*Digital Health and Informatics Innovations for Sustainable Health Care Systems*
*J. Mantas et al. (Eds.)*


*doi:10.3233/SHTI240797*

# Supporting the Episodic Application of Clinical Guidelines over Significant Time Periods

Bruria BEN SHAHAR[a,1], Yuval SHAHAR[a], Shai JAFFE[b,c], Odeya COHEN[b,c], Erez SHALOM[a], Maya SELIVANOVA[d], Ephraim RIMON[d] and Ayelet GOLDSTEIN[e]

[a] *Department of Software and Information Systems Engineering, Ben-Gurion University, Israel*
[b] *Department of Nursing, Faculty of Health Sciences, Ben-Gurion University, Israel*
[c] *Nursing Informatics lab, Faculty of Health Sciences, Ben-Gurion University, Israel*
[d] *Herzfeld Geriatric Rehabilitation Medical Center, Clalit Health Services, Israel*
[e] *Department of Computer Science, Hadassah Academic College Jerusalem, Israel*

ORCiD ID: Bruria BEN SHAHAR https://orcid.org/0009-0003-2631-9132,
Yuval SHAHAR https://orcid.org/0000-0003-0328-2333,
Shai JAFE https://orcid.org/0009-0000-9996-0367,
Odeya COHEN https://orcid.org/0000-0002-2427-6381,
Erez SHALOM https://orcid.org/0000-0002-3471-3143,
Ayelet GOLDSTEIN https://orcid.org/0000-0003-0794-3653

**Abstract.** Medical errors contribute significantly to morbidity and mortality, emphasizing the critical role of *Clinical Guidelines* (GLs) in patient care. Automating GL application can enhance GL adherence, improve patient outcomes, and reduce costs. However, several barriers exist to GL implementation and real-time automated support. Challenges include creating a formalized, machine-comprehensible GL representation, and an *episodic* decision-support system for sporadic treatment advice. This system must accommodate the non-continuous nature of care delivery, including partial actions or partially met treatment goals. We describe the design and implementation of an episodic GL-based clinical decision support system and its retrospective technical evaluation using patient records from a geriatric center. Initial evaluation scores of the e-Picard system were promising, with a mean 94% correctness and 90% completeness based on 50 random pressure ulcer patients. Errors were mainly due to knowledge specification, algorithmic issues, and missing data. Post-corrections, scores improved to 100% correctness and a mean 97% completeness, with missing data still affecting completeness. The results validate the system's capability to assess guideline adherence and provide quality recommendations. Despite initial limitations, we have demonstrated the feasibility of providing, through the e-Picard episodic algorithm, realistic medical decision-making support for noncontinuous, intermittent consultations.

**Keywords.** Clinical Guidelines, CDSSs, Episodic support, Chronic Patients, Quality Assessment, Quality Assurance

[1] Corresponding Author: Bruria Ben Shahar; E-mail: bruriabe@bgu.ac.il.

## 1. Introduction

Clinical Decision Support Systems (CDSSs) based on evidence-based *clinical guidelines* (GLs), such as the Picard engine [1] enhance medical decision-making in hospitals [2] and home-based care, as shown in the EU MobiGuide project [3]. These systems provide continuous real-time guidance and personalized treatment recommendations to chronic patients and their care providers, improving GL adherence, patient outcomes, and cost efficiency [3]. Intelligent monitoring methods like *Knowledge-Based Temporal Abstraction* (KBTA) [4], transform raw clinical data into meaningful concepts, aiding both humans and CDSS in understanding patient medical history and treatment needs.

CDSSs use in hospitals is often intermittent, requiring an adaptive CDSS model for irregular consultation patterns. We propose an episodic CDSS—e-Picard—which extends the continuous Picard model, to evaluate care quality intermittently and provide timely recommendations. At each consultation session, the CDSS assess care quality and advises on continuing care, considering GL adherence and what should be performed next. Addressing incomplete implementation of GL-recommended actions is crucial, requiring re-evaluation in subsequent consultations. To tackle the challenge of assessing medical decision-making quality, particularly with the frequent occurrence of these partial implementations, we employ fuzzy logic [5], which manages value and time-related uncertainties, facilitating nuanced assessment of treatment actions and flexible interpretation of GLs. This fuzzy logic-informed framework, previously evaluated for retrospective quality-assessment [6], is integral to our episodic CDSS, enhancing its capability to evaluate and guide GL adherence progressively.

We retrospectively processed medical records from a geriatric center, focusing on pressure ulcers (PUs) and diabetes management, to analyze the episodic CDSS quality assessments and recommendations. Future research will include a large-scale retrospective assessment of gaps between actual care and validated CDSS recommendations, optimal consultation frequencies and real-time provision of episodic GL-based recommendations to staff.

## 2. Methods

### *2.1 Computational Methods*

We acquire the guideline's knowledge from the medical team, including both *declarative* ["What Is"] knowledge (e.g., what is systolic High Blood Pressure), *procedural* ["How To"] knowledge (e.g., how to administer regular insulin, regarding route, dosage, and frequency) and Quality Assessment (QA) knowledge, which involves (1) treatment intentions ( [care] process and [patient] outcomes), (2) conditions for still performing incomplete or missed actions, and (3) defining [0..1] QA scores for each action using fuzzy temporal logic to address GL ambiguities, particularly in timing and value-based decisions. The default fuzzy membership function is linear. For example, administering insulin in a diabetes management GL involves three constraints: Frequency (e.g., three times daily), Time-Gap (e.g., within 40 minutes post-meal), and Value (e.g., 4-8 units). Each constraint can have full, partial or no compliance. Full-compliance occurs if insulin is given within 40 minutes post-meal, partial between 40 minutes to an hour and none after one hour. Compliance levels are represented as points in a trapezoid: A = (0,1), B

= (40,1), C = (60,0). Between A and B, y = f(x) = 1; between B and C, y = $-\frac{1}{20}x + 3$. For instance, insulin given 57 minutes after a meal, has a score of $-\frac{1}{20} * 57 + 3 = 0.15$ .

Our system maintains continuous support when online and switches to an episodic mode during offline periods, as illustrated in Figure 1.

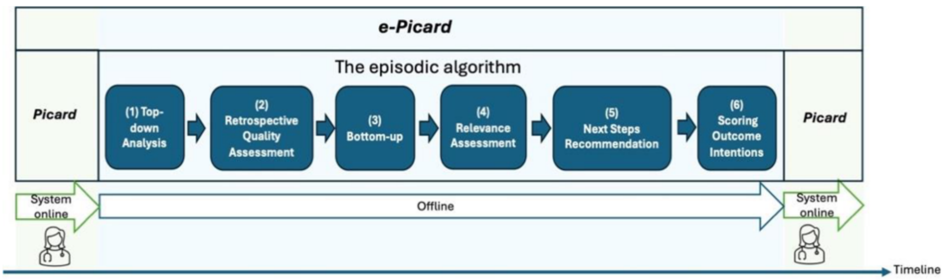


**Figure 1.** Flowchart of the Episodic Guideline Application Algorithm: (1) *Top-down analysis*, determining what should have been done and when; (2) *Retrospective Quality Assessment* to scores actions, using fuzzy temporal logic, allowing for partial performance; (3) *Bottom-up* review identifying unnecessary actions; (4) *Relevant Actions Assessment* identifying relevant actions - past actions still relevant and immediate required actions; (5) *Next Steps Recommendation*, recommending current and future guideline-based actions; (6) *Scoring Outcome intentions* evaluating the fulfillment of the guideline's intentions regarding patient's state.

In the *top-down analysis*, the system assesses required actions from the GL considering the patient's record up to the point of CDSS consultation, including periods without consultation. The retrospective *quality assessment* then assigns a QA performance score to each action. A *bottom-up* review identifies unnecessary actions, followed by a *relevance assessment* evaluating the need for repeating actions due to previous omission or incomplete execution. The GL's knowledge base defines *conditions* for these repetitions and outlines actions that must continue uninterrupted, as detailed in Figure 2's pseudo-code. In *next steps recommendations,* we provide context-specific recommendations for current and future actions as advised by the GL (e.g. "action should be repeated three times a day over the next five days"). Finally, in *scoring outcome intentions,* achievement of the patient's *outcome* goals is evaluated, informing clinicians of achieved objectives or care deficiencies. After each session, the episodic CDSS compiles a report detailing these insights.

```
Find_Relevant_Actions(QA_Actions)

relevant_actions ← ∅
Foreach action ∈ QA_Actions
    if(action.QA_score < action.threshold_to_repeat)//if QA fuzzy score less than threshold
        if(ask_if_exist_now(action.extensive_filter_conditions_reperformed))//check if still valid
            relevant_actions.add(action)
    else
        if(action.is_rigid)//rigid action cannot be stopped in the middle of execution
            relevant_actions.add(action)
return relevant_actions //actions relevant NOW
```

**Figure 2.** The Pseudo code of the Episodic GL application algorithm's "*Find Relevant Actions*" module, whose objective is to find the currently [still] relevant actions.

### *2.2 Evaluation Methods*

The efficacy of e-Picard was evaluated using historical patient data for PUs at a Geriatric Medical Center in Israel, with IRB approval (#0114-21-KMC). PUs were chosen due to their prevalence and the care complexities they present in geriatric patients. In the technical assessment phase, the system's effectiveness was evaluated across crucial clinical scenarios, with approximately ten patients per scenario over the past three years, totaling around 50 patients. With the clinical staff, we identified scenarios for *Monitoring* (Norton scale > 14), *Prevention* (Norton scale < 14, no PUs), *Treatment* (presence of PUs of three severity levels), and *Transition* (from prevention to treatment). We measured performance using real patient data, calculating *correctness* [soundness of the system's recommendations, given the GL and patient data] and *completeness* [proportion of relevant GL-based recommendations made] of retrospective QA scores and recommendations. Through iterative updates, we refined the system as needed.

## 3. Results and Discussion

Results for correctness and completeness, our primary concerns regarding the validity of the system's assessments and recommendations, are detailed in Table 1. After receiving the initial evaluation results, we conducted an error analysis to understand the system's performance. The analysis revealed some errors due to incorrect action definitions during the knowledge acquisition phase; for example, actions intended for daily performance were erroneously set for a longer term in the Prevention stage. After this analysis, we corrected the algorithm's code significantly improving correctness and increasing completeness. However, achieving full completeness was not possible due to incomplete patient records, such as absent Norton scale documentation, leading to recommendation failures, due to the system's high reliance on such data.

**Table 1.** Correctness and Completeness (in %) of the episodic-application algorithm's assessments and recommendations for each stage in the pressure-ulcer guideline, Before (B) [left column] and After (A) [right column] the error analysis in which several knowledge-base and algorithmic errors were fixed. In the *Weighted QA Score*, only correctness is calculated, as completeness is not relevant in this context.

| **Stage** | **Fully Actions** | | **Missing Actions** | | **Current Relevant Actions** | | | | **Weighted QA Score** | | **Average Score** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cr | Cm | Cr | Cm | Cr | | Cm | | Cr | | Cr | | Cm | |
| | | | | | B | A | B | A | B | A | B | A | B | A |
| F-U | 100 | 100 | 100 | 100 | 83 | 100 | 64 | 92 | 90 | 100 | 93 | 100 | 88 | 97 |
| P | 100 | 100 | 100 | 100 | 84 | 100 | 71 | 92 | 62 | 100 | 87 | 100 | 90 | 97 |
| T | 100 | 100 | 100 | 100 | 90 | 100 | 80 | 88 | 100 | 100 | 98 | 100 | 93 | 96 |
| P+T | 100 | 100 | 100 | 100 | 94 | 100 | 71 | 98 | 100 | 100 | 99 | 100 | 90 | 99 |
| WA | 100 | 100 | 100 | 100 | 88 | 100 | 71 | 92 | 88 | 100 | 94 | 100 | 90 | 97 |

F-U: Follow-Up; P: Prevention; T: Treatment; P+T: Prevention and Treatment; WA: Weighted Average; Cr: Correctness; Cm: Completeness; B: Before; A: After.

We developed an episodic algorithm for intermittent CDSS use by medical and nursing staff, capable of calculating partial-adherence scores using fuzzy temporal logic, and evaluated it in PU management. Initial results were highly encouraging. Previous surveys of guideline-based CDSSs [7] and recent updates regarding the knowledge represented by them [8] typically ignore the aspects of intermittent care and partial adherence.

With respect to limitations, one limitation is that the system was evaluated using a GL focused on a single disease (PUs), without handling additional comorbidities. However, the GL we used was a complex one, with multiple stages and actions, and can be expanded to include additional comorbidities (e.g., treating PUs in diabetic patients would add another sub-plan), overcoming the limitation of a single morbidity. In addition, data protection regulations restricted us to structured data, omitting free-text medical summaries. We expect to use NLP for text access in the future. Despite these limitations, the preliminary evaluation indicates a potential for improvement in compliance to the GL, and thus for enhanced treatment outcomes. The next, functional evaluation will be on a larger scale, but "silent" [i.e., without affecting care] running e-Picard on 1000 patient records collected over three years, assessing the correctness and completeness of the actual care, when compared to the system's recommendations. This phase will also assess the system's cost-effectiveness and determine optimal consultation frequency (e.g., once every three days). The final functional evaluation will present real-time recommendations to clinicians, benchmarked against a control group not using CDSS. This method, previously used with simulated cases for the Picard system, will test our system's effectiveness in a clinical setting.

## 4. Conclusions

The e-Picard episodic algorithm is a significant advance in healthcare technology, supporting realistic medical decision-making through intermittent consultations. Despite initial limitations, our technical evaluation shows the system's ability to provide accurate retrospective quality assessment scores and GL-based recommendations. Upcoming silent and functional evaluations will further validate its effectiveness, cost-efficiency, and optimal usage frequency. Continual refinements through these stages are expected to optimize our episodic CDSS, ultimately enhancing patient care outcomes.

**Acknowledgements:** This study was supported by the Israeli National Health Policy research grant No. 2020/284, and by Israel Precision Medicine Partnership (IPMP) grant No. 3543/21.

## References

[1] Shalom E, Shahar Y, Lunenfeld E. An architecture for a continuous, user-driven, and data-driven application of clinical guidelines and its evaluation. J Biomed Inform. 2016;59:130-148.
[2] Shalom E, Shahar Y, Parmet Y, Lunenfeld E. A multiple-scenario assessment of the effect of a continuous-care, guideline-based decision support system on clinicians' compliance to clinical guidelines. Int J Med Inform. 2015;84(4):248-262.
[3] Peleg M, Shahar Y, Quaglini S, et al. Assessment of a personalized and distributed patient guidance system. Int J Med Inform. 2017;101:108–130.
[4] Shahar Y. A framework for knowledge-based temporal abstraction. Artif Intell. 1997;90(1–2):79-133.
[5] Zadeh LA. Fuzzy algorithms. Inf Control. 1968;12(2):94-102.
[6] Shalom E, Goldstein A, Wais R, et al. Implementation and Evaluation of a System for Assessment of The Quality of Long-Term Management of Patients at a Geriatric Hospital. arXiv:2211.12904.
[7] Peleg M. Computer-interpretable clinical guidelines: A methodological review. J Biomed Inform. 2013;46(4):744-763.
[8] Riaño D, Peleg M, ten Teije A. Ten years of knowledge representation for health care (2009–2018): Topics, trends, and challenges. Artif Intell Med. 2019;100:101713.