$\begin{cases} 0 & \text{if } 0 < q \leq q_0 \\ u_1(q) & \text{if } q_0 < q \leq q_1 \\ 1 & \text{if } q_1 < q < 1 \end{cases}$	$\begin{cases} 1 & \text{if } 0 < q < q_1 \\ u_1(q) & \text{if } q_1 \leq q < q_0 \\ 0 & \text{if } q_0 \leq q < 1 \end{cases}$
$m_I(q) =$	$\begin{cases} 0 & \text{if } 0 < q \leq q_0 \\ 1 & \text{if } q_0 < q \leq q_1 \\ 0 & \text{if } q_1 < q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < q < q_1 \\ 1 & \text{if } q_1 \leq q < q_0 \\ 0 & \text{if } q_0 \leq q < 1 \end{cases}$

In contrast to crisp reformulated hypotheses, fuzzy reformulated hypotheses are formulated using both hypothesized fraction values q_0, q_1. Accordingly, it is to distinguish between crisp comparison operators ($<, \leq, \geq, >$) and fuzzy comparison operators ($\boxed{<}, \boxed{\leq}, \boxed{\geq}, \boxed{>}$) in the formulation of the hypotheses. For example, the operator $\leq$ denotes "crisp" lower equal, while the operator $\boxed{\leq}$ stands for "fuzzy" lower equal. Thus, the fuzziness of the hypotheses is caused by a gradual fuzzification of the indifference zone I in terms of fuzzified intervals $(q_0, q_1]$ (right-tailed test) and $[q_1, q_0)$ (left-tailed test), respectively (see Figure 3). Considering exemplary a right-tailed test for the median, linguistic formulations of H_0 and H_1 like

H_0: The population proportion q is approximately lower than or equal to 0.5
H_1: The population proportion q is approximately larger than 0.5

could appropriately be modeled via a symmetric specification of q_0 and q_1 regarding the threshold 0.5 by means of the hypothesized deviation value from this 50%-point, whose range expresses the magnitude of uncertainty regarding q, for example, given an uncertainty level of 10% one would choose the width of the indifference zone (and thus of the fuzzy areas) as $w = q_1 - q_0 = 0.1$, that is, it holds $q_0 = 0.45$ and $q_1 = 0.55$.

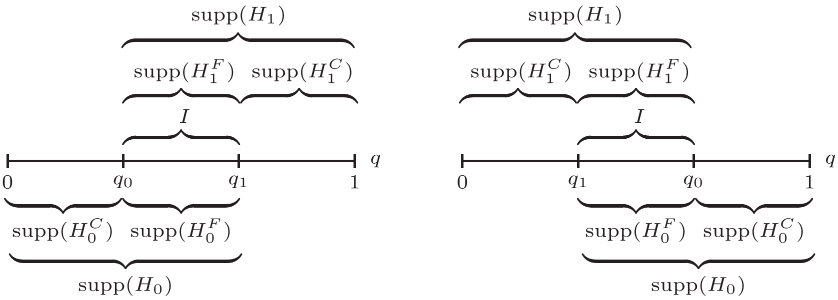

FIGURE 3 Representation of fuzzy reformulated hypotheses H_0, H_1 for the right- and left-tailed case

Due to the generalization of the crisp sets H_0, H_1 to fuzzy sets H_0, H_1, they now consist of a crisp and a fuzzy set: $H_0 = H_0^C \cup H_0^F$, $H_1 = H_1^C \cup H_1^F$, which are referred to as crisp and fuzzy areas of the respective hypothesis (see Table 4, rows 1–4). The edge elements of the nonempty supports of H_0^F and H_1^F are thereby based on the hypothesized values q_0, q_1. In contrast to the normal crisp areas H_0^C and H_1^C (see Table 4, rows 5–6), the fuzzy areas $H_0^F \neq \emptyset$ and $H_1^F \neq \emptyset$ are subnormal. The membership functions $u_0(q)$ and $u_1(q)$ for these areas (see Table 4, rows 7–8) are (strictly) monotonically decreasing or increasing functions depending on whether the membership degree of any element of the parameter set has a decreasing or an increasing trend in relation to the respective hypothesis.

Since the fuzzy sets H_0, H_1 are formulated as union of their crisp and fuzzy areas ($\text{supp}(H_0) = \text{supp}(H_0^C) \cup \text{supp}(H_0^F)$, $\text{supp}(H_1) = \text{supp}(H_1^C) \cup \text{supp}(H_1^F)$), the membership functions $m_{H_0}(q)$ and $m_{H_1}(q)$ result as follows (see also Table 3, rows 9–10, and Table 4, rows 5–8):

$$m_{H_0}(q) = m_{H_0^C \cup H_0^F}(q) = \max\{m_{H_0^C}(q), m_{H_0^F}(q)\}$$
$$m_{H_1}(q) = m_{H_1^C \cup H_1^F}(q) = \max\{m_{H_1^C}(q), m_{H_1^F}(q)\}$$

Consequently, these functions are monotonically increasing or decreasing. In addition, the membership functions $m_{H_0}(q)$ and $m_{H_1}(q)$ can be defined either complementary, that is $m_{H_0}(q) = 1 - m_{H_1}(q)$, or noncomplementary, that is $m_{H_0}(q) \leq 1 - m_{H_1}(q)$. As for non-complementary hypotheses, there is at least one $q \in \Theta$ with $m_{H_0}(q) < 1 - m_{H_1}(q)$. However, a consideration of another type of noncomplementary membership functions, that is, $m_{H_0}(q) \geq 1 - m_{H_1}(q)$, is less relevant in practice and is therefore not considered in this paper.

It should be noted that the specification of $q_0 = q_1$ implies crisp complementary hypotheses with $m_{H_0}(q) = m_{H_0^C}(q)$, $m_{H_1}(q) = m_{H_1^C}(q)$ (due to $H_0^F = \emptyset$, $H_1^F = \emptyset$). In the cases $q_0 < q_1$ (right-tailed test) or $q_1 < q_0$ (left-tailed test), both fuzzy complementary and noncomplementary hypotheses can be modeled.

Now, we exemplary consider piecewise linear, s-, convex- or concave-shaped membership functions $m_{H_0}(q)$, $m_{H_1}(q)$ with $u_0(q)$, $u_1(q)$ given in Table 5. In addition, Figures 4–5 illustrate these membership functions and demonstrate the gradual fuzzification of the indifference zone (fuzzy complementary hypotheses). We refer for a sensitivity analysis regarding the impact of various shapes of membership functions in the context of fuzzy hypothesis testing to Chukhrova and Johannssen.[24]

In the second step, depending on the respective test situation, the user specifies the sample size n, $n \in \mathbb{N}$, and the significance level α, $\alpha \in (0, 1)$. In the third step, under H_0 the test statistic $S^{(-)}$ can

TABLE 4 Crisp and fuzzy areas of the hypotheses and their membership functions (using category "negative signs")

	Right-tailed test	Left-tailed test
Set H_0^C	$\{(q; m_{H_0^C}(q))\|q \in \Theta, m_{H_0^C}(q) \in \{0,1\}\}$ $\text{supp}(H_0^C) = \{q \in \Theta \| q \le q_0\}$	$\{(q; m_{H_0^C}(q))\|q \in \Theta, m_{H_0^C}(q) \in \{0,1\}\}$ $\text{supp}(H_0^C) = \{q \in \Theta \| q \ge q_0\}$
Set H_1^C	$\{(q; m_{H_1^C}(q))\|q \in \Theta, m_{H_1^C}(q) \in \{0,1\}\}$ $\text{supp}(H_1^C) = \{q \in \Theta \| q > q_1\}$	$\{(q; m_{H_1^C}(q))\|q \in \Theta, m_{H_1^C}(q) \in \{0,1\}\}$ $\text{supp}(H_1^C) = \{q \in \Theta \| q < q_1\}$
Set H_0^F	$\{(q; m_{H_0^F}(q))\|q \in \Theta, m_{H_0^F}(q) \in [0,1]\}$ $\text{supp}(H_0^F) = \{q \in \Theta \| q_0 < q < q_1\}$	$\{(q; m_{H_0^F}(q))\|q \in \Theta, m_{H_0^F}(q) \in [0,1]\}$ $\text{supp}(H_0^F) = \{q \in \Theta \| q_1 < q < q_0\}$
Set H_1^F	$\{(q; m_{H_1^F}(q))\|q \in \Theta, m_{H_1^F}(q) \in [0,1]\}$ $\text{supp}(H_1^F) = \{q \in \Theta \| q_0 < q \le q_1\}$	$\{(q; m_{H_1^F}(q))\|q \in \Theta, m_{H_1^F}(q) \in [0,1]\}$ $\text{supp}(H_1^F) = \{q \in \Theta \| q_1 \le q < q_0\}$
$m_{H_0^C}(q) =$	$\begin{cases} 1 & \text{if } 0 < q \le q_0 \\ 0 & \text{if } q_0 < q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < q < q_0 \\ 1 & \text{if } q_0 \le q < 1 \end{cases}$
$m_{H_1^C}(q) =$	$\begin{cases} 0 & \text{if } 0 < q \le q_1 \\ 1 & \text{if } q_1 < q < 1 \end{cases}$	$\begin{cases} 1 & \text{if } 0 < q < q_1 \\ 0 & \text{if } q_1 \le q < 1 \end{cases}$
$m_{H_0^F}(q) =$	$\begin{cases} 0 & \text{if } 0 < q \le q_0 \\ u_0(q) & \text{if } q_0 < q < q_1 \\ 0 & \text{if } q_1 \le q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < q \le q_1 \\ u_0(q) & \text{if } q_1 < q < q_0 \\ 0 & \text{if } q_0 \le q < 1 \end{cases}$
$m_{H_1^F}(q) =$	$\begin{cases} 0 & \text{if } 0 < q \le q_0 \\ u_1(q) & \text{if } q_0 < q \le q_1 \\ 0 & \text{if } q_1 < q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < q < q_1 \\ u_1(q) & \text{if } q_1 \le q < q_0 \\ 0 & \text{if } q_0 \le q < 1 \end{cases}$

be calculated via $\sum_{i=1}^{n} m_{C^{(-)}}(X_i)$ by using a sample of size n. In addition, a crisp test decision is to make in the fourth step by comparing the α-level to the generalized p-value. The specifications of the respective test decisions for the generalized test are identical to those for the conventional test.

As for the particular determination of the generalized p-value, the respective results from Section 3 can be adopted only for elements from the crisp area of H_0 but not from the fuzzy area (indifference zone). Therefore, the definition of the p-value is to generalize in relation to the fuzzy area of H_0, particularly in compliance with the definition of the generalized type I error criterion,

$$E_1(n, c^{(-)}) = \sup_{q \in H_0} \left\{ (m_{H_0}(q) - m_{H_1}(q)) G_{n,c^{(-)}}(q) \right\} \le \alpha. \tag{4}$$

The generalized error criterion E_1 is defined as the supremum of weighted probabilities for false rejection of H_0. In particular, the probabilities correspond to the power function $G_{n,c^{(-)}}(q)$, while the weight function is defined as the difference between the membership of an element $q \in \Theta$ to the fuzzy null hypothesis and the membership of an element $q \in \Theta$ to the fuzzy alternative hypothesis, that is $m_{H_0}(q) - m_{H_1}(q)$ for all $q \in \Theta$ (see also Arnold[25,26]). Thus, with respect to the type I error, these weights consider the loss amount, which results depending on the membership of a particular element of the parameter space to the respective fuzzy hypothesis. For example, for $q \in \text{ncl}(H_0)$, the weight is constantly equal to one (as also for crisp formulated hypotheses):

$$m_{H_0}(q) - m_{H_1}(q) = 1 - 0 = 1 \quad \forall q \in \text{ncl}\ (H_0) \text{ and thus } \forall q \in \text{supp}\left(H_0^C\right)$$

In addition, definition (4) can be reformulated to

$$\begin{aligned} E_1(n, c^{(-)}) &= \sup_{q \in \text{supp}(H_0)} \left\{(m_{H_0}(q) - m_{H_1}(q))G_{n,c^{(-)}}(q)\right\} \\ &= \max\left\{\begin{matrix} \sup\limits_{q \in \text{supp}\left(H_0^C\right)} \left\{(m_{H_0}(q) - m_{H_1}(q))G_{n,c^{(-)}}(q)\right\} \\ \sup\limits_{q \in \text{supp}\left(H_0^F\right)} \left\{(m_{H_0}(q) - m_{H_1}(q))G_{n,c^{(-)}}(q)\right\} \end{matrix}\right\} \\ &= \max\left\{\begin{matrix} G_{n,c^{(-)}}(q_0) \\ \sup\limits_{q \in \text{supp}\left(H_0^F\right)} \left\{(m_{H_0}(q) - m_{H_1}(q))G_{n,c^{(-)}}(q)\right\} \end{matrix}\right\} \end{aligned}$$

due to the fact that only those elements of the null hypothesis, which lead to positive weights of the power function, are of interest for the determination of the supremum of weighted probabilities. Therefore, the definition of the supremum can be focused on elements of the support of the null hypothesis that corresponds to the union of the supports of its crisp and fuzzy areas. Given that the power function $G_{n,c^{(-)}}(q)$ is increasing for right-tailed test problems or decreasing for left-tailed test problems on the one hand as well as the relationship

$$m_{H_0}(q) - m_{H_1}(q) = 1 \quad \forall q \in \text{supp}\left(H_0^C\right)$$

holds on the other hand, E_1 has a supremum for the elements of the support of H_0^C at the point $q = q_0$. Finally, it should be noted that in the case of crisp hypotheses the support of the fuzzy area corresponds to the empty set, which implies $\text{supp}(H_0) = \text{supp}(H_0^C)$. Thus we obtain the results presented in Section 3.

Due to the monotonicity of the weight and power functions, for $q \in \text{supp}(H_0^F)$ and $\text{supp}(H_0^F) \neq \emptyset$ the generalized error criterion also has a supremum, which is in general not representable in closed form. Instead, this supremum can be determined numerically, and afterward it is to compare with the supremum of the support of the crisp area.

Applying the above results to the definition of the p-value in the case of fuzzy hypotheses, the generalized p-values (i.e., weighted p-values) for a right- and a left-tailed test are then given as follows:

$$\begin{aligned} p_r(n, s^{(-)}) &= \sup_{q \in H_0}\left\{(m_{H_0}(q) - m_{H_1}(q))\mathbb{P}_{q \in H_0}(S^{(-)} \geq s^{(-)})\right\} \\ &= \sup_{q \in H_0}\left\{(m_{H_0}(q) - m_{H_1}(q)) \sum_{m=s^{(-)}}^{n} \binom{n}{m} q^m (1-q)^{n-m}\right\} \\ p_l(n, s^{(-)}) &= \sup_{q \in H_0}\left\{(m_{H_0}(q) - m_{H_1}(q))\mathbb{P}_{q \in H_0}(S^{(-)} \leq s^{(-)})\right\} \\ &= \sup_{q \in H_0}\left\{(m_{H_0}(q) - m_{H_1}(q)) \sum_{m=0}^{s^{(-)}} \binom{n}{m} q^m (1-q)^{n-m}\right\} \end{aligned}$$

TABLE 5 Exemplary shapes of membership functions (using category "negative signs")

Shape	Right-tailed test	Left-tailed test
Linear	$u_0(q) = \frac{q_1 - q}{q_1 - q_0}$ if $q_0 < q \le q_1$ $u_1(q) = \frac{q - q_0}{q_1 - q_0}$ if $q_0 < q \le q_1$	$u_0(q) = \frac{q - q_1}{q_0 - q_1}$ if $q_1 \le q < q_0$ $u_1(q) = \frac{q_0 - q}{q_0 - q_1}$ if $q_1 \le q < q_0$
s (polynomial)	$u_0(q) = \begin{cases} 1 - 2\left(\frac{q - q_0}{q_1 - q_0}\right)^2 & \text{if } q_0 < q \le \frac{q_0 + q_1}{2} \\ 2\left(\frac{q_1 - q}{q_1 - q_0}\right)^2 & \text{if } \frac{q_0 + q_1}{2} < q \le q_1 \end{cases}$ $u_1(q) = \begin{cases} 2\left(\frac{q - q_0}{q_1 - q_0}\right)^2 & \text{if } q_0 < q \le \frac{q_0 + q_1}{2} \\ 1 - 2\left(\frac{q_1 - q}{q_1 - q_0}\right)^2 & \text{if } \frac{q_0 + q_1}{2} < q \le q_1 \end{cases}$	$u_0(q) = \begin{cases} 2\left(\frac{q - q_1}{q_0 - q_1}\right)^2 & \text{if } q_1 \le q < \frac{q_0 + q_1}{2} \\ 1 - 2\left(\frac{q_0 - q}{q_0 - q_1}\right)^2 & \text{if } \frac{q_0 + q_1}{2} \le q < q_0 \end{cases}$ $u_1(q) = \begin{cases} 1 - 2\left(\frac{q - q_1}{q_0 - q_1}\right)^2 & \text{if } q_1 \le q < \frac{q_0 + q_1}{2} \\ 2\left(\frac{q_0 - q}{q_0 - q_1}\right)^2 & \text{if } \frac{q_0 + q_1}{2} \le q < q_0 \end{cases}$
Convex	$u_0(q) = \exp\left(\frac{-6(q - q_0)}{q_1 - q_0}\right)$ if $q_0 < q \le q_1$ $u_1(q) = \exp\left(\frac{-6(q_1 - q)}{q_1 - q_0}\right)$ if $q_0 < q \le q_1$	$u_0(q) = \exp\left(\frac{-6(q_0 - q)}{q_0 - q_1}\right)$ if $q_1 \le q < q_0$ $u_1(q) = \exp\left(\frac{-6(q - q_1)}{q_0 - q_1}\right)$ if $q_1 \le q < q_0$
Concave	$u_0(q) = 1 - \exp\left(\frac{-6(q_1 - q)}{q_1 - q_0}\right)$ if $q_0 < q \le q_1$ $u_1(q) = 1 - \exp\left(\frac{-6(q - q_0)}{q_1 - q_0}\right)$ if $q_0 < q \le q_1$	$u_0(q) = 1 - \exp\left(\frac{-6(q - q_1)}{q_0 - q_1}\right)$ if $q_1 \le q < q_0$ $u_1(q) = 1 - \exp\left(\frac{-6(q_0 - q)}{q_0 - q_1}\right)$ if $q_1 \le q < q_0$

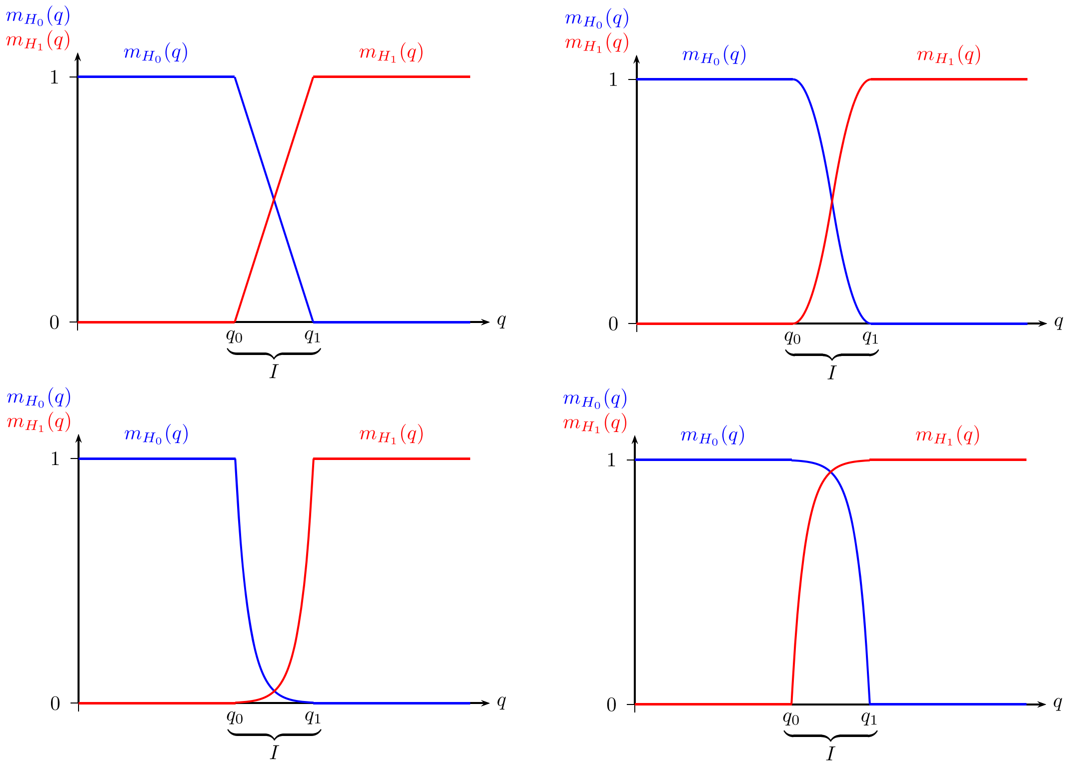

FIGURE 4 Different shapes of membership functions (piecewise linear, s-shaped, convex, concave), right-tailed case [Color figure can be viewed at wileyonlinelibrary.com]

Considering in addition the distinction between crisp and fuzzy areas, we obtain in the right- and left-tailed case:

$$p_r(n, s^{(-)}) = \max \begin{cases} \sum_{m=s^{(-)}}^{n} \binom{n}{m} q_0^m (1-q_0)^{n-m} \\ \sup_{q \in \operatorname{supp}(H_0^F)} \left\{ (m_{H_0}(q) - m_{H_1}(q)) \sum_{m=s^{(-)}}^{n} \binom{n}{m} q^m (1-q)^{n-m} \right\} \end{cases}$$
$$p_l(n, s^{(-)}) = \max \begin{cases} \sum_{m=0}^{s^{(-)}} \binom{n}{m} q_0^m (1-q_0)^{n-m} \\ \sup_{q \in \operatorname{supp}(H_0^F)} \left\{ (m_{H_0}(q) - m_{H_1}(q)) \sum_{m=0}^{s^{(-)}} \binom{n}{m} q^m (1-q)^{n-m} \right\} \end{cases} \tag{5}$$

Finally, let us interpret the generalized p-value under separate consideration of crisp and fuzzy areas. If the global maximum of the generalized p-value stems from the crisp area of H_0, then the respective conditional probability (weighted by one) that the observed or more extreme (in terms of H_0) sampling results occur when H_0 is true, is of probabilistic nature as in common statistical hypothesis testing. If the global maximum originates from the fuzzy area, then the respective weighted conditional probability falls below the (unweighted) respective p-value in terms of an

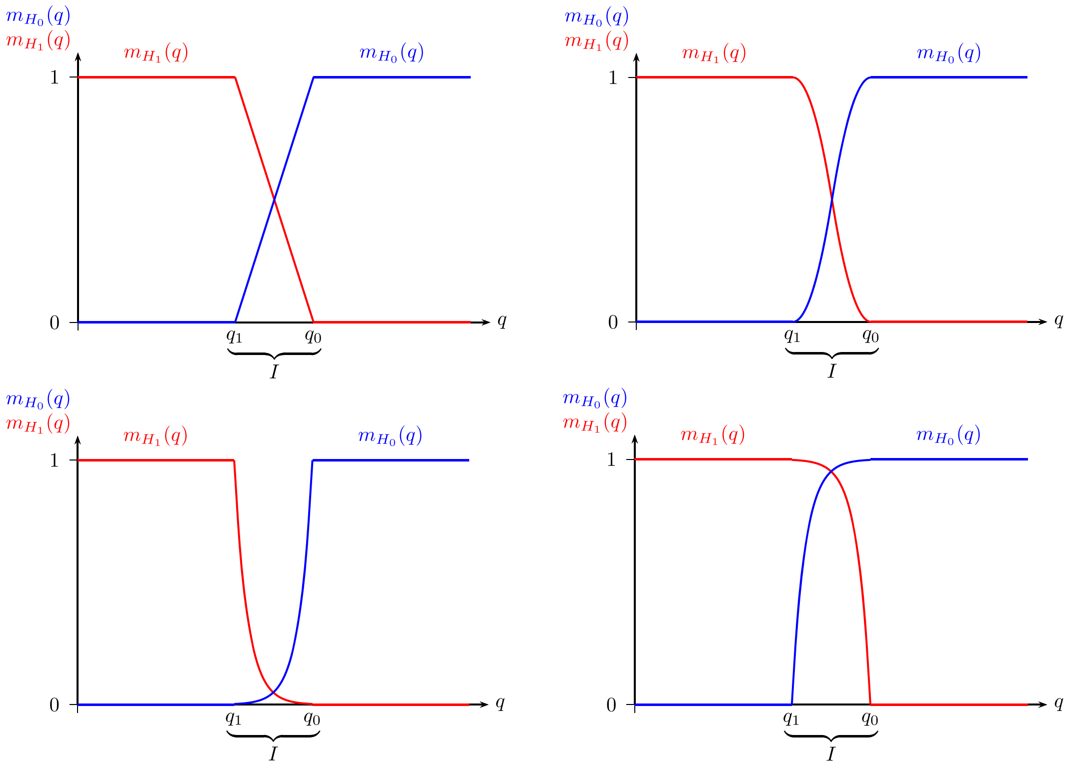

FIGURE 5 Different shapes of membership functions (piecewise linear, s-shaped, convex, concave), left-tailed case [Color figure can be viewed at wileyonlinelibrary.com]

appropriate extent constituted by the fuzziness of hypotheses. In this case, the generalized p-value is in turn of probabilistic nature but adequately adjusted by a subjective component, which reflects exemplary integrating of matters of economic or possibilistic nature into the consideration, to close, reduce, or express existing uncertainties or imprecisions of linguistic statements.

4.2 | Outline for the case of the category "positive signs"

Using the alternative test statistic $S^{(+)}$, defined as the number of positive signs in the sample, the adjustments for the generalized binomial test are as follows:

1. Hypotheses formulation (see Table 6):
   While the preliminary hypotheses $H_0^{'}$ and $H_1^{'}$ remain unchanged (see Table 3, rows 1–2), the reformulated fuzzy hypotheses H_0 and H_1 with respective indicator functions (see Table 3, rows 3–11) should be alternatively remodeled to statements on $1 - q$ (instead of statements on q) with crisp and fuzzy areas (see Table 7) in compliance with the remodeled test statistic.
2. Shapes of membership functions (see Table 8)
3. Generalized type I error criterion:

TABLE 6 Fuzzy subsets of the parameter space and their membership functions (using category "positive signs")

	Preliminary left-tailed test	Preliminary right-tailed test
Hypothesis H_0'	$k_q \geq k$	$k_q \leq k$
Hypothesis H_1'	$k_q < k$	$k_q > k$
	Left-tailed test	Right-tailed test
Hypothesis H_0	$1-q \boxed{\geq} 1-q_1, 1-q \geq 1-q_0$	$1-q \leq 1-q_0, 1-q \boxed{\leq} 1-q_1$
Hypothesis H_1	$1-q < 1-q_1, 1-q \boxed{<} 1-q_0$	$1-q \boxed{>} 1-q_0, 1-q > 1-q_1$
With	$0 < q_0 \leq q_1 < 1$	$0 < q_1 \leq q_0 < 1$
Set H_0	$\{(1-q; m_{H_0}(1-q)) \mid 1-q \in \Theta, m_{H_0}(1-q) \in [0,1]\}$ $\text{ncl}(H_0) = \{1-q \in \Theta \mid 1-q \geq 1-q_0\}$	$\{(1-q; m_{H_0}(1-q)) \mid 1-q \in \Theta, m_{H_0}(1-q) \in [0,1]\}$ $\text{ncl}(H_0) = \{1-q \in \Theta \mid 1-q \leq 1-q_0\}$
Set H_1	$\{(1-q; m_{H_1}(1-q)) \mid 1-q \in \Theta, m_{H_1}(1-q) \in [0,1]\}$ $\text{ncl}(H_1) = \{1-q \in \Theta \mid 1-q < 1-q_1\}$	$\{(1-q; m_{H_1}(1-q)) \mid 1-q \in \Theta, m_{H_1}(1-q) \in [0,1]\}$ $\text{ncl}(H_1) = \{1-q \in \Theta \mid 1-q > 1-q_1\}$
Set I	$\{(1-q; m_I(1-q)) \mid 1-q \in \Theta, m_I(1-q) \in \{0,1\}\}$	$\{(1-q; m_I(1-q)) \mid 1-q \in \Theta, m_I(1-q) \in \{0,1\}\}$
$m_{H_0}(1-q)=$	$\begin{cases} 0 & \text{if } 0 < 1-q < 1-q_1 \\ m_{H_0^C}(1-q) & \text{if } 1-q = 1-q_1 \\ u_0(1-q) & \text{if } 1-q_1 < 1-q < 1-q_0 \\ 1 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 1 & \text{if } 0 < 1-q \leq 1-q_0 \\ u_0(1-q) & \text{if } 1-q_0 < 1-q < 1-q_1 \\ m_{H_0^C}(1-q) & \text{if } 1-q = 1-q_1 \\ 0 & \text{if } 1-q_1 < 1-q < 1 \end{cases}$
$m_{H_1}(1-q)=$	$\begin{cases} 1 & \text{if } 0 < 1-q < 1-q_1 \\ u_1(1-q) & \text{if } 1-q_1 \leq 1-q < 1-q_0 \\ 0 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_0 \\ u_1(1-q) & \text{if } 1-q_0 < 1-q \leq 1-q_1 \\ 1 & \text{if } 1-q_1 < 1-q < 1 \end{cases}$
$m_I(1-q)=$	$\begin{cases} 0 & \text{if } 0 < 1-q < 1-q_1 \\ 1 & \text{if } 1-q_1 \leq 1-q < 1-q_0 \\ 0 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_0 \\ 1 & \text{if } 1-q_0 < 1-q \leq 1-q_1 \\ 0 & \text{if } 1-q_1 < 1-q < 1 \end{cases}$

$$
\begin{aligned}
E_1(n, c^{(+)}) &= \sup_{1-q \in \text{supp}(H_0)} \left\{ (m_{H_0}(1-q) - m_{H_1}(1-q)) G_{n,c^{(+)}}(1-q) \right\} \\
&= \max \left\{ \begin{array}{c} G_{n,c^{(+)}}(1-q_0) \\ \sup\limits_{1-q \in \text{supp}\left(H_0^F\right)} \left\{ (m_{H_0}(1-q) - m_{H_1}(1-q)) G_{n,c^{(+)}}(1-q) \right\} \end{array} \right\}
\end{aligned}
$$

4. Generalized p-value for left- and right-tailed case:

$$
p_l(n, s^{(+)}) = \max \left\{ \begin{array}{c} \sum\limits_{m=0}^{s^{(+)}} \binom{n}{m} (1-q_0)^m q_0^{n-m} \\ \sup\limits_{1-q \in \text{supp}(H_0^F)} \left\{ (m_{H_0}(1-q) - m_{H_1}(1-q)) \sum\limits_{m=0}^{s^{(+)}} \binom{n}{m} (1-q)^m q^{n-m} \right\} \end{array} \right.
$$

$$
p_r(n, s^{(+)}) = \max \left\{ \begin{array}{c} \sum\limits_{m=s^{(+)}}^{n} \binom{n}{m} (1-q_0)^m q_0^{n-m} \\ \sup\limits_{1-q \in \text{supp}(H_0^F)} \left\{ (m_{H_0}(1-q) - m_{H_1}(1-q)) \sum\limits_{m=s^{(+)}}^{n} \binom{n}{m} (1-q)^m q^{n-m} \right\} \end{array} \right.
$$

TABLE 7 Crisp and fuzzy areas of the hypotheses and their membership functions (using category "positive signs")

	Left-tailed test	Right-tailed test
Set H_0^C	$\{(1-q;\, m_{H_0^C}(1-q)) \mid 1-q \in \Theta,\, m_{H_0^C}(1-q) \in \{0,1\}\}$ $\text{supp}(H_0^C) = \{1-q \in \Theta \mid 1-q \geq 1-q_0\}$	$\{(1-q;\, m_{H_0^C}(1-q)) \mid 1-q \in \Theta,\, m_{H_0^C}(1-q) \in \{0,1\}\}$ $\text{supp}(H_0^C) = \{1-q \in \Theta \mid 1-q \leq 1-q_0\}$
Set H_1^C	$\{(1-q;\, m_{H_1^C}(1-q)) \mid 1-q \in \Theta,\, m_{H_1^C}(1-q) \in \{0,1\}\}$ $\text{supp}(H_1^C) = \{1-q \in \Theta \mid 1-q < 1-q_1\}$	$\{(1-q;\, m_{H_1^C}(1-q)) \mid 1-q \in \Theta,\, m_{H_1^C}(1-q) \in \{0,1\}\}$ $\text{supp}(H_1^C) = \{1-q \in \Theta \mid 1-q > 1-q_1\}$
Set H_0^F	$\{(1-q;\, m_{H_0^F}(1-q)) \mid 1-q \in \Theta,\, m_{H_0^F}(1-q) \in [0,1]\}$ $\text{supp}(H_0^F) = \{1-q \in \Theta \mid 1-q_1 < 1-q < 1-q_0\}$	$\{(1-q;\, m_{H_0^F}(1-q)) \mid 1-q \in \Theta,\, m_{H_0^F}(1-q) \in [0,1]\}$ $\text{supp}(H_0^F) = \{1-q \in \Theta \mid 1-q_0 < 1-q < 1-q_1\}$
Set H_1^F	$\{(1-q;\, m_{H_1^F}(1-q)) \mid 1-q \in \Theta,\, m_{H_1^F}(1-q) \in [0,1]\}$ $\text{supp}(H_1^F) = \{1-q \in \Theta \mid 1-q_1 \leq 1-q < 1-q_0\}$	$\{(1-q;\, m_{H_1^F}(1-q)) \mid 1-q \in \Theta,\, m_{H_1^F}(1-q) \in [0,1]\}$ $\text{supp}(H_1^F) = \{1-q \in \Theta \mid 1-q_0 < 1-q \leq 1-q_1\}$
$m_{H_0^C}(1-q)=$	$\begin{cases} 0 & \text{if } 0 < 1-q < 1-q_0 \\ 1 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 1 & \text{if } 0 < 1-q \leq 1-q_0 \\ 0 & \text{if } 1-q_0 < 1-q < 1 \end{cases}$
$m_{H_1^C}(1-q)=$	$\begin{cases} 1 & \text{if } 0 < 1-q < 1-q_1 \\ 0 & \text{if } 1-q_1 \leq 1-q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_1 \\ 1 & \text{if } 1-q_1 < 1-q < 1 \end{cases}$
$m_{H_0^F}(1-q)=$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_1 \\ u_0(1-q) & \text{if } 1-q_1 < 1-q < 1-q_0 \\ 0 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_0 \\ u_0(1-q) & \text{if } 1-q_0 < 1-q < 1-q_1 \\ 0 & \text{if } 1-q_1 \leq 1-q < 1 \end{cases}$
$m_{H_1^F}(1-q)=$	$\begin{cases} 0 & \text{if } 0 < 1-q < 1-q_1 \\ u_1(1-q) & \text{if } 1-q_1 \leq 1-q < 1-q_0 \\ 0 & \text{if } 1-q_0 \leq 1-q < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < 1-q \leq 1-q_0 \\ u_1(1-q) & \text{if } 1-q_0 < 1-q \leq 1-q_1 \\ 0 & \text{if } 1-q_1 < 1-q < 1 \end{cases}$

TABLE 8 Exemplary shapes of membership functions (using category "positive signs")

Shape	Left-tailed test	Right-tailed test
Linear	$u_0(1-q) = \frac{1-q-(1-q_1)}{1-q_0-(1-q_1)}$ if $1-q_1 \leq 1-q < 1-q_0$	$u_0(1-q) = \frac{1-q_1-(1-q)}{1-q_1-(1-q_0)}$ if $1-q_0 < 1-q \leq 1-q_1$
	$u_1(1-q) = \frac{1-q_0-(1-q)}{1-q_0-(1-q_1)}$ if $1-q_1 \leq 1-q < 1-q_0$	$u_1(1-q) = \frac{1-q-(1-q_0)}{1-q_1-(1-q_0)}$ if $1-q_0 < 1-q \leq 1-q_1$
s (polynomial)	$u_0(1-q) = \begin{cases} 2\left(\frac{1-q-(1-q_1)}{1-q_0-(1-q_1)}\right)^2 & \text{if } 1-q_1 \leq 1-q < \frac{2-q_0-q_1}{2} \\ 1-2\left(\frac{1-q_0-(1-q)}{1-q_0-(1-q_1)}\right)^2 & \text{if } \frac{2-q_0-q_1}{2} \leq 1-q < 1-q_0 \end{cases}$	$u_0(1-q) = \begin{cases} 1-2\left(\frac{1-q-(1-q_0)}{1-q_1-(1-q_0)}\right)^2 & \text{if } 1-q_0 < 1-q \leq \frac{2-q_0-q_1}{2} \\ 2\left(\frac{1-q_1-(1-q)}{1-q_1-(1-q_0)}\right)^2 & \text{if } \frac{2-q_0-q_1}{2} < 1-q \leq 1-q_1 \end{cases}$
	$u_1(1-q) = \begin{cases} 1-2\left(\frac{1-q-(1-q_1)}{1-q_0-(1-q_1)}\right)^2 & \text{if } 1-q_1 \leq 1-q < \frac{2-q_0-q_1}{2} \\ 2\left(\frac{1-q_0-(1-q)}{1-q_0-(1-q_1)}\right)^2 & \text{if } \frac{2-q_0-q_1}{2} \leq 1-q < 1-q_0 \end{cases}$	$u_1(1-q) = \begin{cases} 2\left(\frac{1-q-(1-q_0)}{1-q_1-(1-q_0)}\right)^2 & \text{if } 1-q_0 < 1-q \leq \frac{2-q_0-q_1}{2} \\ 1-2\left(\frac{1-q_1-(1-q)}{1-q_1-(1-q_0)}\right)^2 & \text{if } \frac{2-q_0-q_1}{2} < 1-q \leq 1-q_1 \end{cases}$
Convex	$u_0(1-q) = \exp\left(\frac{-6(1-q_0-(1-q))}{1-q_0-(1-q_1)}\right)$ if $1-q_1 \leq 1-q < 1-q_0$	$u_0(1-q) = \exp\left(\frac{-6(1-q-(1-q_0))}{1-q_1-(1-q_0)}\right)$ if $1-q_0 < 1-q \leq 1-q_1$
	$u_1(1-q) = \exp\left(\frac{-6(1-q-(1-q_1))}{1-q_0-(1-q_1)}\right)$ if $1-q_1 \leq 1-q < 1-q_0$	$u_1(1-q) = \exp\left(\frac{-6(1-q_1-(1-q))}{1-q_1-(1-q_0)}\right)$ if $1-q_0 < 1-q \leq 1-q_1$
Concave	$u_0(1-q) = 1-\exp\left(\frac{-6(1-q-(1-q_1))}{1-q_0-(1-q_1)}\right)$ if $1-q_1 \leq 1-q < 1-q_0$	$u_0(1-q) = 1-\exp\left(\frac{-6(1-q_1-(1-q))}{1-q_1-(1-q_0)}\right)$ if $1-q_0 < 1-q \leq 1-q_1$
	$u_1(1-q) = 1-\exp\left(\frac{-6(1-q_0-(1-q))}{1-q_0-(1-q_1)}\right)$ if $1-q_1 \leq 1-q < 1-q_0$	$u_1(1-q) = 1-\exp\left(\frac{-6(1-q-(1-q_0))}{1-q_1-(1-q_0)}\right)$ if $1-q_0 < 1-q \leq 1-q_1$

In addition, analogously to the crisp case, the representations (2) and (3) can be applied to the corresponding representations of the generalized "left-tailed" p-value, that is,

$$p_l(n, s^{(+)}) = \sup_{1-q \in H_0} \{(m_{H_0}(1-q) - m_{H_1}(1-q))\mathbb{P}_{1-q \in H_0}(S^{(+)} \leq s^{(+)})\} \quad \text{(preliminary left-tailed test)} \tag{6}$$

$$p_l(n, s^{(-)}) = \sup_{q \in H_0} \{(m_{H_0}(q) - m_{H_1}(q))\mathbb{P}_{q \in H_0}(S^{(-)} \leq s^{(-)})\} \quad \text{(preliminary right-tailed test)} \tag{7}$$

5 | FUZZY CATEGORIES

It is well-known that the sign of a single observation of a continuous variable X_i, $i = 1, \ldots, n$, with respect to a given hypothesized quantile value k admits two possibilities:

$$\text{negative } (X_i - k < 0) \text{ and positive } (X_i - k > 0)$$

These both possibilities constitute two respective categories: "negative signs" ($C^{(-)}$) and "positive signs" ($C^{(+)}$), which are essential to construct the test statistic in the framework of the sign test. In particular, the test statistic considers only one category, exemplary "negative signs," and thus provides the absolute frequency (number) of negative signs for a given sample.

Further, a disjunctive coding by 0/1 of a continuous variable X_i allows its convenient treatment as a categorical variable with the following indicator functions under H_0:

$$m_{C^{(-)}}(X_i) = \begin{cases} 1 & \text{if } X_i < k \\ 0 & \text{if } X_i > k \end{cases} \quad \text{and} \quad m_{C^{(+)}}(X_i) = \begin{cases} 0 & \text{if } X_i < k \\ 1 & \text{if } X_i > k \end{cases}$$

However, such a disjunctive coding leads to an exclusive choice of a category, which may seem to be quite natural at first sight, but has two serious drawbacks for the sign test in practice:

- loss of information regarding the magnitude of sample observations $x_1, \ldots, x_n$ (i.e., their ranking);
- generation of discontinuities at the border point k of both categories, that is, null differences (ties).

As for the handling of ties ($X_i = k$), their classification to both categories in an equal magnitude of 0.5 is recommended (see Dixon and Mood[15]). Such a neutral handling of ties is more promising than a reduction of the sample size due to the elimination of ties or classification to the under-represented category. However, such an approach can not resolve the general problem of missing ranking of observations especially near the border point of both categories.

Consequently, a disjunctive coding (exclusive categories) is too restrictive and should be replaced by a more appropriate one, like a fuzzy coding (i.e., fuzzy categories "rather positive signs" and "rather negative signs"). In particular, using fuzzy instead of interval partition (see also Figure 6), fuzzy coding allows to specify a degree of conviction of a sample observation for each category, and simultaneously keeps the advantage of treatment as a categorical variable X_i with membership functions $m_{C^{(-)}}(X_i), m_{C^{(+)}}(X_i) \in [0, 1]$ instead of the underlying continuous variable $X_i \in \mathbb{R}$.

Having two fuzzy categories $(C^{(-)}, C^{(+)})$ with complementary membership functions (as in Figure 6) or noncomplementary membership functions, that is, $m_{C^{(-)}}(X_i) = 1 - m_{C^{(+)}}(X_i)$ or $m_{C^{(-)}}(X_i) \leq 1 - m_{C^{(+)}}(X_i)$, the first (second) category $C^{(-)}$ $(C^{(+)})$ may be identified with success exemplary if one of the following options occurs:

1. $m_{C^{(-)}}(X_i) \in [0.5, 1]$ $(m_{C^{(+)}}(X_i) \in [0.5, 1])$.
2. $m_{C^{(-)}}(X_i) \in (0, 1]$ $(m_{C^{(+)}}(X_i) \in (0, 1])$.

Then it can be implemented into the test statistic defined as the number of successes in the sample. Given complementary membership functions, option 1. provides a gradual consideration regarding category $C^{(-)}$ $(C^{(+)})$ to the left (right) of the hypothesized quantile value k, while option 2. allows to implement a degree of conviction for fuzzy category $C^{(-)}$ $(C^{(+)})$ in the whole border zone around k, that is between x_l and x_r. In fact, there is no justification to neither prefer the items X_i with membership degree ≥ 0.5 nor discriminate the items X_i with membership degree <0.5 regarding the category of interest, so that both options appear appropriate.

In the following, we provide a definition of the test statistic S using both options for right- and left-tailed considerations in combination with $C^{(-)}$ and $C^{(+)}$, respectively, that is,

$$\begin{aligned} S_1^{(-)} &= \sum_{i=1}^{n} m_{C^{(-)}}(X_i) \text{ with } X_i \in {}_{0.5}C^{(-)} \quad \text{vs} \\ S_1^{(+)} &= \sum_{i=1}^{n} m_{C^{(+)}}(X_i) \text{ with } X_i \in {}_{0.5}C^{(+)} \quad (1.\ \text{ option}) \end{aligned} \tag{8}$$

$$\begin{aligned} S_2^{(-)} &= \sum_{i=1}^{n} m_{C^{(-)}}(X_i) \text{ with } X_i \in C^{(-)} \quad \text{vs} \\ S_2^{(+)} &= \sum_{i=1}^{n} m_{C^{(+)}}(X_i) \text{ with } X_i \in C^{(+)} \quad (2.\ \text{ option}) \end{aligned} \tag{9}$$

where $C^{(-)} := \{(X_i; m_{C^{(-)}}(X_i))|X_i \in \mathbb{R}\}$ and $C^{(+)} := \{(X_i; m_{C^{(+)}}(X_i))|X_i \in \mathbb{R}\}$ are the fuzzy sets regarding the category "(rather) negative signs" and "(rather) positive signs" as well as

$$\begin{aligned} {}_{0.5}C^{(-)} &= \left\{\left(X_i; \min\left\{m_{C_{0.5}^{(-)}}(X_i), m_{C^{(-)}}(X_i)\right\}\right)|X_i \in \mathbb{R}\right\} \quad \text{and} \\ {}_{0.5}C^{(+)} &= \left\{\left(X_i; \min\left\{m_{C_{0.5}^{(+)}}(X_i), m_{C^{(+)}}(X_i)\right\}\right)|X_i \in \mathbb{R}\right\} \end{aligned}$$

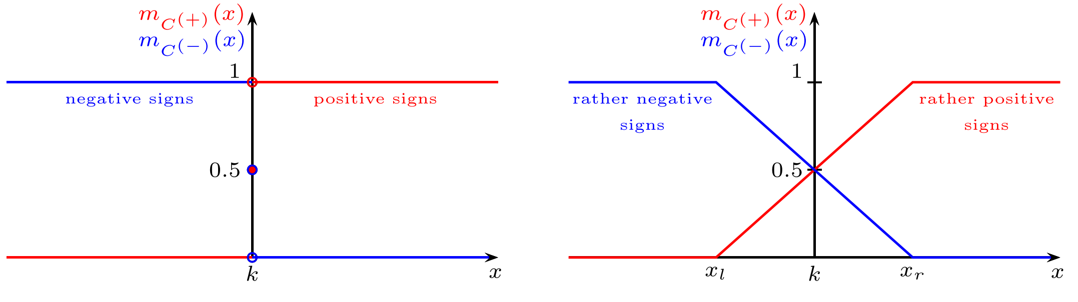

FIGURE 6 Disjunctive and fuzzy coding of a continuous variable X_i [Color figure can be viewed at wileyonlinelibrary.com]

are the fuzzy sets based on the 0.5-cut of $C^{(-)}$ and $C^{(+)}$, respectively. Modeling of $m_{C^{(-)}}$ and $m_{C^{(+)}}$ can be proposed exemplary with help of piecewise linear complementary membership functions:

$$m_{C^{(-)}}(X_i) = \begin{cases} 1 & \text{if } X_i < x_l \\ \dfrac{x_r - X_i}{x_r - x_l} & \text{if } x_l \leq X_i \leq x_r \\ 0 & \text{if } X_i > x_r \end{cases} \quad m_{C^{(+)}}(X_i) = \begin{cases} 0 & \text{if } X_i < x_l \\ \dfrac{X_i - x_l}{x_r - x_l} & \text{if } x_l \leq X_i \leq x_r \\ 1 & \text{if } X_i > x_r \end{cases}$$

It should be noted that the provided example of membership functions is not the only possible and reasonable alternative and can be adjusted to specific application fields, see Chukhrova and Johannssen,[20] for a detailed discussion on this topic. As for the specification of the target values x_l and x_r, there are a lot of possibilities, for instance via symmetric modeling with respect to a given hypothesized quantile value k, which seems to be a natural one and in compliance with ordering on the real line. In contrast, an asymmetric modeling of target values x_l, x_r could provide a more appropriate mapping of grades of preference (based e.g., on cost aspects) attributed to the respective category.

As for a symmetric modeling, we introduce an approach using the median absolute deviation (MAD) of individual observations X_i, $i = 1, \ldots, n$, from the hypothesized quantile value k, that is,

$$\text{MAD} = \text{median}\{|X_1 - k|, \ldots, |X_n - k|\}$$

for specification of x_l and x_r as $x_l = k - \text{MAD}$ and $x_r = k + \text{MAD}$, respectively. The MAD is a robust measure of dispersion being more resilient to outliers in the data compared to the standard deviation (see Pham-Gia and Hung,[27] for further properties of the MAD).

It should be emphasized that fuzzy categories lead to a crisp test statistic (S is a crisp number with $S \in \mathbb{R}_+$ instead of $S \in \mathbb{N}_0$ as in the crisp categories case), independently of specified membership functions. This important benefit provides the possibility of a direct and natural implementation of fuzzy categories into the classical sign test or into the generalized sign test with fuzzy hypotheses introduced in Section 4.

In the following, we propose to link the specific formulation of the preliminary null hypothesis, that is,

(a) H_0': $k_q \leq k$ (there are no more positive than negative signs),
(b) H_0': $k_q \geq k$ (there are no more negative than positive signs),

to the specific formulation of the test statistic, which considers the category of interest enclosed in H_0', that is,

(a) $S^{(-)}$ regarding "rather negative signs",
(b) $S^{(+)}$ regarding "rather positive signs".

Such a formulation of the test statistic allows to aggregate fuzzy coding into hypothesis testing in a more compatible way and always leads to left-tailed modeling of reformulated hypotheses H_0 and H_1, and thus to the respective p-value ($p_l(n, s^{(+)})$, $p_l(n, s^{(-)})$, see 2, 3, 6, and 7) of a left-tailed binomial test. While the right-tailed sign test (based on the test statistic regarding "(rather) negative signs") is discussed in Sections 3.1 and 4.1, the left-tailed test (based on the test statistic regarding "(rather) positive signs") can be found in Sections 3.2 and 4.2.

Applying the 1. option (or crisp categories), the implementation of a real-valued test statistic into the binomial probability function for calculating the p-value can be obtained via gamma function Γ (that interpolates the factorial function to noninteger values) instead of using binomial coefficients (that are limited to positive integers), that is,

$$p_l\left(n, s_1^{(-)}\right) = \mathbb{P}_{q \in H_0}\left(S_1^{(-)} \leq s_1^{(-)}\right) = \sum_{m=0}^{s_1^{(-)}} \frac{\Gamma(n+1)}{\Gamma(m+1)\Gamma(n-m+1)} q^m (1-q)^{n-m} \tag{10}$$

$$p_l\left(n, s_1^{(+)}\right) = \mathbb{P}_{1-q \in H_0}\left(S_1^{(+)} \leq s_1^{(+)}\right) = \sum_{m=0}^{s_1^{(+)}} \frac{\Gamma(n+1)}{\Gamma(m+1)\Gamma(n-m+1)} (1-q)^m q^{n-m} \tag{11}$$

with

$$\Gamma(z+1) = \int_0^{\infty} t^z e^{-t}\, \mathrm{d}t, \quad z \in \mathbb{R}_+$$

for right- and left-tailed preliminary considerations, respectively. These definitions are valid when formulating crisp hypotheses, while the generalized p-values (6) and (7) are to adjust in an analogous way when formulating fuzzy hypotheses.

Considering both options (8) and (9) for the formulation of the test statistic, it holds that $S_1^{(-)} \leq S_2^{(-)}$, since the support of ${}_{0.5}C^{(-)}$ regarding $S_1^{(-)}$ contains the same items either in the crisp or fuzzy categories case, that is, $\{X_i | X_i \leq k\}$, while the support of $C^{(-)}$ regarding $S_2^{(-)}$ differs, that is,

$$\{X_i | X_i < x_r\}\ (\text{fuzzy categories case}) \quad \text{versus} \quad \{X_i | X_i \leq k\}\ (\text{crisp categories case})$$

Therefore, this deviation should be appropriately compensated when using the 2. option. Such a compensation can be provided via adjustment of the sample size n by adding a deviant number of items, say

$$\mathrm{a}^{(-)} = |\{X_i | k < X_i < x_r\}|,$$

where $\mathrm{a}^{(-)} \in \mathbb{N}_0$. This modeling is based on the maximization of the conviction approach, that is, full consideration of an additional item with characteristic $X_i \in (k, x_r)$ in the test statistic. It should be noted that this approach ensures the appropriate adjustment of the number of items $X_i, i = 1, \ldots, n$, which is crucial for the calculation of the test statistic, independently of the type (crisp/fuzzy) and the sign $(-/+)$ of the category of interest. On the one hand, if we deal with the crisp category "negative (positive) signs", we consider solely elements X_i at and on the left (right) of the hypothesized quantile value k, say

$$n^{(-)} = n^{(-)}_{X_i<k} + n^{(-)}_{X_i=k} + \underbrace{n^{(-)}_{X_i>k}}_{=0} \quad \left(n^{(+)} = \underbrace{n^{(+)}_{X_i<k}}_{=0} + n^{(+)}_{X_i=k} + n^{(+)}_{X_i>k} \right)$$

with $n^{(-)}_{X_i=k} = n^{(+)}_{X_i=k}$. The same is valid when formulating the fuzzy category "rather negative (positive) signs" and using the 1. option. On the other hand, if we deal with the 2. option, we additionally consider the elements X_i between k and x_r (x_l and k), that is, the number $n^{(-)}_{X_i>k} \geq 0$ ($n^{(+)}_{X_i<k} \geq 0$) of $X_i, i = 1, \ldots, n$, is a so called adjustment term, which ensures the implementation of additional elements into the test statistic S_2 and thus constitutes the definition of $\mathrm{a}^{(-)}$ ($\mathrm{a}^{(+)}$), where $\mathrm{a}^{(+)} = |\{X_i | x_l < X_i < k\}|$, $\mathrm{a}^{(+)} \in \mathbb{N}_0$.

In addition, the above presented possibility to calculate the p-value using gamma function can be adjusted for the 2. option by implementing of $\mathrm{a}^{(-)}$ and $\mathrm{a}^{(+)}$,

$$p_l(n + \mathrm{a}^{(-)}, s^{(-)}) = \mathbb{P}_{q \in H_0}\left(S_2^{(-)} \leq s_2^{(-)}\right) = \sum_{m=0}^{s_2^{(-)}} \frac{\Gamma(n + \mathrm{a}^{(-)} + 1)}{\Gamma(m + 1)\Gamma(n + \mathrm{a}^{(-)} - m + 1)} q^m (1 - q)^{n+\mathrm{a}^{(-)}-m} \tag{12}$$

$$p_l(n + \mathrm{a}^{(+)}, s^{(+)}) = \mathbb{P}_{1-q \in H_0}\left(S_2^{(+)} \leq s_2^{(+)}\right) = \sum_{m=0}^{s_2^{(+)}} \frac{\Gamma(n + \mathrm{a}^{(+)} + 1)}{\Gamma(m + 1)\Gamma(n + \mathrm{a}^{(+)} - m + 1)} (1 - q)^m q^{n+\mathrm{a}^{(+)}-m} \tag{13}$$

for right- and left-tailed preliminary considerations, respectively. These definitions are valid when formulating crisp hypotheses, while the generalized p-values (6) and (7) are to adjust in an analogous way when formulating fuzzy hypotheses.

6 | CASE STUDY: COVID-19 IN HIV-INFECTED INDIVIDUALS

In this section, we present a case study considering a real data set on COVID-19 in HIV-infected individuals to emphasize the benefits of the proposed methodology in practical applications.

6.1 | The data set

One week after the outbreak of pneumonia of unknown origin in Wuhan, China, on December 31, 2019, its cause has been identified as severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2). On March 11, 2020, due to a continuously increasing number of COVID-19 cases worldwide, the World Health Organization (WHO) declared a pandemic. Spain, and especially the Community of Madrid, has been one of the most affected countries worldwide with 203,715 confirmed cases as of April 30, 2020 (see Vizcarra et al.[28]).

The data set underlying this case study is taken from Vizcarra et al.[28] and includes data about consecutive HIV-infected individuals aged 18 years or older with a diagnosis of suspected or confirmed COVID-19 (as of April 30, 2020). These data result from an observational prospective study at the Hospital Universitario Ramon y Cajal, a tertiary university hospital with 1100 beds and 2873 adult patients with HIV on regular follow-up at the monographic HIV clinics. In particular, there are $n = 51$ patients diagnosed with COVID-19 among the 2873 HIV-infected individuals (i.e., a COVID-19 infection rate of 1.78%). The authors of the study investigated whether there is an increased risk (e.g., due to comorbidities, lower CD4 cell counts, or unsuppressed HIV RNA viral load) or a decreased risk (e.g., due to immunosuppression or regular use of antiretrovirals) in HIV-infected individuals of SARS-CoV-2 infection.

6.2 | The role of human body temperature and related measurement problems

One of the most common characteristics of HIV and COVID-19 is an increased body temperature (hyperthermia). It should be noted that hyperthermia is the most frequent symptom of COVID-19 (besides dry cough and dyspnoea), while it is an accompanying symptom of HIV. The variable "human body temperature" is a continuous one, but commonly treated as categorical variable in medical context and usually specified in an interval-valued way. Considering babies and children and adults, the average body temperature ranges from 36.6°C (97.9°F) to 37.2°C (99°F) and from 36.1°C (97°F) to 37.2°C (99°F), respectively. In addition, the following classification for human body temperature is often used for adults (see e.g., Hutchison et al.,[29] Laupland,[30] and Grunau et al.[31]):

Hypothermia	$<35.0^\circ\mathrm{C}$ ($<95.0^\circ\mathrm{F}$)
Normal	$36.5-37.5^\circ\mathrm{C}$ ($97.7-99.5^\circ\mathrm{F}$)
Hyperthermia	$37.5-38.0^\circ\mathrm{C}$ ($99.5-100.4^\circ\mathrm{F}$)
Fever	$38.0-39.0^\circ\mathrm{C}$ ($100.4-102.2^\circ\mathrm{F}$)
High fever	$39.0-40.0^\circ\mathrm{C}$ ($102.2-104.0^\circ\mathrm{F}$)
Hyperpyrexia	$>40.0^\circ\mathrm{C}$ ($>104.0^\circ\mathrm{F}$)

Note that the intervals are right half-open, and $x^\circ\mathrm{C}$ can be easily converted into $y^\circ\mathrm{F}$ by using $y = \frac{9}{5}x + 32$.

In practice the body temperature is measured real-valued. Depending on the problem, the measured value is often compared with either the lower bound, midpoint or upper bound of the corresponding category to interpret the measured value. This approach is simple, but has severe disadvantages: On the one hand, measurements of body temperature are generally imprecise. Human body temperature varies depending on age, emotions, exertion level (according to activities), health status (e.g., illness, menstruation), sex, state of consciousness (e.g., waking, sleeping, sedated), time of day (due to a person's circadian rhythm), type of medical thermometer, and what part of the body the measurement is taken at (e.g., oral, rectal, axillary, tympanic) (see e.g., Moran and Mendal[32]). Therefore, each type of measurement has a specific range of temperatures. On the other hand, temperature fluctuation within a category and membership of boundary values between categories are not adequately represented. For instance, if a temperature of $37.49^\circ\mathrm{C}$ is measured by means of a digital thermometer, the value is rounded up to $37.5^\circ\mathrm{C}$ and falls into the category of hyperthermia, although it is actually more likely to be within a gray zone between the categories "normal" and "hyperthermia."

These measurement problems could be addressed by replacing crisp categories by fuzzy categories for human body temperature X. With regard to the frequently observed symptom of hyperthermia in HIV and COVID-19 infections, that is, $X \geq 37.5^\circ\mathrm{C}$ and $k = 37.5$, two exemplary crisp categories

Category I (negative signs regarding k): nonincreased body temperature ($X < 37.5^\circ\mathrm{C}$);
Category II (positive signs regarding k): increased body temperature ($X > 37.5^\circ\mathrm{C}$);

with indicator functions

$$m_{C^{(-)}}(X) = \begin{cases} 1 & \text{if } X < 37.5 \\ 0 & \text{if } X > 37.5 \end{cases} \quad m_{C^{(+)}}(X) = \begin{cases} 0 & \text{if } X < 37.5 \\ 1 & \text{if } X > 37.5 \end{cases} \tag{14}$$

and $m_{C^{(-)}}(X) = m_{C^{(+)}}(X) = 0.5$ for $X = 37.5$ (due to an indifference argument) could be replaced by fuzzy categories:

Category I (rather negative signs regarding k): rather nonincreased body temperature ($X \leq 36.9^\circ\mathrm{C}, X \boxed{\leq} 38.1^\circ\mathrm{C}$).

Category II (rather positive signs regarding k): rather increased body temperature ($X \boxed{\geq} 36.9°$C, $X \geq 38.1$°C).

with piecewise linear membership functions (see also Figure 7)

$$m_{C^{(-)}}(X) = \begin{cases} 1 & \text{if } X \leq 36.9 \\ \dfrac{38.1 - X}{38.1 - 36.9} & \text{if } 36.9 < X < 38.1 \\ 0 & \text{if } X \geq 38.1 \end{cases}$$

$$m_{C^{(+)}}(X) = \begin{cases} 0 & \text{if } X \leq 36.9 \\ \dfrac{X - 36.9}{38.1 - 36.9} & \text{if } 36.9 < X < 38.1 \\ 1 & \text{if } X \geq 38.1 \end{cases} \tag{15}$$

where $x_l = 36.9$ and $x_r = 38.1$ are calculated with help of MAD $= 0.6$.

6.3 | Performing the sign test with crisp hypotheses and crisp versus fuzzy categories

In the first step, we consider the conventional sign test with crisp hypotheses and compare the impact of crisp versus fuzzy categories on the test decision. Let us assume that it is our goal to check whether the median of human body temperature regarding COVID-19- and HIV-infected individuals is significantly above 37.5°C . Then we can formulate the following preliminary and reformulated hypotheses:

$$H_0' : k_q \leq k \quad \text{versus} \quad H_1' : k_q > k \quad \text{with } k = 37.5$$
$$H_0 : q \geq q_0 \quad \text{versus} \quad H_1 : q < q_0 \quad \text{with } q_0 = 0.5$$

Since the test performance generally improves with an increasing value of n, we do not limit our considerations to a single sample size, but perform an additional sensitivity analysis regarding various values of n. In particular, we choose $n \geq 10$ due to the numerical experience regarding the performance of the sign test and successively specify n as $n = 10, 11, ..., 51$.

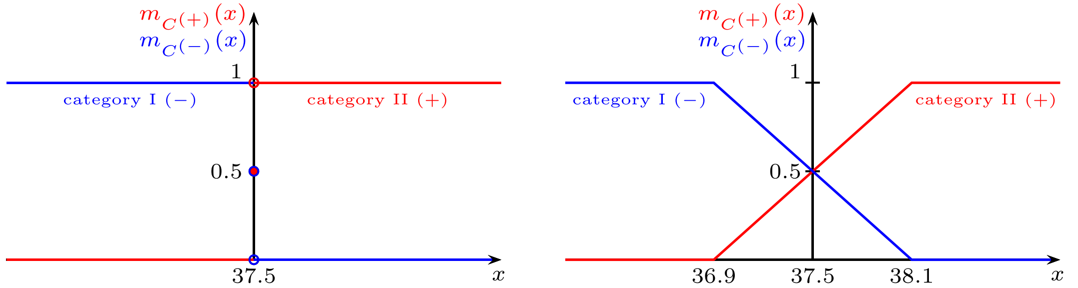

FIGURE 7 Crisp and fuzzy categories for human body temperature [Color figure can be viewed at wileyonlinelibrary.com]

Considering random sampling and depending on the specified sample size, we use the respective segment of the data set of 51 HIV-infected individuals with diagnosis of COVID-19, that is, $n = 10, 11, \ldots, 51$ items of the data set.

The data set (see Table 10, columns 1–2) contains eight observations that are equal to the hypothesized quantile value $k = 37.5$, that is, we are confronted with the problem of ties. To soften the difficulties in handling ties when employing the common sign test, we propose to classify them to both categories (negative/positive signs) with an equal magnitude of 0.5. In particular, we distinguish three cases concerning the categories:

- *Case* 1: crisp categories
- *Case* 2: fuzzy categories, piecewise linear membership functions, membership values ≥0.5 (1. option, see Section 5)
- *Case* 3: fuzzy categories, piecewise linear membership functions, adjusted n (2. option, see Section 5)

To clarify the procedure and to simplify the interpretation of the results given in Table 10, we also present examples on how to calculate the test statistic $s^{(-)}$ and the corresponding p-value for both crisp and fuzzy categories.

Example 6.1 (Calculation of test statistic and p-value in the crisp categories case). As for the determination of the test statistic $s^{(-)}$ and the corresponding p-value (3) in the crisp categories case (Case 1), let us calculate these values exemplary for $n = 25$. Thus, we have

$$s^{(-)} = \sum_{i=1}^{25} m_{C^{(-)}}(x_i) = 1 + 0 + 0.5 + \cdots + 0.5 + 0 + 0 = 9.5$$

(see Table 9, row 4) with $m_{C^{(-)}}(X_i)$ determined by (14) as well as

$$p_l(n, s^{(-)}) = \mathbb{P}_{q\in H_0}(S^{(-)} \leq 9.5) = 0.5^{25} \sum_{m=0}^{9.5} \frac{\Gamma(26)}{\Gamma(m+1)\Gamma(26-m)} = 0.1934.$$

Example 6.2 (Calculation of test statistic and p-value in the fuzzy categories case). Considering fuzzy categories, we apply both options (discussed in Section 5) to determine the test statistic $s^{(-)}$ and the p-value, that is, we implement the above Cases 2 and 3. In the first step, we calculate MAD for $n = 25$ as median $\{1.3, 0.3, \ldots, 1.9, 0.3\} = 0.6$ (see Table 9, row 3). Thus, using $x_l = k - \text{MAD} = 37.5 - 0.6 = 36.9$ and $x_r = k + \text{MAD} = 37.5 + 0.6 = 38.1$, we obtain piecewise linear membership functions for fuzzy categories (see (15) and Figure 7). The test statistic $s^{(-)}$ is then calculated by using either the whole fuzzy set $C^{(-)}$ (2. option) or its subset ${}_{0.5}C^{(-)}$ based on the 0.5-cut (1. option). The respective membership values of sample items with respect to both options are given in Table 9, rows 5–6. The test statistic $s^{(-)}$ and the corresponding p-value for both cases (see 10 and 12) are calculated as follows:

TABLE 9 Membership values when considering crisp categories (Case 1) and fuzzy categories (Cases 2/3)

	i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
	x_i	36.2	37.8	37.5	39.5	38	38.1	36.7	39.4	35.5	37.8	38
	$\lvert x_i - k \rvert$	1.3	0.3	0	2	0.5	0.6	0.8	1.9	2	0.3	0.5
Case 1	$m_{C^{(-)}}(x_i)$	1	0	0.5	0	0	0	1	0	1	0	0
Case 2	$m_{{}_{0.5}C^{(-)}}(x_i)$	1	0	0.5	0	0	0	1	0	1	0	0
Case 3	$m_{C^{(-)}}(x_i)$	1	0.25	0.5	0	0.0833	0	1	0	1	0.25	0.0833

	i	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
	x_i	36.7	39	38	38.3	37.3	37	39.5	38.5	37.3	37.5	35.7	37.5	39.4	37.8
	$\lvert x_i - k \rvert$	0.8	1.5	0.5	0.8	0.2	0.5	2	1	0.2	0	1.8	0	1.9	0.3
Case 1	$m_{C^{(-)}}(x_i)$	1	0	0	0	1	1	0	0	1	0.5	1	0.5	0	0
Case 2	$m_{{}_{0.5}C^{(-)}}(x_i)$	1	0	0	0	0.6667	0.9167	0	0	0.6667	0.5	1	0.5	0	0
Case 3	$m_{C^{(-)}}(x_i)$	1	0	0.0833	0	0.6667	0.9167	0	0	0.6667	0.5	1	0.5	0	0.25

$$s^{(-)} = \sum_{i=1}^{25} m_{{}_{0.5}C^{(-)}}(x_i) = 1 + 0 + 0.5 + \cdots + 0.5 + 0 + 0 = 8.75 \qquad (\text{Case 2})$$

$$s^{(-)} = \sum_{i=1}^{25} m_{C^{(-)}}(x_i) = 1 + 0.25 + 0.5 + \cdots + 0.5 + 0 + 0.25 = 9.75 \quad (\text{Case 3})$$

$$p_l(n, s^{(-)}) = \mathbb{P}_{q \in H_0}(S^{(-)} \leq 8.75) = 0.5^{25} \sum_{m=0}^{8.75} \frac{\Gamma(26)}{\Gamma(m+1)\Gamma(26-m)} = 0.1066 \, (\text{Case 2})$$

$$p_l(n + a^{(-)}, s^{(-)}) = \mathbb{P}_{q \in H_0}(S^{(-)} \leq 9.75) = 0.5^{31} \sum_{m=0}^{9.75} \frac{\Gamma(32)}{\Gamma(m+1)\Gamma(32-m)} = 0.0319 \, (\text{Case 3})$$

Note that in Case 3, we have $\mathrm{a}^{(-)} = 6$ and thus $n + \mathrm{a}^{(-)} = 31$ since there are six additional membership values below 0.5 (which add up to one, see 8.75 vs. 9.75) compared to Case 2.

We can resume as follows: While the p-values in Case 1 (see Example 6.1) and Case 2 are above commonly used significance levels ($\mathbb{P}_{q \in H_0}(S^{(-)} \leq s^{(-)}) > 10\%$), the p-value in Case 3 is below the 5%-level, that is, $H_0 : q \geq 0.5$ can be rejected only in Case 3 and we can conclude that human body temperature is significantly above 37.5°C at the 5% level of significance. That is, we are confronted with hyperthermia or worse.

We refer to Table 10, columns 4–10, for complete results in the crisp hypotheses case. Although we are only interested in results for $n \geq 10$, the results for lower values of n are also given for the sake of completeness. Nonetheless, the following interpretations regularly refer to $n \geq 10$.

Considering the empirical median (Table 10, column 3) and the MAD (Table 10, column 4), we also refer to Figure 8, where the empirical median $\widehat{k}_{0.5}$, the hypothesized quantile/median value k and $x_l = k - \text{MAD}$, $x_r = k + \text{MAD}$ are plotted as functions of the sample size n. It can be seen that the empirical median is mostly equal to 37.8 and the MAD lies between 0.6 and 0.8 (compare also $\widehat{k}_{0.5} = 37.7$ and $\text{MAD} = 0.7$ for the entire data set with $n = 51$). Note that the numerical results are displayed as line plots rather than scatter plots for better output illustration.

As for varying values of n, columns 5–7 and 8–10 of Table 10 contain the test statistics $s^{(-)}$ and p-values regarding crisp hypotheses for three cases, respectively, each calculated for $n = 1, \ldots, 51$. Considering the test statistic $s^{(-)}$, we observe the following order: $s^{(-)}_{\text{Case 2}} \leq s^{(-)}_{\text{Case 1}} < s^{(-)}_{\text{Case 3}}$. Analyzing and comparing p-values lead to the implications that Case 1 (crisp categories) generally leads to the highest p-values, and p-values are considerably lower in Case 3 compared to Case 2 (both fuzzy categories). As a rule, the larger n, the smaller the p-value, and moreover, it can be seen that p-values are mostly below 5% as of $n = 14$ (Case 3) and below 10% as of $n = 30$ (Case 2). Considering Case 1, there is only one p-value below 10% (for $n = 37$). Overall, there are no p-values below 1% in the crisp hypotheses case. This behavior can also be seen in Figure 9A.

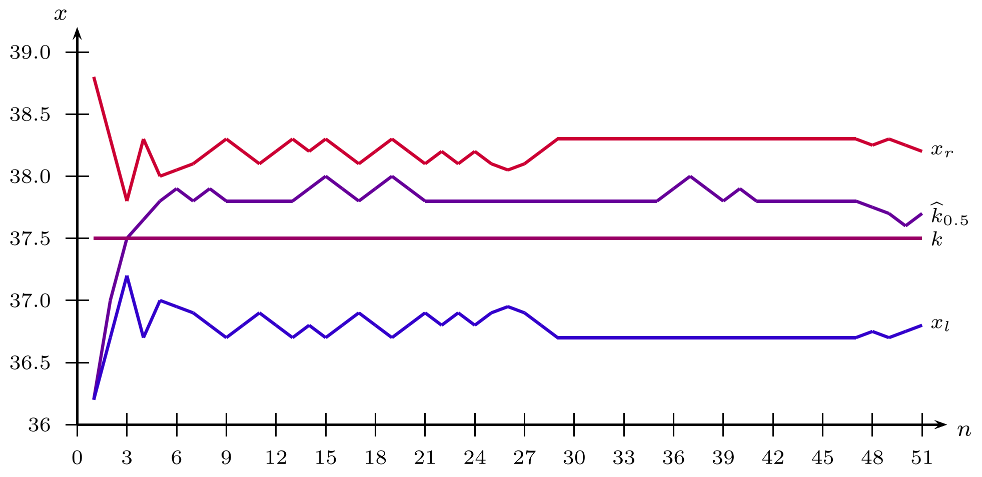

FIGURE 8 The behavior of the empirical median and $x_l = k - \text{MAD}$, $x_r = k + \text{MAD}$ for increasing values of n [Color figure can be viewed at wileyonlinelibrary.com]

6.4 | Performing the sign test with fuzzy hypotheses and crisp versus fuzzy categories

In this section, we consider fuzzy reformulated hypotheses H_0 and H_1 using crisp and fuzzy categories. In the first step, let us again assume that we would like to check whether the median of human body temperature regarding COVID-19- and HIV-infected individuals is significantly above 37.5°C (right-tailed sign test). However, instead of crisp specified 50%-fraction, we use fuzzy statements on the unknown fraction q, that is, $\mathbb{P}(X \leq 37.5) \approx \mathbb{P}(X \geq 37.5) \approx 50\%$. In particular, the fuzzy reformulated hypotheses can be stated as follows:

$$H_0 : q \boxed{\geq} q_1, q \geq q_0 \quad \text{versus} \quad H_1 : q < q_1, q \boxed{<} q_0 \quad \text{with } 0 < q_1 < q_0 < 1$$

Since we perform a test for the median, that is, the 50%-quantile, it is appropriate to construct symmetric membership functions around the value 0.5: $|q_0 - 0.5| = |q_1 - 0.5|$, $m_{H_1}(q) = 1 - m_{H_0}(q)$. Let us additionally assume that the hypothesized median value 37.5 deviates from the true median by at most 2.5%, that is, we specify $q_0 = 0.525$ and $q_1 = 0.475$. Linguistic interpretations of H_0 and H_1 are then given by:

- H_0: The population proportion is approximately larger than or equal to 0.5, in particular fuzzy (crisp) larger than or equal to 0.475 (0.525).
- H_1: The population proportion is approximately lower than 0.5, in particular fuzzy (crisp) lower than 0.525 (0.475).

We propose the usage of piecewise polynomial (s-shaped) functions for modeling the fuzzy areas of both hypotheses, that is (q_1, q_0), within the membership functions (see Section 4). Considering $q_0 = 0.525$ and $q_1 = 0.475$ as before, $m_{H_0}(q)$ and $m_{H_1}(q)$ are defined by

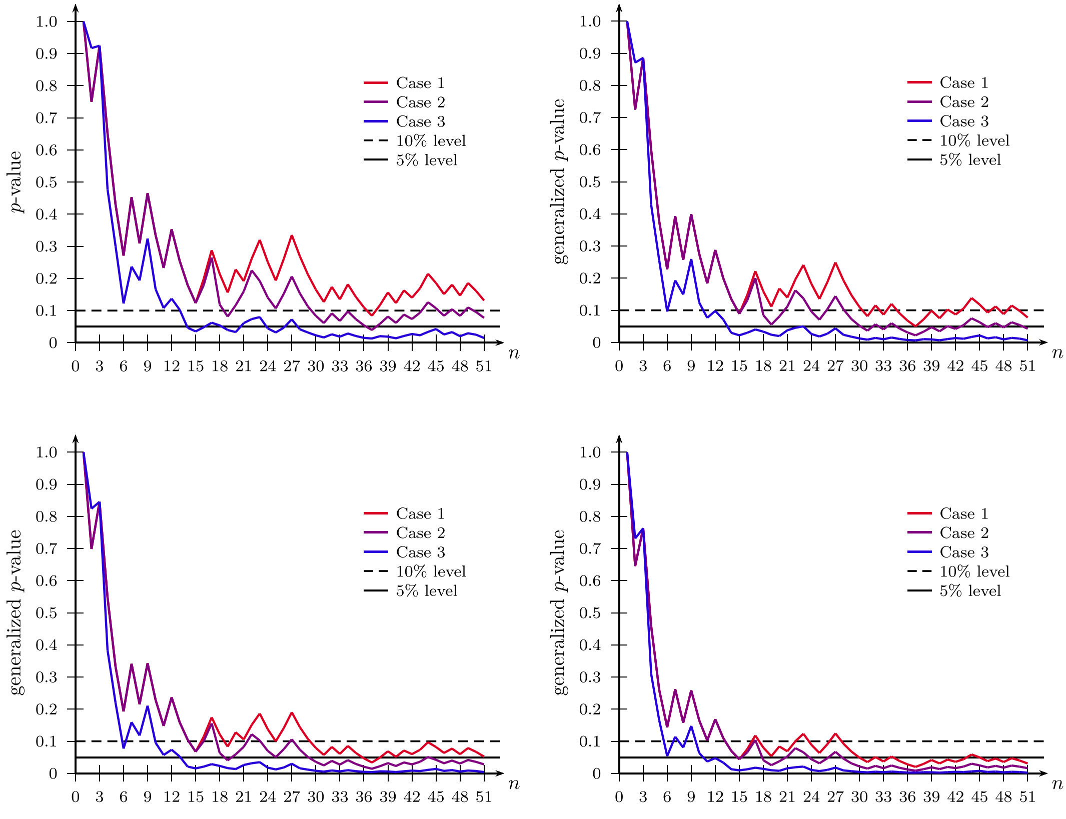

FIGURE 9 The behavior of (generalized) p-values for increasing values of n, each considering four cases. (A) Crisp hypotheses, (B) fuzzy hypotheses with $w = 0.05$, (C) fuzzy hypotheses with $w = 0.1$, (D) fuzzy hypotheses with $w = 0.2$ (from top left to bottom right) [Color figure can be viewed at wileyonlinelibrary.com]

$$m_{H_0}(q) = \begin{cases} 1 & \text{if } 0 < q \leq q_1 \\ 2\left(\dfrac{q - q_1}{q_0 - q_1}\right)^2 & \text{if } q_1 < q \leq \dfrac{q_0 + q_1}{2} \\ 1 - 2\left(\dfrac{q_0 - q}{q_0 - q_1}\right)^2 & \text{if } \dfrac{q_0 + q_1}{2} < q \leq q_0 \\ 0 & \text{if } q_0 < q < 1 \end{cases}$$

$$m_{H_1}(q) = \begin{cases} 0 & \text{if } 0 < q \leq q_1 \\ 1 - 2\left(\dfrac{q - q_1}{q_0 - q_1}\right)^2 & \text{if } q_1 < q \leq \dfrac{q_0 + q_1}{2} \\ 2\left(\dfrac{q_0 - q}{q_0 - q_1}\right)^2 & \text{if } \dfrac{q_0 + q_1}{2} < q \leq q_0 \\ 1 & \text{if } q_0 < q < 1 \end{cases}$$

Such a choice can be justified as follows:

- On the one hand, the slope of the power function is not that steep in the fuzzy area (due to the symmetry of the binomial distribution for $q = 0.5$ in combination with smaller values of n).

- On the other hand, piecewise linear, convex, or concave membership functions in combination with plausibly rather narrow width of the fuzzy area lead to weight functions that are too steep within the indifference zone. The slope of the weighted power function thus results in suprema which are not larger as in the crisp area.

In contrast, by employing polynomial functions these problems can be avoided, and moreover, a concave-convex-shaped function is better suited to ensure a "smooth transition" within the fuzzy area around the value 0.5.

Example 6.3 (Calculation of generalized p-values when considering fuzzy hypotheses). As for the calculation of generalized p-values via (5) adjusted by means of (10) and (12) when considering fuzzy hypotheses we exemplary choose $n = 25$, $q_0 = 0.525$, $q_1 = 0.475$ (thus, $w = 0.05$), and s-shaped membership functions. Then we get the following p-values for Cases 1–3 (see also Table 10, columns 11–13, row 25):

$$p_l(25, 9.5) = \max\left\{ \begin{array}{c} \sum_{m=0}^{9.5} \frac{\Gamma(26)}{\Gamma(m+1)\Gamma(26-m)} 0.525^m (1-0.525)^{25-m} \\ \sup_{q \in (0.475, 0.525)} \left\{ (m_{H_0}(q) - m_{H_1}(q)) \sum_{m=0}^{9.5} \frac{\Gamma(26)}{\Gamma(m+1)\Gamma(26-m)} q^m (1-q)^{25-m} \right\} \end{array} \right\} = 0.1353$$

$$p_l(25, 8.75) = \max\left\{ \begin{array}{c} \sum_{m=0}^{8.75} \frac{\Gamma(26)}{\Gamma(m+1)\Gamma(26-m)} 0.525^m (1-0.525)^{25-m} \\ \sup_{q \in (0.475, 0.525)} \left\{ (m_{H_0}(q) - m_{H_1}(q)) \sum_{m=0}^{8.75} \frac{\Gamma(26)}{\Gamma(m+1)\Gamma(26-m)} q^m (1-q)^{25-m} \right\} \end{array} \right\} = 0.0707$$

$$p_l(31, 9.75) = \max\left\{ \begin{array}{c} \sum_{m=0}^{9.75} \frac{\Gamma(32)}{\Gamma(m+1)\Gamma(32-m)} 0.525^m (1-0.525)^{31-m} \\ \sup_{q \in (0.475, 0.525)} \left\{ (m_{H_0}(q) - m_{H_1}(q)) \sum_{m=0}^{9.75} \frac{\Gamma(32)}{\Gamma(m+1)\Gamma(32-m)} q^m (1-q)^{31-m} \right\} \end{array} \right\} = 0.0183$$

Note that the respective test statistics $s^{(-)}$ as well as the adjustment term $\mathrm{a}^{(-)}$ are the same as in the case of crisp hypotheses.

In the following, we perform a sensitivity analysis in the framework of the sign test regarding the sample size n (i.e., we successively use $n = 10, 11, \ldots, 51$ items of the data set) and various widths $w = q_0 - q_1$ of the indifference zone (and thus various widths of fuzzy areas) when considering fuzzy hypotheses.

As in the crisp hypotheses case, the results of the analysis are given in Table 10. In particular, Table 10 contains generalized p-values regarding fuzzy hypotheses for three cases, each calculated for $n = 1, \ldots, 51$. These p-values are determined using different values of w: $w = 0.05$ (columns 11–13), $w = 0.1$ (columns 14–16), and $w = 0.2$ (columns 17–19).

We analyze the p-values considering the above three cases and crisp versus fuzzy hypotheses. Figure 9 shows the behavior of (generalized) p-values for increasing values of n and is subdivided into four subfigures. While Figure 9A contains p-values for all the cases based on crisp hypotheses, Figures 9B–D present the respective cases based on fuzzy hypotheses for

TABLE 10 Data set and results of the sensitivity analysis

							Crisp hypotheses			Fuzzy hypotheses								
										$q_0 = 0.525$ and $q_1 = 0.475$			$q_0 = 0.55$ and $q_1 = 0.45$			$q_0 = 0.6$ and $q_1 = 0.4$		
				Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
i	x_i	$\widehat{k}_{0.5}$	MAD	$s^{(-)}$	$s^{(-)}$	$s^{(-)}$	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value
1	36.2	37.5	1.3	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	37.8	37.65	0.8	1	1.0000	1.3125	0.7500	0.7500	0.9171	0.7246	0.7246	0.8711	0.6986	0.6986	0.8241	0.6454	0.6454	0.7313
3	37.5	37.5	0.3	1.5	1.5000	1.5000	0.9244	0.9244	0.9244	0.8858	0.8858	0.8858	0.8455	0.8455	0.8455	0.7631	0.7631	0.7631
4	39.5	38.5	0.8	1.5	1.5000	1.8125	0.6520	0.6520	0.4776	0.5986	0.5986	0.4284	0.5479	0.5479	0.3837	0.4583	0.4583	0.3097
5	38	38	0.5	1.5	1.5000	1.7000	0.4300	0.4300	0.3006	0.3777	0.3777	0.2562	0.3321	0.3321	0.2195	0.2605	0.2605	0.1658
6	38.1	38.05	0.55	1.5	1.5000	1.7727	0.2711	0.2711	0.1220	0.2278	0.2278	0.0960	0.1929	0.1929	0.0772	0.1434	0.1434	0.0538
7	36.7	38	0.6	2.5	2.5000	2.8333	0.4529	0.4529	0.2370	0.3930	0.3930	0.1932	0.3412	0.3412	0.1593	0.2620	0.2620	0.1143
8	39.4	38.05	0.7	2.5	2.5000	3.0000	0.3092	0.3092	0.1938	0.2570	0.2570	0.1492	0.2154	0.2154	0.1179	0.1576	0.1576	0.0804
9	35.5	38	0.8	3.5	3.5000	4.1250	0.4656	0.4656	0.3243	0.3994	0.3994	0.2597	0.3428	0.3428	0.2107	0.2585	0.2585	0.1478
10	37.8	37.9	0.7	3.5	3.5000	4.2857	0.3347	0.3347	0.1666	0.2755	0.2755	0.1239	0.2286	0.2286	0.0954	0.1650	0.1650	0.0633
11	38	38	0.6	3.5	3.5000	4.1667	0.2327	0.2327	0.1077	0.1837	0.1837	0.0765	0.1478	0.1478	0.0572	0.1027	0.1027	0.0369
12	36.7	37.9	0.7	4.5	4.5000	5.4286	0.3530	0.3530	0.1372	0.2878	0.2878	0.0981	0.2367	0.2367	0.0736	0.1687	0.1687	0.0476
13	39	38	0.8	4.5	4.5000	5.6250	0.2552	0.2552	0.1026	0.1998	0.1998	0.0714	0.1594	0.1594	0.0527	0.1097	0.1097	0.0336
14	38	38	0.7	4.5	4.5000	5.5714	0.1795	0.1795	0.0464	0.1350	0.1350	0.0298	0.1047	0.1047	0.0211	0.0700	0.0700	0.0130
15	38.3	38	0.8	4.5	4.5000	5.8125	0.1233	0.1233	0.0352	0.0891	0.0891	0.0222	0.0674	0.0674	0.0155	0.0439	0.0439	0.0095
16	37.3	38	0.7	5.5	5.1429	6.2143	0.1983	0.1789	0.0478	0.1480	0.1312	0.0301	0.1141	0.1001	0.0210	0.0757	0.0657	0.0128
17	37	38	0.6	6.5	6.0833	6.8333	0.2880	0.2651	0.0622	0.2219	0.2005	0.0404	0.1746	0.1557	0.0287	0.1181	0.1040	0.0177
18	39.5	38	0.7	6.5	6.0000	7.0714	0.2143	0.1189	0.0539	0.1588	0.0848	0.0335	0.1217	0.0635	0.0233	0.0803	0.0409	0.0141
19	38.5	38	0.8	6.5	5.9375	7.2500	0.1560	0.0812	0.0395	0.1113	0.0555	0.0239	0.0832	0.0405	0.0164	0.0537	0.0256	0.0098
20	37.3	38	0.7	7.5	6.6429	7.7143	0.2282	0.1168	0.0330	0.1679	0.0808	0.0197	0.1280	0.0594	0.0135	0.0839	0.0377	0.0080

TABLE 10 (Continued)

							Crisp hypotheses			Fuzzy hypotheses								
										$q_0 = 0.525$ and $q_1 = 0.475$			$q_0 = 0.55$ and $q_1 = 0.45$			$q_0 = 0.6$ and $q_1 = 0.4$		
				Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
i	x_i	$\widehat{k}_{0.5}$	MAD	$s^{(-)}$	$s^{(-)}$	$s^{(-)}$	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value
21	37.5	38	0.6	8	7.2500	8.0000	0.1917	0.1596	0.0610	0.1395	0.1117	0.0377	0.1057	0.0826	0.0261	0.0688	0.0527	0.0158
22	35.7	37.9	0.7	9	8.1429	9.2143	0.2617	0.2253	0.0735	0.1961	0.1621	0.0456	0.1511	0.1218	0.0316	0.1000	0.0786	0.0191
23	37.5	37.8	0.6	9.5	8.7500	9.5000	0.3202	0.1922	0.0797	0.2412	0.1370	0.0501	0.1863	0.1023	0.0349	0.1234	0.0657	0.0212
24	39.4	37.9	0.7	9.5	8.6429	9.7143	0.2512	0.1405	0.0444	0.1823	0.0960	0.0263	0.1375	0.0699	0.0178	0.0891	0.0440	0.0106
25	37.8	37.8	0.6	9.5	8.7500	9.7500	0.1934	0.1066	0.0319	0.1353	0.0707	0.0183	0.0998	0.0506	0.0123	0.0634	0.0314	0.0072
26	37.1	37.8	0.55	10.5	9.6818	10.5000	0.2608	0.1523	0.0467	0.1881	0.1037	0.0272	0.1411	0.0753	0.0183	0.0910	0.0472	0.0109
27	36.4	37.8	0.6	11.5	10.5833	11.5833	0.3353	0.2064	0.0727	0.2491	0.1441	0.0438	0.1903	0.1060	0.0299	0.1246	0.0672	0.0179
28	38.4	37.8	0.7	11.5	10.4286	11.7857	0.2694	0.1539	0.0416	0.1931	0.1029	0.0237	0.1442	0.0740	0.0158	0.0925	0.0460	0.0093
29	39	37.8	0.8	11.5	10.3125	11.9375	0.2128	0.1130	0.0319	0.1471	0.0725	0.0177	0.1076	0.0510	0.0117	0.0678	0.0312	0.0068
30	38.4	37.9	0.8	11.5	10.3125	11.9375	0.1654	0.0836	0.0229	0.1103	0.0518	0.0124	0.0791	0.0359	0.0081	0.0491	0.0216	0.0047
31	38.3	38	0.8	11.5	10.3125	11.9375	0.1266	0.0610	0.0163	0.0815	0.0365	0.0085	0.0574	0.0249	0.0055	0.0351	0.0149	0.0032
32	36.7	37.9	0.8	12.5	11.3125	12.9375	0.1739	0.0914	0.0260	0.1154	0.0564	0.0140	0.0825	0.0389	0.0091	0.0510	0.0234	0.0053
33	39	38	0.8	12.5	11.3125	12.9375	0.1347	0.0677	0.0188	0.0863	0.0404	0.0098	0.0606	0.0274	0.0063	0.0370	0.0163	0.0036
34	36.8	37.9	0.8	13.5	12.2500	13.8750	0.1819	0.0976	0.0285	0.1200	0.0598	0.0152	0.0855	0.0411	0.0099	0.0527	0.0247	0.0057
35	39.2	38	0.8	13.5	12.2500	13.8750	0.1424	0.0731	0.0208	0.0908	0.0433	0.0108	0.0635	0.0293	0.0069	0.0387	0.0174	0.0040
36	38.6	38	0.8	13.5	12.2500	13.8750	0.1100	0.0541	0.0150	0.0678	0.0310	0.0075	0.0467	0.0207	0.0048	0.0281	0.0122	0.0027
37	38	38	0.8	13.5	12.2500	14.0625	0.0839	0.0396	0.0130	0.0500	0.0219	0.0063	0.0339	0.0145	0.0040	0.0202	0.0084	0.0022
38	36.8	38	0.8	14.5	13.1875	15.0000	0.1168	0.0586	0.0202	0.0716	0.0334	0.0100	0.0492	0.0222	0.0064	0.0295	0.0130	0.0036
39	37	38	0.8	15.5	14.0000	15.8125	0.1565	0.0815	0.0188	0.0988	0.0475	0.0093	0.0688	0.0319	0.0059	0.0416	0.0188	0.0034
40	38.5	38	0.8	15.5	14.0000	15.8125	0.1233	0.0615	0.0137	0.0753	0.0346	0.0066	0.0516	0.0230	0.0041	0.0309	0.0134	0.0023

(Continues)

TABLE 10 (Continued)

							Crisp hypotheses			Fuzzy hypotheses								
										$q_0 = 0.525$ and $q_1 = 0.475$			$q_0 = 0.55$ and $q_1 = 0.45$			$q_0 = 0.6$ and $q_1 = 0.4$		
				Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
i	x_i	$\hat{k}_{0.5}$	MAD	$s^{(-)}$	$s^{(-)}$	$s^{(-)}$	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value
41	36.5	38	0.8	16.5	15.0000	16.8125	0.1631	0.0875	0.0209	0.1025	0.0508	0.0104	0.0711	0.0340	0.0066	0.0430	0.0201	0.0037
42	37.5	37.9	0.8	17	15.5000	17.3125	0.1400	0.0737	0.0271	0.0869	0.0422	0.0135	0.0600	0.0281	0.0086	0.0361	0.0165	0.0049
43	37.5	37.8	0.8	17.5	16.0000	17.8125	0.1693	0.0933	0.0231	0.1059	0.0539	0.0114	0.0733	0.0361	0.0072	0.0442	0.0212	0.0041
44	36.5	37.8	0.8	18.5	17.0000	18.8125	0.2149	0.1261	0.0335	0.1383	0.0751	0.0170	0.0969	0.0508	0.0108	0.0590	0.0301	0.0062
45	37.5	37.8	0.8	19	17.5000	19.3125	0.1856	0.1071	0.0425	0.1179	0.0628	0.0217	0.0821	0.0422	0.0138	0.0497	0.0249	0.0079
46	38	37.8	0.8	19	17.5000	19.5000	0.1510	0.0838	0.0256	0.0928	0.0476	0.0124	0.0637	0.0316	0.0078	0.0382	0.0185	0.0044
47	37.5	37.8	0.8	19.5	18.0000	20.0000	0.1807	0.1044	0.0329	0.1121	0.0599	0.0161	0.0771	0.0399	0.0102	0.0463	0.0234	0.0057
48	37.7	37.75	0.75	19.5	18.1333	20.3333	0.1466	0.0834	0.0202	0.0880	0.0466	0.0095	0.0597	0.0307	0.0059	0.0355	0.0179	0.0033
49	36.3	37.7	0.8	20.5	19.0000	21.3750	0.1861	0.1097	0.0293	0.1148	0.0627	0.0141	0.0789	0.0417	0.0088	0.0472	0.0245	0.0050
50	37.5	37.6	0.75	21	19.6333	21.8333	0.1611	0.0954	0.0249	0.0981	0.0540	0.0118	0.0670	0.0358	0.0074	0.0400	0.0209	0.0041
51	38	37.7	0.7	21	19.7857	21.9286	0.1312	0.0773	0.0147	0.0774	0.0427	0.0066	0.0522	0.0280	0.0041	0.0308	0.0163	0.0023

TABLE 11 Fractions of (generalized) p-values of maximum 1%, 5%, and 10% for the considered cases ($n \geq 10$)

Significant at level (%)	Crisp hypotheses			Fuzzy hypotheses ($w = 0.05$)			Fuzzy hypotheses ($w = 0.1$)			Fuzzy hypotheses ($w = 0.2$)		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
1	0	0	0	0	0	0.2143	0	0	0.4524	0	0.0238	0.6667
5	0	0.0238	0.7619	0.0238	0.2857	0.8810	0.0714	0.5238	0.9048	0.4762	0.7381	0.9762
10	0.0238	0.4524	0.9048	0.3095	0.6905	0.9762	0.5952	0.7619	1	0.8333	0.8810	1

various values of the width w of the indifference zone. In addition, Table 11 gives the fractions of (generalized) p-values that are below or at 1%, 5%, and 10% for the considered cases. Analyzing the results in Tables 10 and 11 and Figure 9, we get the following insights:

- The case of crisp categories (Case 1) generally leads to the highest p-values.
- Considering fuzzy categories (Cases 2 and 3), we observe in general considerably lower p-values in Case 3 compared to Case 2.
- p-values are generally higher for crisp hypotheses (columns 8–10) compared to fuzzy hypotheses (case by case comparison).
- Considering fuzzy hypotheses, p-values are the lower, the larger the width w of the indifference zone (case by case comparison).
- As a rule, the larger n, the smaller the p-value. This behavior is shown in Figure 9, and moreover, it can be seen that p-values are on average below 5% as of $n = 14$ (Case 3, crisp and fuzzy hypotheses) and $n = 30$ (Case 2, fuzzy hypotheses).
- Considering Case 1, there are no p-values of maximum 1% as well as only a few p-values of maximum 5% when hypotheses are fuzzy ($w = 0.05, 0.1$) and no one when hypotheses are crisp. As for Case 3, more than 90% of p-values are below 10% and there are strongly increasing fractions of p-values below 1% for larger widths w of the indifference zone, while there is no p-value below 1% when hypotheses are crisp.

Summarized, the classical sign test for the median (with crisp categories and crisp hypotheses) generally does not lead to a rejection of H_0 and thus to statistical validation of hyperthermia symptom related to COVID-19 in HIV-infected individuals. This is due to the fact that the hypothesized median value $37.5°\text{C}$ is chosen very close to the true value (see also the empirical median $\widehat{k}_{0.5} = 37.7$ that is, in turn very close to 37.5), so there is no strong evidence against the validity of H_0. In contrast, the extended sign test for the median mostly leads to a rejection of H_0 and thus to statistical validation of hyperthermia through the abandonment of exclusive categories and an incorporation of the uncertainty level regarding the 50%-percentile specified by the border temperature value of 37.5°C.

7 | CONCLUSIONS

In this paper, we have presented an extended sign test under implementation of fuzzy categories and fuzzy formulated hypotheses. It has been shown that this generalized test procedure improves the generality, versatility, and practicability of the common sign test. In particular, the generalized sign test provides the following benefits compared to the usual test procedure with respect to fuzzy categories (1.–2.), fuzzy hypotheses (3.–4.), and both of them (5.):

1. The magnitude of the deviations between observations and hypothesized quantile value k is no longer ignored but incorporated into the test procedure. In particular, the formulation of membership functions for both categories helps to adequately map the magnitude of the observations in the border area, that is, via decreasing/increasing membership degrees to the respective category. In addition, there is also the possibility to consider the order of observations beyond the border area via full or no membership (1 vs. 0) to the respective category in compliance with the classical approach.

2. Several problems of the common sign test related to ties in the data are considerably mitigated by a neutral handling of ties (i.e., classification of ties to both categories in an equal magnitude of 0.5) and simultaneous implementation of the untransformed real-valued test statistic in the definition of the p-value by using gamma functions instead of binomial coefficients.
3. The imprecision of linguistic statements on fractions q regarding the underlying quantiles k_q is adequately modeled by implementing fuzzy sets, that is, the magnitude of uncertainty is implemented by means of width of the indifference zone (and thus of the fuzzy areas of the hypotheses with respective (non-)complementary decreasing/increasing membership functions, whose shapes include possibilistic aspects).
4. The formulation of fuzzy hypotheses on $q \in (0, 1)$ instead of $k_q \in \mathbb{R}$ is more natural and simpler for the practitioner/researcher due to the fact that the underlying test is of binomial type and specific deviations (e.g., 2.5%, 5%, and 10%) from the hypothesized fraction value (like 50% when testing for the median) can be coherently embedded both into single applications and automated processes like big data knowledge extraction.
5. The combination of fuzzy categories (with real-valued test statistic) and fuzzy hypotheses (with expanded sets of hypothesized fractions) rather enables a rejection of H_0 in cases where the hypothesized quantile value is specified very close to the true one.

With all these advantages, the generalized sign test is neat in theory and practice, provides an interpretation of the generalized p-value in line with the classical probabilistic p-value, and ensures a crisp (binary) test decision, that is, “reject/do not reject H_0”. This is not self-evident, since fuzzy tests often (1) lead to difficulties in practical applications due to the complexity of the underlying methodology (like modeling of fuzzy data, fuzzy random variables, fuzzy parameters) and (2) result in fuzzy test decisions supported by interpretations of the observed and theoretical key measures in a possibilistic sense, and therefore lack a sound basis for decision-making that is essential for the practitioner.

To show the benefits of the proposed methodology in practical applications, we have performed an extensive case study on hyperthermia symptom related to COVID-19 in HIV-infected individuals. In particular, we have replaced the crisp category “(non-)increased temperature” by a fuzzy one due to the fact that (human body) temperature is a categorical variable, and moreover, linguistic statements on the 50%-threshold regarding the underlying median were embedded into the test procedure by employing fuzzy hypotheses. The test results clearly indicate that in situations with rigidness originated from exclusive categories as well as with uncertainties regarding the fractions of the underlying quantiles

- fuzzy categories are preferable to crisp categories and
- fuzzy hypotheses are preferable to crisp hypotheses.

An important direction for future research is the generalization of the common two-tailed sign test in compliance with the extended statistical model proposed in this paper, whose methodological complexity is considerably increased. Finally, it should be emphasized that although the paper is focused on improving the sign test, the introduced methodology can be transferred to other nonparametric and parametric test procedures.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support provided by the German Research Foundation (DFG) in the framework of the research project no. 409030527. The authors are also grateful to the student assistant Marcel Gaweda for his engaged support. Finally, the authors thank both anonymous reviewers for their valuable feedback and suggestions, which were important and helpful to improve the paper.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

ORCID

Nataliya Chukhrova https://orcid.org/0000-0002-4105-7033
Arne Johannssen https://orcid.org/0000-0002-4265-3996

REFERENCES

1. Grzegorzewski P. K-sample median test for vague data. *Int J Intell Syst*. 2009;24(5):529-539.
2. Shafiq M, Atif M, Viertl R. Generalized likelihood ratio test and Cox's F-test based on fuzzy lifetime data. *Int J Intell Syst*. 2017;32(1):3-16.
3. Chukhrova N, Johannssen A. Fuzzy hypothesis testing: systematic review and bibliography. *Appl Soft Comput*. 2021;106:107331.
4. Grzegorzewski P. Statistical inference about the median from vague data. *Control Cybernet*. 1998;27(3): 447-464.
5. Grzegorzewski P. Distribution-free tests for vague data. In: Lopez-Diaz M, Gil MA, Grzegorzewski P, Hryniewicz O, Lawry J, eds. *Soft Methodology and Random Information Systems*. Vol 26. Heidelberg: Springer; 2004:495-502. Advances in Intelligent and Soft Computing.
6. Grzegorzewski P, Spiewak M. The sign test for interval-valued data. In: Ferraro MB, Giordani P, Vantaggi B, Gagolewski M, Gil MA, Grzegorzewski P, Hryniewicz O, eds. *Soft Methods for Data Science*. Vol 456. Springer; 2016:269-276. Advances in Intelligent Systems and Computing.
7. Grzegorzewski P, Spiewak M. The sign test and the signed-rank test for interval-valued data. *Int J Intell Syst*. 2019;34(9):2122-2150.
8. Grzegorzewski P, Spiewak M. Two-sample median test for interval-valued data. In: Atanassov KT, Atanassova V, Kacprzyk J, Kaluszko A, Krawczak M, Owsinski JW, Sotirov S, Sotirova E, Szmidt E, Zadrozny S, eds. *Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives*. Vol 1081. Springer; 2021:113-128. Advances in Intelligent Systems and Computing.
9. Hesamian G, Chachi J. Fuzzy sign test for imprecise quantities: A p-value approach. *J Intell Fuzzy Syst*. 2014;27(6):3159-3167.
10. Hesamian G, Taheri SM. Credibility theory oriented sign test for imprecise observations and imprecise hypotheses. In: Kruse R, Berthold M, Moewes C, Gil MA, Grzegorzewski P, Hryniewicz O, eds. *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*. Vol 190. Berlin: Springer; 2013: 153-163. Advances in Intelligent Systems and Computing.
11. Kahraman C, Bozdag CF, Ruan D, AF Özok. Fuzzy sets approaches to statistical parametric and non-parametric tests. *Int J Intell Syst*. 2004;19(11):1069-1078.
12. Momeni F, Sadeghpour-Gildeh B. Nonparametric tests for median in fuzzy environment. *Int J Fuzzy Syst*. 2016;18(1):130-139.
13. Shams M, Hesamian G. Sign test for fuzzy random variables. *J Algorithm Comput*. 2018;50(2):121-139.
14. Couso I, Dubois D. Statistical reasoning with set-valued information: ontic vs. epistemic views. *Int J Approx Reason*. 2014;55(7):1502-1518.
15. Dixon WJ, Mood AM. The statistical sign test. *J Am Stat Assoc*. 1946;41(236):557-566.
16. De Garibay VG. Behaviour of fuzzy ANOVA. *Kybernetes*. 1987;16(2):107-112.

17. Grzegorzewski P, Szymanowski H. Chi-square test for homogeneity with fuzzy data. In: Grzegorzewski P, Gagolewski M, Hryniewicz O, Gil MA, eds. *Strengthening Links Between Data Analysis and Soft Computing*. Vol 315. Cham: Springer; 2015:151-158. Advances in Intelligent Systems and Computing.
18. Taheri SM, Hesamian G. Goodman-Kruskal measure of association for fuzzy-categorized variables. *Kybernetika*. 2011;47(1):110-122.
19. Wu B, Chang S-K. On testing hypothesis of fuzzy sample mean. *Japan J Industr Appl Math*. 2007;24(2):197-209.
20. Chukhrova N, Johannssen A. Fuzzy hypothesis testing for a population proportion based on set-valued information. *Fuzzy Sets and Systems*. 2020;387:127-157.
21. Zimmermann H-J. *Fuzzy Set Theory and its Applications*. 4th ed. Dordrecht: Kluwer Academic Publishers; 2001.
22. Chukhrova N, Johannssen A. Inspection tables for single acceptance sampling with crisp and fuzzy formulation of quality limits. *Int J Qual Reliab Manag*. 2018;35(9):1755-1791.
23. Chukhrova N, Johannssen A. Randomized vs. non-randomized hypergeometric hypothesis testing with crisp and fuzzy hypotheses. *Statistical Papers*. 2020;61(6):2605-2641.
24. Chukhrova N, Johannssen A. Generalized one-tailed hypergeometric test with applications in statistical quality control. *J Quality Technol*. 2020;52(1):14-39.
25. Arnold BF. An approach to fuzzy hypothesis testing. *Metrika*. 1996;44(1):119-126.
26. Arnold BF. Testing fuzzy hypotheses with crisp data. *Fuzzy Sets Syst*. 1998;94(3):323-333.
27. Pham-Gia T, Hung TL. The mean and median absolute deviations. *Math Comput Model*. 2001;34(7-8):921-936.
28. Vizcarra P, Perez-Elias MJ, Quereda C, et al. Description of COVID-19 in HIV-infected individuals: a single-centre, prospective cohort. *The Lancet HIV*. 2020;7(8):e554-e564.
29. Hutchison JS, Ward RE, Lacroix J, et al. Hypothermia therapy after traumatic brain injury in children. *New England J Med*. 2008;358(23):2447-2456.
30. Laupland KB. Fever in the critically ill medical patient. *Critical Care Med*. 2009;37(7):273-278.
31. Grunau BE, Wiens MO, Brubacher JR. Dantrolene in the treatment of MDMA-related hyperpyrexia: a systematic review. *Canadian J Emergency Med*. 2010;12(5):435-442.
32. Moran DS, Mendal L. Core temperature measurement: methods and current insights. *Sports Med*. 2002;32(14):879-885.

How to cite this article: Chukhrova N, Johannssen A. Nonparametric fuzzy hypothesis testing for quantiles applied to clinical characteristics of COVID-19. *Int J Intell Syst*. 2021;36:2922–2963. https://doi.org/10.1002/int.22407