

Fuzzy-Logic-Driven Scenario Steering for Adaptive Whole-of-Society Tabletop Exercises

Niklas Tschäschke

TH Köln – University of Applied Sciences
niklas.tschaeschke@th-koeln.de

Lennart Landsberg

TH Köln – University of Applied Sciences
lennart_walter.landsberg@th-koeln.de

Christoph Schneider

TH Köln – University of Applied Sciences
christoph.schneider@th-koeln.de

Ompe Aimé Mudimu

TH Köln – University of Applied Sciences
ompe_aime.mudimu@th-koeln.de

ABSTRACT

Conventional tabletop exercises (TTEs) assume linear scenario trajectories, while contemporary crises evolve through nonlinear interactions between governance decisions and societal dynamics such as trust, compliance, and coordination. Iterative refinement between exercises is insufficient for whole-of-society preparedness in complex health crises. This paper proposes an adaptive TTE design that operationalizes societal feedback as structured inputs for in-exercise scenario steering. Developed within the European PREPSHIELD project, the approach combines causal-loop modeling with a fuzzy-rule-based inference mechanism linking participant behavior to dynamic scenario state updates. Findings from an initial Hamburg pilot demonstrate feasibility and reveal critical design challenges, particularly that overly abstract measurement instruments risk obscuring relevant social dynamics. The paper contributes a candidate methodological framework for adaptive whole-of-society exercises, supported by empirically refined rule patterns and a multidimensional evaluation strategy. As a work-in-progress study, the design is undergoing validation. While developed for health crises, the mechanism is designed for domain-specific adaptation.

Keywords

Tabletop exercises, adaptive scenario steering, whole-of-society preparedness, fuzzy inference systems, crisis exercise design

INTRODUCTION

Tabletop exercises (TTEs) are an established instrument in crisis management for testing coordination, decision-making, and inter-organizational learning under simulated conditions (Emaliyawati et al., 2025). TTEs are typically structured as scenario-based activities organized into predefined rounds. A storybook specifies the development of the crisis situation, while facilitators guide participants through discussion phases followed by debriefings and after-action reviews. Improvements are commonly introduced between exercise iterations by refining scenarios, roles, or facilitation strategies (Baumann et al., 2025). This iterative structure has proven valuable for testing procedures and strengthening inter-organizational understanding.

This assumption of scenario stability is inconsistent with evidence on contemporary crisis dynamics. Pandemics, climate-related emergencies, and cyber incidents are characterized by compound risks where health impacts, misinformation, resource constraints, and governance challenges interact across sectors (Kruczkiewicz et al., 2021). Recent evidence confirms this complexity at the operational level. A systematic review of reviews mapping healthcare facility challenges during past outbreaks in Europe found that challenges were distributed across all dimensions of the 4S surge capacity framework - staff, stuff, space, and system - with system-level issues such as unclear protocols, inconsistent communication, and telemedicine adoption barriers being most frequently reported (Rubini et al., 2026). Crisis outcomes emerge from nonlinear social dynamics, including shifts in trust toward authorities, changing compliance behavior, and evolving coordination patterns among governmental, civil society, and citizen actors (Stivas & Cole, 2023). These dynamics are central to whole-of-society crisis management but remain difficult to anticipate in advance. While TTEs often discuss such dynamics during reflection phases, most

exercise designs do not allow the scenario itself to change in response to participant behavior during execution (Emaliyawati et al., 2025).

European research initiatives such as the PREPSHIELD project increasingly use tabletop exercises to explore whole-of-society preparedness for complex health crises (Baumann et al., 2025; Emaliyawati et al., 2025). In this project, TTEs follow an iterative multi-pilot design; however, adaptation primarily occurs between exercises rather than during them. Emerging social dynamics such as declining trust, coordination breakdowns, or unexpected compliance patterns are therefore typically analyzed retrospectively instead of shaping the ongoing scenario.

This observation reflects a broader methodological gap in crisis exercise design. Despite advances in complex systems thinking and adaptive governance research, most tabletop exercises rely on predefined scenario trajectories and facilitator improvisation when unexpected developments occur (Emaliyawati et al., 2025). Even exercises addressing compound risks often treat adaptivity informally, which limits transparency, reproducibility, and systematic evaluation of how social variables influence scenario evolution (Sydnes et al., 2025). Rubini et al. (2026) demonstrate that the vast majority of reviewed literature focused exclusively on COVID-19, with challenges from other infectious disease outbreaks such as H1N1, measles, or Mpox remaining largely undocumented. This underscores the need for preparedness instruments - including tabletop exercises - that are designed for transferability across hazard types rather than being calibrated to a single pandemic experience, aligning with the all-hazards approach advocated in the broader preparedness literature.

Against this background, this paper makes three contributions to the methodology of adaptive crisis exercises within the Command and Control domain. First, a conceptual contribution: We reframe the role of societal actors in tabletop exercises from exogenous scenario conditions to endogenous causal inputs. This shift reconceptualizes adaptive exercises as socio-technical systems in which governance decisions and societal responses jointly determine scenario evolution through transparent, rule-based inference - advancing beyond facilitator improvisation toward reproducible scenario steering. Second, a methodological contribution: We specify and operationalize a fuzzy-logic-based steering mechanism that translates qualitative participant assessments of trust, compliance, and coordination into structured scenario adjustments between exercise rounds. The mechanism combines causal-loop-derived variable structures with Mamdani-type inference, providing a formal yet linguistically interpretable bridge between observed participant behavior and scenario state changes. Third, an empirical contribution (work-in-progress): We report design-informing observations from an initial pilot exercise that reveal a critical measurement challenge - the misalignment between academically derived assessment constructs and participants' situated interpretive frames - and present concrete design refinements for subsequent validation.

The proposed approach connects to central concerns in the Command and Control literature. Effective C2 depends on shared situational awareness, coordinated decision-making under uncertainty, and the capacity to adapt operational responses as situations evolve (Comfort, 2007; Wolbers et al., 2013). The adaptive mechanism proposed here creates a closed-loop decision environment in which the quality of command and coordination processes has observable, traceable consequences for crisis trajectories - positioning tabletop exercises not merely as discussion forums but as structured decision support environments that approximate the feedback dynamics characteristic of real-world C2 operations.

As a work-in-progress study, the paper focuses on conceptual specification and preliminary design evaluation rather than summative validation. The adaptive mechanism is presented as a candidate framework undergoing iterative empirical refinement.

RELATED WORK

Tabletop exercises are widely used for crisis preparedness, coordination testing, and inter-organizational learning (Alakrawi et al., 2024; Emaliyawati et al., 2025). However, the literature on exercise methodology reveals a persistent gap between the complexity of contemporary crises and the structural rigidity of most exercise formats. Emaliyawati et al. (2025) confirm that TTEs typically follow predefined scenario scripts with facilitator-led adjustments. Sydnes et al. (2025) demonstrate that interorganizational coordination - a central exercise objective - depends on dynamic information exchange that static scenarios cannot fully represent. While some exercise frameworks incorporate branching narratives or conditional injects, these adaptations typically rely on facilitator judgment without formal inference mechanisms, limiting transparency and reproducibility.

The broader crisis simulation literature has explored computational approaches to scenario generation and adaptive training environments. Agent-based models and serious games enable dynamic scenario evolution but require computational infrastructure that exceeds typical tabletop settings. The C2 training literature emphasizes the importance of realistic decision environments that produce consequential feedback, yet few frameworks

operationalize this principle for discussion-based exercise formats. The present study occupies the intersection of these traditions: it applies formal inference mechanisms within tabletop - rather than simulation - settings, maintaining the accessibility and participatory character of TTEs while introducing structured adaptivity.

Fuzzy logic has been applied in various crisis-adjacent domains, including risk assessment, infrastructure vulnerability analysis, and organizational performance evaluation (Higashi et al., 2024; Thalmeiner et al., 2025). Its capacity to handle linguistic variables, graded membership, and expert-defined rule bases makes it well suited for domains where precise quantification is infeasible. Vonglao (2017) demonstrates that fuzzy approaches improve the interpretability of Likert-scale data - directly relevant for translating participant assessments into exercise inputs. However, to our knowledge, fuzzy inference has not previously been applied as a real-time steering mechanism within crisis tabletop exercises.

The reviewed literature reveals a methodological gap at the intersection of adaptive exercise design and formal inference mechanisms. While the need for more dynamic and responsive exercise formats is widely acknowledged, existing approaches either rely on informal facilitator adaptation or require full simulation environments. No established framework provides transparent, rule-based, yet linguistically interpretable scenario steering within discussion-based tabletop formats. The present study addresses this gap.

ADAPTIVE TTE DESIGN: FROM STATIC STORYBOOKS TO DYNAMIC SCENARIOS

Operationalizing Whole-of-Society Dynamics through Exercise Variables

Recent work on whole-of-society preparedness conceptualizes crisis outcomes as emergent properties of coordination, communication, and trust among state actors, civil society, and citizens, extending beyond institutional capacity (Ruggiero et al., 2024). In compound and prolonged crises, trajectories are shaped by how actors perceive authorities, coordinate actions, and adhere to recommended measures (Žilinskas & Sakalauskas, 2021). These social dynamics are central for crisis evolution but cannot automatically be transferred into tabletop exercise design (Alakrawi et al., 2024; Wu et al., 2022).

Structural factors such as historical trust legacies or socio-economic inequalities remain only indirectly observable in exercise settings and are difficult to connect to scenario adjustments. For this reason, the present study adopts a parsimonious variable selection strategy.

Exercise variables must meet three criteria. First, they must be empirically linked to crisis outcomes such as stabilization, escalation, or response effectiveness. Second, they must be sensitive to intervention through communication strategies, governance decisions, or operational measures. Third, they must be observable during the exercise through participant decisions, structured assessments, or interaction patterns (Baumann et al., 2025; Emaliyawati et al., 2025).

Based on these criteria, the adaptive design focuses on trust, compliance, and coordination. Empirical research from recent health crises shows that trust in authorities strongly predicts adherence to protective measures and shapes crisis dynamics (Han et al., 2023; Seyd & Bu, 2022; Shanka & Menebo, 2022). Research on disaster response highlights cross-organizational coordination as a determinant of effective crisis management, particularly in multi-actor environments (Sydnes et al., 2025). Compliance represents the behavioral translation of trust and coordination into action. Together, these variables capture the behavioral and perceptual mechanisms through which whole-of-society crises propagate, while remaining measurable within tabletop formats.

Whole-of-society crises are characterized by feedback structures linking health dynamics, economic activity, governance measures, and societal responses. Systems research demonstrates that these interdependencies unfold through interacting feedback loops rather than linear cause-effect chains (Strelkovskii & Rovenskaya, 2021). A comparative review of COVID-19 causal loop diagrams confirms that health, economic, governmental, and social perception variables appear across the majority of published models and form tightly coupled feedback structures (Sahin et al., 2020; Strelkovskii & Rovenskaya, 2021).

Building on this literature, the present design clusters CLD-derived variables into four subsystems: Health, Economy, Measures, and Sentiment in Society. This grouping reduces system complexity while preserving dominant feedback relationships. The four subsystems are not exhaustive, but they represent dimensions that repeatedly emerge as central drivers of crisis evolution and are directly relevant for decision-making in exercises.

In adaptive tabletop exercises, these subsystems become scenario state variables observable by participants. Each subsystem is influenced by a defined set of input variables elicited from decisions, interaction patterns, and structured assessments. Changes in one subsystem propagate to others: restrictive measures reduce economic activity, which influences societal sentiment and compliance, feeding back into health outcomes.

By grounding exercise variables in empirically documented feedback structures, adaptive scenario changes follow plausible system dynamics rather than ad hoc facilitator choices. This establishes a transparent link between observed participant behavior and structured scenario adjustment.

Table 1. Input Variables and Output Dimensions in Adaptive TTEs

Subsystem	Example Input Variables (measured in TTE)	Output Dimension (Scenario State)
Health	Hygiene practices, infected population, testing campaign, health system capacity	Overall health impact
Economy	Business closures, unemployment rate, economic activity	Economic stability
Measures	Restrictions on businesses, clarity of messages, government stimulus package, awareness campaigns	Intensity and coherence of measures
Sentiment in Society	Public trust, panic and fear, non-compliance, misinformation, perceived effectiveness of measures, social interaction	Societal sentiment

Note. Each subsystem corresponds to a scenario state variable (e.g., 'Health system capacity: low/medium/high') that evolves based on input variables (e.g., trust, compliance) measured during the exercise.

Fuzzy Logic as Mechanism for Adaptive Scenario Steering

While the Section above defines *what* is measured, the question remains *how* these measurements are translated into scenario adjustments. This section introduces fuzzy logic as the formal mechanism for this translation.

Adaptive in-exercise steering requires a method that translates qualitative participant assessments and partly ambiguous observer data into reproducible scenario adjustments without imposing false precision. Traditional facilitator-based steering lacks transparency and comparability. Purely binary trigger mechanisms oversimplify gradual social dynamics.

Fuzzy logic addresses this requirement. It enables rule-based integration of multiple inputs while allowing graded membership in linguistic categories such as low, medium, and high. Variables are represented through membership functions that assign partial belonging to overlapping sets. This allows gradual transitions instead of abrupt thresholds and combines linguistic reasoning with formal computation. (Mamdani & Assilian, 1975; Zadeh, 1965)

In this paper, adaptivity refers exclusively to structured scenario modifications that occur between exercise rounds within a single session. After each round, facilitator observations, Crisis-Management-Team (CMT) decision logs, and Market Group (MG) feedback are processed through the fuzzy inference system, producing updated subsystem scores that guide inject selection for the subsequent round. This within-session steering is distinct from iteration, which describes improvements applied between separate exercise events (e.g., between TTE1 and TTE2). The adaptive mechanism does not operate continuously during discussion phases; it processes accumulated round-level data at defined transition points.

This capacity for graded classification is well suited to tabletop exercises. Social perceptions are rarely binary: trust, coordination, and compliance shift gradually and unevenly. Fixed thresholds such as “trust below 50 percent triggers protest inject ignore these nuances. Fuzzy logic allows a value measured on a Likert scale to activate multiple categories simultaneously, enabling smoother and more plausible scenario transitions. (Higashi et al., 2024; Vonglao, 2017)

Within the adaptive TTE design, fuzzy logic links CMT performance and MG feedback to scenario evolution. Structured observer assessments, MG evaluations, and decision logs are transformed into fuzzy input values. A predefined rule base connects combinations of input states to subsystem-level scenario adjustments. (Higashi et al., 2024; Thalmeiner et al., 2025)

Design of Linguistic Variables and Membership Functions

All input and output variables are defined on a continuous scale from 0 to 10 and mapped to three overlapping triangular membership functions: Low [0,0,5], Medium [2,5,8], and High [5,10,10] (Figure 1).

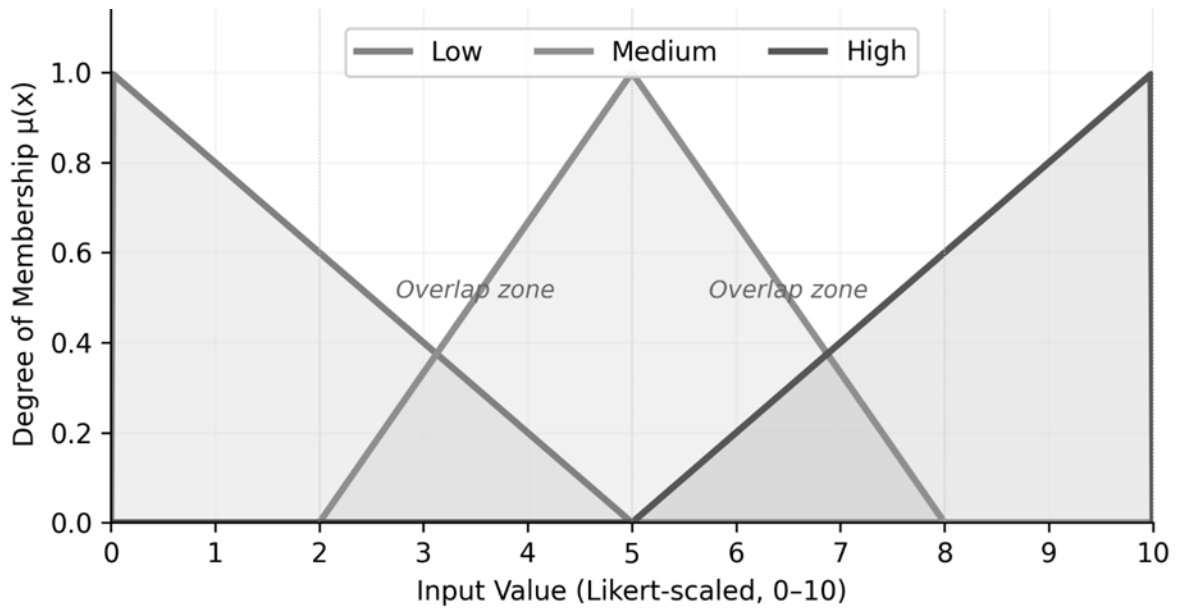


Figure 1. Standardized triangular membership functions for all input and output variables in the [anonymized] fuzzy inference system

Note. All variables are defined on a continuous [0, 10] scale with three linguistic categories: Low [0, 0, 5], Medium [2, 5, 8], and High [5, 10, 10]. Overlap zones between adjacent categories (approximately [2, 5] and [5, 8]) ensure that intermediate values activate multiple membership functions simultaneously, enabling graded rather than discrete classification.

For example, a trust assessment of 3.5 activates both Low ($\mu \approx 0.30$) and Medium ($\mu \approx 0.50$), reflecting ambiguity in participant judgments rather than forcing discrete classification.

During exercises, heterogeneous data from MG assessments, observer instruments, and CMT logs are normalized to the 0-10 scale before entering the inference system.

System Architecture and Rule Base

The adaptive mechanism consists of four parallel Mamdani-type fuzzy inference systems corresponding to the subsystems Health, Economy, Measures, and Sentiment in Society. Each subsystem processes a defined subset of input variables and produces one defuzzified output score representing the current scenario state.

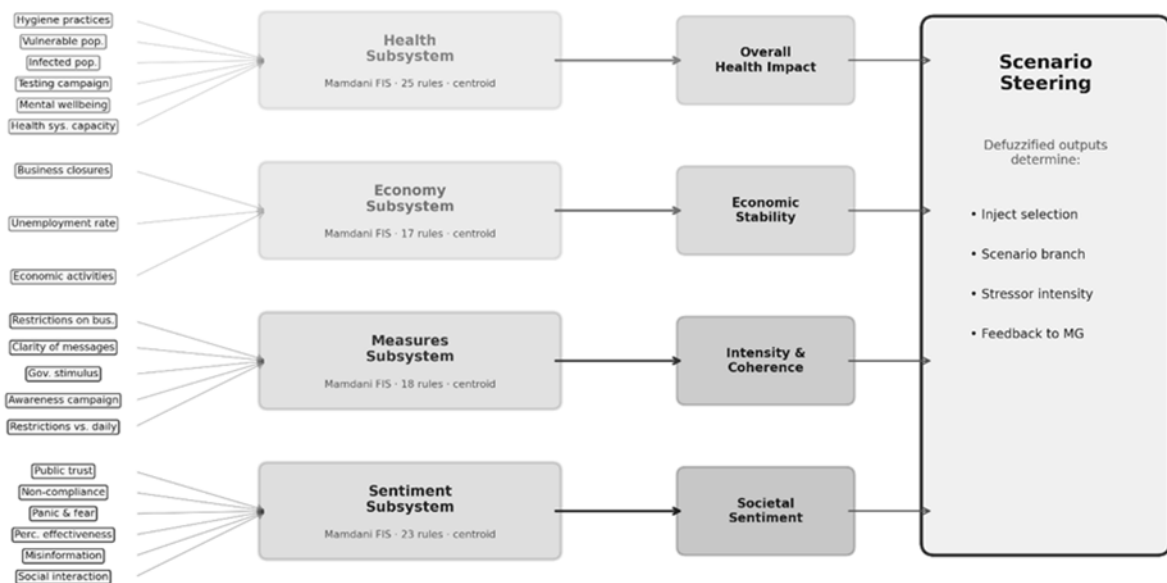


Figure 2. Architecture of the PREPSHIELD adaptive steering mechanism comprising four parallel Mamdani-type fuzzy inference subsystems.

Note. Twenty input variables are processed through 83 expert-defined rules distributed across four subsystems (Health: 25 rules, Economy: 17, Measures: 18, Sentiment in Society: 23). Each subsystem produces one defuzzified output score on a [0, 10] scale. Output scores are interpreted by the Directors Team to select scenario injects for subsequent exercise rounds. CMT = Crisis Management Team; MG = Market Group.

Rules are formulated as IF–THEN statements. Antecedents are combined using the minimum operator. Unspecified inputs act as “don’t care” conditions. The 83 rules were developed through a structured expert elicitation process conducted by the research team in collaboration with domain specialists from public health, crisis governance, and civil protection. The initial rule base was derived from published causal loop structures (Sahin et al., 2020; Strelkovskii & Rovenskaya, 2021) by systematically translating documented feedback relationships into IF–THEN rule patterns. Rule plausibility was assessed through iterative review sessions and adjusted during TTE1 preparation based on scenario logic and facilitator feedback. The current rule base represents expert consensus on plausible input-output relationships and is subject to empirical revision across subsequent pilots. Three illustrative rules from the Health subsystem demonstrate the encoded expert knowledge:

R1: IF Hygiene practices = High AND Vulnerable population = Low AND Infected population = Low AND Testing campaign = High AND Mental wellbeing = High AND Health system capacity = Medium THEN Health = High

R3: IF Hygiene practices = Low AND Infected population = High AND Health system capacity = Low THEN Health = Low

R9: IF Hygiene practices = High AND Infected population = Low AND Testing campaign = High AND Health system capacity = Medium THEN Health = High

R1 encodes a favorable constellation. R3 represents an escalation pathway. R9 captures a compensatory pattern where strong hygiene and testing sustain positive outcomes despite unspecified variables.

Across 83 rules, output distributions reflect expert assumptions about crisis dynamics. The Health subsystem contains more rules leading to Low outcomes, reflecting asymmetric health risks. The Measures subsystem contains more High-outcome rules, reflecting the reinforcing nature of coherent policy packages.

Inference and Defuzzification Process

For each exercise round, the inference process consists of four steps. First, fuzzification maps observed input values to membership degrees across linguistic categories. Second, rule evaluation determines firing strengths using the minimum operator. Third, aggregation combines activated rule outputs using the maximum operator. Fourth, defuzzification applies the centroid method to produce a crisp value between 0 and 10. This value represents the current state of the subsystem and is used by the Directors Team to select the next scenario path. A fully worked example is provided in the following Section.

Illustrative Adaptive Scenario Branching: Mini-Scenario from the Hamburg Pilot

To demonstrate adaptivity in practice, this section presents a simplified two-round scenario derived from the Hamburg influenza exercise. The example illustrates how identical starting conditions lead to diverging trajectories once societal feedback is treated as an endogenous driver.

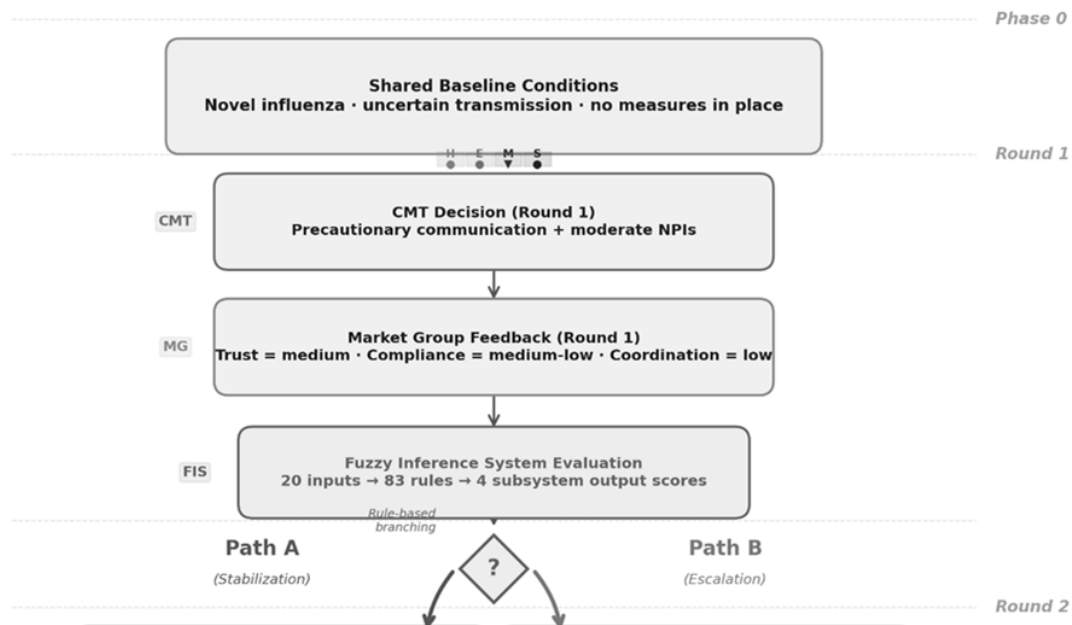


Figure 3. Adaptive scenario branching in the Hamburg pilot exercise, illustrating divergent crisis trajectories from identical baseline conditions.

Note. All participants begin from the same Phase 0 baseline (moderate health impact, stable economy, neutral sentiment, no measures). CMT decisions and MG feedback are evaluated through the fuzzy inference system, producing subsystem state profiles that guide scenario inject selection. Path A represents a stabilization trajectory (Trust ↑, Compliance ↑, Coordination ↑); Path B represents an escalation trajectory (Trust ↓, Compliance ↓, Coordination ↓). The branching mechanism is rule-based and reproducible: identical input values produce identical steering outputs. CMT = Crisis Management Team; MG = Market Group; FIS = Fuzzy Inference System.

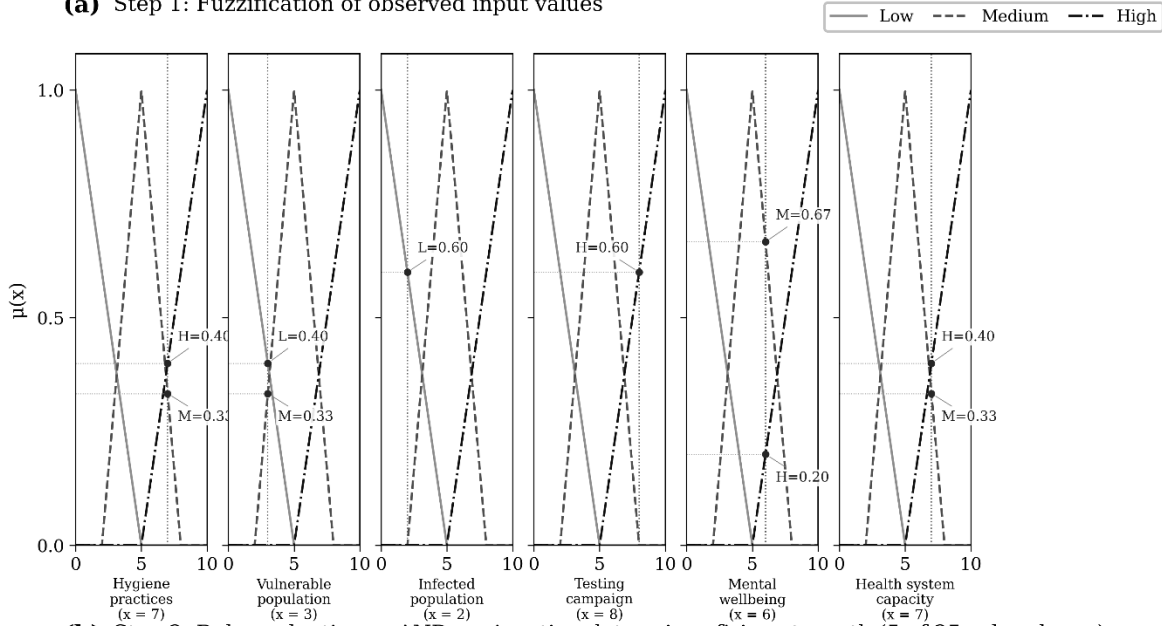
This worked example illustrates the target inference logic rather than a direct record of TTE1 execution, where moderators translated qualitative MG feedback into crisp input values using structured but partly subjective interpretation (see Section “Pilot Observations: Design Refinement Evidence from the Hamburg Pilot”). Subsequent pilots aim to standardize this translation through digital infrastructure.

All participants begin with identical baseline conditions: a novel influenza virus with uncertain transmission characteristics and no measures in place. The Directors Team assesses initial subsystem states as moderate health impact, stable economy, neutral societal sentiment, and absence of restrictions.

In Round 1, the CMT introduces precautionary communication and moderate non-pharmaceutical interventions. The MG provides structured feedback indicating Trust = medium with approximately 60 percent membership, Compliance = medium-low, and Coordination = low. These assessments are translated into fuzzy input values according to the operationalization described in Section “Fuzzy Logic as a Mechanism for Adaptive Scenario Steering”.

Figure 4 traces the full Mamdani inference process for the Health subsystem using values from the pilot. Hygiene practices = 7, Vulnerable population = 3, Infected population = 2, Testing campaign = 8, Mental wellbeing = 6, and Health system capacity = 7 are fuzzified and mapped to overlapping membership degrees.

(a) Step 1: Fuzzification of observed input values



(b) Step 2: Rule evaluation — AND-conjunction determines firing strength (5 of 25 rules shown)

Rule	Hygiene	Vuln. Pop.	Infected	Testing	Mental WB	Health Cap.	→ Output	Firing Strength
R1	High	Low	Low	High	High	Med.	High	0.20
R9	High	—	Low	High	—	Med.	High	0.33
R16	High	—	—	—	High	High	High	0.20
R3	Low	—	High	—	—	Low	Low	0.00
R13	—	—	Med.	—	Med.	Med.	Medium	0.00

(c) Steps 3 & 4: Aggregation, centroid defuzzification, and scenario steering interpretation

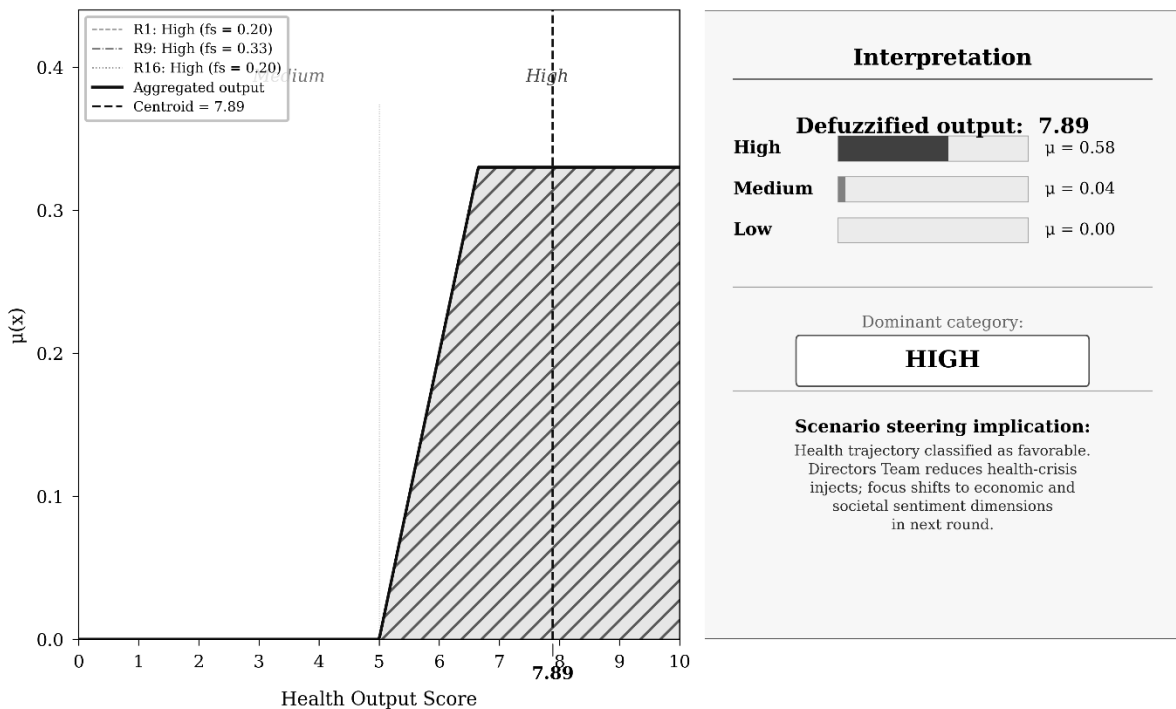


Figure 4. Worked example of the Mamdani fuzzy inference process for the Health subsystem using input values derived from the pilot exercise.

Note. Panel (a): Fuzzification of six input variables (Hygiene practices = 7, Vulnerable population = 3, Infected population = 2, Testing campaign = 8, Mental wellbeing = 6, Health system capacity = 7) with membership degree annotations. Panel (b): Rule activation table showing firing strengths for five selected rules from the 25-rule Health base; rules with firing strength > 0 contribute to the aggregated output. Panel (c): Aggregated output distribution with centroid defuzzification producing a crisp output of 7.89, corresponding to dominant High membership ($\mu = 0.58$). The Directors Team interprets this result as a stable health trajectory, reducing the likelihood of health-crisis injects in the subsequent round. The example represents the target inference logic; TTE1 implementation relied on manual moderator translation.

The rule evaluation step shows that Rules R1 and R9, both associated with High health outcomes, are activated with firing strengths of 0.20 and 0.33 respectively, while Rule R3, representing an escalation pathway, remains inactive because its antecedent conditions are not satisfied. After aggregation and centroid defuzzification, the Health subsystem produces a crisp output value of 7.89, corresponding to dominant High membership ($\mu = 0.58$). The Directors Team interprets this value as indicating a stable health trajectory, which reduces the likelihood of health-related escalation injects in the following round.

Applying the same inference process across all four subsystems generates a multidimensional scenario state profile that can be interpreted holistically. For example, a Health output of 7.89 combined with an Economy output around 4.2 and a Sentiment output around 3.5 indicates favorable epidemiological dynamics alongside emerging societal tension. Such constellations guide targeted inject selection, for instance scenarios related to protest dynamics or institutional trust erosion.

For the Sentiment subsystem, rules such as IF Public trust = Medium AND Non-compliance = Medium AND Panic and fear = Medium THEN Sentiment = Medium and IF Public trust = Low AND Non-compliance = High AND Panic and fear = High AND Fake news = High THEN Sentiment = Low are evaluated against the fuzzified MG inputs. The resulting Sentiment output demonstrates how societal feedback translates into differentiated scenario steering through explicit rule activation.

Project Context

The adaptive mechanism is developed within the European PREPSHIELD project, which follows a three-step cycle (Analyze, Co-create, Evaluate) implemented across pilot sites in Hamburg, Novara (Italy), and Romania. TTEs serve as structured testing environments for refining adaptive mechanisms, digital tools, and stakeholder integration. The integration of adaptivity follows a staged implementation: In TTE1, adaptivity was implemented manually by facilitators applying predefined fuzzy rules. TTE2 introduces a hybrid configuration with platform-generated rule-based suggestions while facilitators retain decision authority. TTE3 extends adaptivity through integration with agent-based simulation components. Across all stages, TTEs function both as crisis preparedness tools and as testing environments for refining adaptive mechanisms.

EVALUATION STRATEGY AND WORK-IN-PROGRESS INSIGHTS

The evaluation strategy builds on the evaluation framework. In line with the work-in-progress character of this study, the evaluation specifies how adaptive mechanisms can be assessed systematically across iterative pilots.

The approach combines structured measurement instruments with preliminary empirical insights from the first pilot exercise (TTE1). Evaluation focuses on the practical feasibility of adaptive mechanisms and the quality of decision processes under dynamically evolving scenario conditions.

Evaluation Dimensions for Adaptive Whole-of-Society Exercises

The evaluation design follows a project-developed framework adapted to the methodological requirements of adaptive tabletop exercises. Decision quality and coordination are treated as a combined dimension because adaptive steering depends on how coordination structures influence interpretation and execution of decisions. A new dimension, adaptive responsiveness, is introduced to evaluate whether rule-based scenario steering operates as intended. Learning outcomes remain aligned with the original framework.

Rather than presenting abstract definitions, the evaluation dimensions operationalize specific observable mechanisms relevant for adaptive exercises:

- Decision-making and coordination quality examines how the CMT interprets evolving information, aligns responsibilities across roles, and produces timely decisions under changing scenario conditions. Indicators include coordination structure activation, clarity of task allocation, and

temporal sequencing of decisions documented through observer protocols and time-stamped artifacts.

- Inclusiveness and whole-of-society engagement evaluates whether societal input is integrated into decision processes. The analysis focuses on the Market Group as a structured representation of societal perspectives. Indicators capture participation intensity, frequency of interaction between MG and CMT, and the degree to which MG signals influence decision trajectories.
- Adaptive responsiveness assesses whether rule-based adaptive steering generates meaningful scenario evolution. Indicators include rule-trigger logs, timing and type of injects, participant reactions, and observable divergence of scenario trajectories across rounds.
- Learning and preparedness outcomes capture perceived realism, learning effects, and anticipated transferability to real-world crisis contexts. Data sources include pre- and post-exercise questionnaires, structured debrief sessions, and reflection prompts aligned with indicators.

This multi-method design integrates behavioral data (logs and artifacts), perceptual data (participant surveys), and process-oriented observations. Behavioral traces document what participants did, surveys capture how they interpreted events, and observations provide contextual understanding of emergent dynamics. Convergence across data sources increases validity, while divergence reveals measurement challenges or context-specific effects. Using identical instruments across pilots supports cross-site comparison between Pilot 1, Pilot 2 and Pilot 3, where institutional structures and trust environments differ.

Pilot Observations: Design Refinement Evidence from the Hamburg Pilot

The Hamburg pilot (TTE1) was conducted as a proof-of-concept within a design science research approach (Hevner, 2010, pp. 9–22). The primary objective was not summative validation of the adaptive mechanism but design evaluation: testing whether the conceptual architecture could be operationalized in a facilitated exercise setting and identifying design requirements for subsequent iterations. Approximately 20 participants - primarily students and early-career professionals simulating institutional and societal roles - addressed a novel influenza scenario across three discussion-based rounds. At this stage, adaptive steering was implemented manually by facilitators applying the predefined rule logic. The following observations are organized along two axes: operational findings relevant for exercise design refinement, and methodological insights pertaining to the measurement and feedback architecture.

The following observations are organized along two axes: operational findings relevant for exercise facilitation, and methodological insights pertaining to the evaluation dimensions introduced in the Section above.

First, the MG remained largely reactive rather than proactive. Structured observer protocols indicate idle periods exceeding 30 percent of total exercise time in two of three rounds, and MG input rarely influenced CMT decision trajectories directly. This finding indicates that societal participation requires structural embedding into decision workflows rather than optional consultation.

Second, the absence of a minimum shared data pack reduced decision consistency. Participants lacked common reference indicators regarding infection dynamics, health system capacity, and societal sentiment. Combined with observable anchoring effects - where participants advocated strict lockdown measures despite scenario information suggesting lower transmissibility (Tversky & Kahneman, 1974) - this produced diverging decision rationales that were difficult to compare across rounds. The observation highlights the need for explicit decision checkpoints and standardized data baselines.

Two methodological insights are particularly relevant for adaptive exercise design.

The first concerns measurement design. The initial operationalization of social variables proved difficult to apply, especially for MG participants with diverse professional and societal backgrounds. The fuzzy logic assessment framework was perceived as conceptually abstract. Participants struggled to translate experiential judgments into predefined scales for trust and compliance. Feedback was therefore expressed primarily in qualitative narratives rather than structured ratings.

As a result, moderators aggregated feedback subjectively and often assigned neutral values. Trust and compliance ratings clustered around mid-range values not because perceptions were neutral but because the measurement format did not align with participants' interpretive frames. The apparent lack of variation therefore reflects a methodological artifact rather than stable social dynamics. This highlights a broader cross-societal challenge. Inclusion alone does not guarantee meaningful societal input. Measurement instruments that are too abstract or technically framed risk obscuring rather than capturing relevant social dynamics.

The second methodological learning concerns visibility of feedback loops. MG assessments were not directly visible to the CMT, which reduced perceived relevance of societal input. This observation led to the redesign of

feedback mechanisms for subsequent exercises. Planned improvements include app-based notifications linking MG input to CMT decision dashboards, visualization of societal sentiment trends, and rule-based inject logic that translates MG assessments into explicit scenario changes.

TTE1 confirms that rule-based adaptive steering is operationally feasible in facilitated exercise settings but requires substantial redesign of measurement interfaces and feedback visibility to produce valid societal input.

TOWARDS PRACTITIONER GUIDANCE: DESIGN IMPLICATIONS AND NEXT STEPS

Practitioner guidance emerges progressively from the staged implementation of adaptive tabletop exercises across the three pilots. The design evolves through iterative testing of adaptive mechanisms, evaluation instruments, and stakeholder interaction formats. TTE1 provides initial empirical grounding and reveals practical constraints related to measurement, participant interpretation, and integration of societal feedback. Subsequent pilots extend this work by testing digitally supported adaptive steering and cross-site comparability.

The resulting practitioner guidance aims to translate design insights into actionable components for exercise planners. These include structured variable selection, rule-based scenario steering procedures, and facilitation approaches that embed societal actors as active contributors to scenario development.

Measurements Refinement and Input Validity

The garbage-in-garbage-out concern is inherent to any inference system operating on human judgment data and is not unique to fuzzy logic. Three design features mitigate this risk. First, triangulation across data sources - observer protocols, participant assessments, and decision logs - reduces dependence on any single input channel. Second, fuzzy membership functions explicitly accommodate measurement imprecision through overlapping categories; unlike crisp classification, a fuzzy system does not amplify small measurement errors into categorical misclassification. Third, the iterative pilot structure enables systematic calibration of both measurement instruments and inference rules against observed exercise dynamics. The measurement challenges identified in TTE1 directly inform instrument redesign for subsequent pilots, including behaviorally anchored rating scales, visual analog scales for continuous assessment, structured observation rubrics with inter-rater calibration, and a shared data pack providing common baseline indicators.

CONCLUSION: TOWARD ADAPTIVE WHOLE-OF-SOCIETY EXERCISE DESIGN

The static scenario assumption underlying conventional tabletop exercises leaves a methodological gap when crises evolve through feedback between governance decisions and societal responses. Iterative refinement across exercises improves design quality but does not address how emerging social dynamics can shape scenario development during an ongoing session. This study conceptualizes adaptivity as a complementary methodological layer that enables structured in-exercise steering based on observable social variables.

The primary contribution lies in reframing the role of societal actors within crisis exercises. In conventional designs, society is treated as an exogenous condition rather than a dynamic variable within the exercise. The adaptive mechanism introduced positions societal actors as causal inputs whose assessments of trust, compliance, and coordination directly influence scenario evolution through transparent rule-based inference. This transformation shifts adaptive steering from facilitator improvisation to an explicit socio-technical process that can be documented and compared across exercise settings.

The first pilot exercise revealed a critical design challenge. Societal participants struggled to express experiential judgments using academically derived measurement constructs. The resulting clustering of mid-range values did not reflect stable perceptions but misalignment between formal modeling assumptions and participants' interpretive frames. Effective whole-of-society adaptivity therefore depends on aligning formal assessment structures with participants' situated reasoning; a calibration challenge that extends beyond interface design to the epistemological assumptions embedded in measurement instruments.

The mechanism has been partially tested in an initial pilot with limited participant diversity. Subsequent exercises in Novara (Italy) and Romania will examine cross-context robustness and progressively introduce hybrid and automated configurations. These pilots will provide empirical evidence on how adaptive steering performs across different governance structures and trust environments.

Beyond the project context, adaptive tabletop exercises offer a methodological bridge between simulation-based modeling and participatory crisis learning. Embedding transparent inference mechanisms into exercises enables systematic observation of how trust shifts, compliance patterns evolve, and coordination structures stabilize or fragment under different interventions. If realized, exercises become experimental spaces for understanding

compound crisis dynamics, not merely rehearsals for known scenarios.

ACKNOWLEDGEMENTS

This research was funded by the European Union's Horizon Europe programme under Grant Agreement No. 101168124 (PREPSHIELD). The views expressed are those of the authors and do not necessarily reflect the position of the European Commission.

We thank all authors, program and local committee members, and volunteers for their hard work and contributions to the ISCRAM conference. The layout of this format was originally adapted by Bartel Van de Walle for the ISCRAM 2005 conference in Brussels from a workshop document template created by Ping Zhang. Thank you, Ping, for your kind permission to use your template. Slight changes have been made by Murray Turoff for ISCRAM 2006 in Newark NJ. Further changes were made by Mark Pfaff for ISCRAM 2014, primarily to conform to Google Scholar indexing criteria. Rob Grace made recent changes, including the adoption of APA Style for ISCRAM 2024. The references cited in this paper are included for illustrative purposes only.

REFERENCES

- Alakrawi, G. A., Al-Wathinani, A. M., Gómez-Salgado, J., Alobaid, A. M., Abahussian, M., Alhazmi, R., Mobrad, A., Jebreel, A., Althunayyan, S., & Goniewicz, K. (2024). Evaluating the efficacy of full-scale and tabletop exercises in enhancing paramedic preparedness for external disasters: A quasi-experimental study. *Medicine*, *103*(49), e40777. <https://doi.org/10.1097/MD.00000000000040777>
- Baumann, F., Landsberg, L., Säger, T., & Mudimu, O. A. (2025). Tabletop exercises as a potential to improve "Preparedness for Society in Health Crises and Disasters". *Procedia Computer Science*, *256*, 101–105. <https://doi.org/10.1016/j.procs.2025.02.101>
- Comfort, L. K. (2007). Crisis Management in Hindsight: Cognition, Communication, Coordination, and Control. *Public Administration Review*, *67*(s1), 189–197. <https://doi.org/10.1111/j.1540-6210.2007.00827.x>
- Emaliyawati, E., Ibrahim, K., Trisyani, Y., Aan Nuraeni, Sugiharto, F., Miladi, Q. N., Abdillah, H., Christina, M., Setiawan, D. R., & Sutini, T. (2025). Enhancing Disaster Preparedness Through Tabletop Disaster Exercises: A Scoping Review of Benefits for Health Workers and Students. *Advances in Medical Education and Practice*, *16*, 1–11. <https://doi.org/10.2147/AMEP.S504705>
- Han, Q., Zheng, B., Cristea, M., Agostini, M., Bélanger, J. J., Gützkow, B., Kreienkamp, J., & Leander, N. P. (2023). Trust in government regarding COVID-19 and its associations with preventive health behaviour and prosocial behaviour during the pandemic: A cross-sectional and longitudinal study. *Psychological Medicine*, *53*(1), 149–159. <https://doi.org/10.1017/S0033291721001306>
- Hevner, A. (2010). *Design Research in Information Systems: Theory and Practice. Integrated Series in Information Systems: Vol. 22*. Springer Science+Business Media LLC. <https://doi.org/10.1007/978-1-4419-5653-8>
- Higashi, S., Ampirit, P., Qafzezi, E., Ikeda, M., Matsuo, K., & Barolli, L. (2024). A comparison study of two implemented fuzzy-based models for decision of logical trust. *Journal of High Speed Networks*, *30*(2), 279–295. <https://doi.org/10.3233/JHS-230189>
- Kruczkiewicz, A., Klopp, J., Fisher, J., Mason, S., McClain, S., Sheekh, N. M., Moss, R., Parks, R. M., & Braneon, C. (2021). Opinion: Compound risks and complex emergencies require new approaches to preparedness. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(19). <https://doi.org/10.1073/pnas.2106795118>
- Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, *7*(1), 1–13. [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2)
- Rubini, E., Trentin, M., Maffi, P., Aammar, B., Gaievskiy, S., Bahattab, A., Lamine, H., Kordi-Kaiser, M., Staub, K., & Ragazzoni, L. (2026). Challenges in healthcare facilities' response to past outbreaks: A systematic review of reviews. *BMC Health Services Research*, *26*(1), 141. <https://doi.org/10.1186/s12913-025-13934-9>
- Ruggiero, A., Piotrowicz, W. D., & John, L. (2024). Enhancing societal resilience through the whole-of-society approach to crisis preparedness: Complex adaptive systems perspective – The case of Finland.

- International Journal of Disaster Risk Reduction*, 114, 104944.
<https://doi.org/10.1016/j.ijdr.2024.104944>
- Sahin, O., Salim, H., Suprun, E., Richards, R., MacAskill, S., Heilgeist, S., Rutherford, S., Stewart, R. A., & Beal, C. D. (2020). Developing a Preliminary Causal Loop Diagram for Understanding the Wicked Complexity of the COVID-19 Pandemic. *Systems*, 8(2), 20. <https://doi.org/10.3390/systems8020020>
- Seyd, B., & Bu, F. (2022). Perceived risk crowds out trust? Trust and public compliance with coronavirus restrictions over the course of the pandemic. *European Political Science Review*, 14(2), 155–170. <https://doi.org/10.1017/S1755773922000078>
- Shanka, M. S., & Menebo, M. M. (2022). When and How Trust in Government Leads to Compliance with COVID-19 Precautionary Measures. *Journal of Business Research*, 139, 1275–1283. <https://doi.org/10.1016/j.jbusres.2021.10.036>
- Stivas, D., & Cole, A. (2023). The importance of trust and transparency in managing the COVID-19 pandemic. Evidence from sixteen EU member states. *Journal of Contemporary European Studies*, 1–18. <https://doi.org/10.1080/14782804.2023.2193733>
- Strelkovskii, N., & Rovenskaya, E. (2021). Causal Loop Diagramming of Socioeconomic Impacts of COVID-19: State-of-the-Art, Gaps and Good Practices. *Systems*, 9(3), 65. <https://doi.org/10.3390/systems9030065>
- Sydnés, M., Kjeve, M., & Sydnés, A. K. (2025). Interorganizational coordination during emergencies and crises. *International Journal of Disaster Risk Reduction*, 121, 105310. <https://doi.org/10.1016/j.ijdr.2025.105310>
- Thalmeiner, G., Gáspár, S., & Zéman, Z. (2025). Developing a fuzzy management control model for the performance evaluation of the company digitalization level. *Journal of Open Innovation: Technology, Market, and Complexity*, 11(4), 100633. <https://doi.org/10.1016/j.joitmc.2025.100633>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science (New York, N.Y.)*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Vonglao, P. (2017). Application of fuzzy logic to improve the Likert scale to measure latent variables. *Kasetsart Journal of Social Sciences*, 38(3), 337–344. <https://doi.org/10.1016/j.kjss.2017.01.002>
- Wolbers, J., Groenewegen, P., Mollee, J., & Bím, J. (2013). Incorporating Time Dynamics in the Analysis of Social Networks in Emergency Management. *Jhsem*, 10(2), 555–585. <https://doi.org/10.1515/jhsem-2013-0019>
- Wu, C., Bierman, A., & Schieman, S. (2022). Socioeconomic stratification and trajectories of social trust during COVID-19. *Social Science Research*, 108, 102750. <https://doi.org/10.1016/j.ssresearch.2022.102750>
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Žilinskas, R., & Sakalauskas, L. (2021). Exploring crisis governance: Quest for functional resilience during COVID-19 in Lithuania. *Scientific Journal of the Military University of Land Forces*, 202(4), 763–798. <https://doi.org/10.5604/01.3001.0015.6180>