MP08-09

EVALUATING CHATGPT'S ADHERENCE TO CONTEMPORARY CLINICAL GUIDELINES AMID OUTDATED LEGACY INFORMATION: IMPLICATIONS FOR MEDICAL ACCURACY IN AI TOOLS

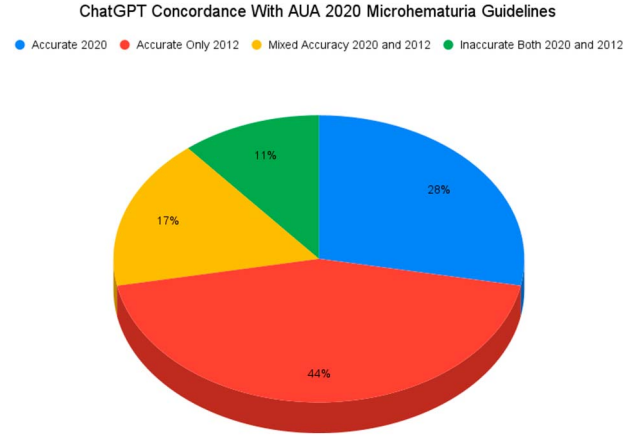
Yitzchak E. Katlowitz, Brooklyn, NY; Kalman Katlowitz, Houston, TX; Nachum Katlowitz, Staten Island, NY*

INTRODUCTION AND OBJECTIVE: The use of generative AI modalities for queries pertaining to medial care is rising from both medical practitioners and the general population. Fortunately, there are often clinical guidelines readily available and accessible to the AI algorithms. These clinical guidelines often undergo periodic updates to incorporate new evidence and improve patient care. As a result, outdated guidelines remain widely accessible, posing a risk for inquiry modalities that cannot properly differentiate between old and new recommendations. The American Urologic Associate (AUA) guidelines for microhematuria underwent an update in 2020 from the 2012 version. Notably, the 2020 guidelines introduced risk stratification for malignancy, whereas the 2012 guidelines broadly recommended diagnostic comprehensiveness via CT urography and cystoscopy for all patients over 35 years without regard for cancer risk. This study evaluates ChatGPT's ability to deliver accurate recommendations consistent with the latest AUA microhematuria guidelines from 2020 as opposed to the 2012 version.

METHODS: ChatGPT was prompted with clinical scenarios and general questions pertaining to microhematuria, specifically instructing the model to respond based upon only the 2020 guidelines. If responses lacked completeness, additional prompts were used to elicit more detail. Responses were categorized into four groups: accurate based upon 2020 guidelines, accurate only based upon 2012 guidelines, accuracy based upon a mix of 2020 and 2012 guidelines, and inaccurate.

RESULTS: Accurate responses based upon 2020 guidelines were observed in 28% of cases, while 44% of the responses were only accurate based upon 2012 guidelines. Responses that had a mix of 2020 and 2012 based recommendations appeared in 17% of cases, and inaccurate recommendations accounted for 11% of case responses.

CONCLUSIONS: While ChatGPT had an impressive overall accuracy of approximately 89% concordance with an official guideline, our findings findings suggest ChatGPT frequently references older guidelines, even when explicitly directed to use updated information. This highlights the potential limitations of generative AI in clinical decision support and the need for caution when using AI, in particular for frequently updated and studied information.



Source of Funding: None

MP08-10

REAL WORLD EVIDENCE OF UROMONITOR FOR NMIBC FOLLOW UP BY THE CUETO GROUP; CENTRALIZED VERSUS LOCAL PERFORMANCE

Guerrero-Ramos Félix, Madrid, Spain; Rubio-Briones Jose, Valencia, Spain; Mercadé Asier, Barcelona, Spain; Miqueleiz Legaz Maite, Pamplona, Spain; Martín Rodríguez Ruth, Madrid, Spain; Alcaraz Antonio, Barcelona, Spain; Bezana Abadía Ines, Pamplona, Spain; Lozano Fernando, Barcelona, Spain; Arrabal Miguel, Granada, Spain; Luis Cardo Adrián, Toledo, Spain; Gonzalo Rodríguez Victoria, Burgos, Spain; Caño Velasco Jorge, Serrano Liesa Miriam, Madrid, Spain; Ferreiro Pareja Cristina, Barcelona, Spain; Burbano Marta, Zaragoza, Spain; Gonzalez Valcárcel Ignacio, Madrid, Spain; Sánchez Zalabardo Daniel, Pamplona, Spain; Villacampa Felipe, Madrid, Spain; Velarde Muñoz Consolación, Cádiz, Spain; Guijarro Ana, Madrid, Spain; Ledo Maria Jose, Cádiz, Spain; Gallioli Andrea, Barcelona, Spain; Sánchez González Alvaro, Madrid, Spain; López González A. Jose, Valencia, Spain; Jalón Monzón Antonio, Oviedo, Spain; Saucedo Jiménez Inés, Jerez De La Frontera, Spain; Moreillo Leticia, Albacete, Spain; Marzullo J. Leopoldo, Valencia, Spain; Campos Sañudo A. Jose, Cantabria, Spain; Bordalba Gómez R. Joan, Lleida, Spain; Mengual Lourdes, Barcelona, Spain; Borque-Fernando Ángel, Esteban M. Luis, Zaragoza, Spain; Palou Joan, Barcelona, Spain; J.A Witjes, Nijmegen, Netherlands; Martínez-Piñeiro Luis, Madrid, Spain*

INTRODUCTION AND OBJECTIVE: Part of the high cost of managing NMIBC is due to the high number of cystoscopies performed for follow-up. Uromonitor®, a urine biomarker test that detects hotspot mutations in three genes (TERT, FGFR3, and KRAS) for the evaluation of disease recurrence. Our main objective was to perform an external validation of its sensitivity (S), specificity (SP), PPV and NPV for the detection of recurrences by WHO 04/16 grade in a real world scenario. As a secondary objective, we wanted to compare the results in a reference laboratory to those obtained in each participating Center.

METHODS: We invited all CUETO Group members to enrol patients in a competitive manner. Inclusion criteria were restricted to any risk group NMIBC once in a FU setting and the initial tumours had to be previously diagnosed over the last 3 months to 2 years. Uromonitor® and urinary cytology were obtained before follow-up cystoscopies. The study had two phases; In phase I, the urine filtration protocol proposed by the supplier of the biomarker kit was followed with centralized analysis in the reference laboratory of IPATIMUP (Porto, Portugal). In phase II, Uromonitor® was obtained from direct urine and processed in each of the participating Centers. Follow-up was intended to be for 2 years. Follow-up was intended to be for 2 years.

RESULTS: 1,178 patients were analyzed, 652 and 526 in phase I and II respectively. Initial tumours had been low or high grade (WHO 04/16) in 50% and 46%, less than T1 and T1 in 59% and 36%, and 12% had concomitant carcinoma in situ. Cystoscopy was reported as normal, with recurrence or doubtful in 993 (84.3%), 136 (11.5%) and 49 (4.2%). When comparing Phase I and II, Uromonitor® was invalid in 0.6% and 6.1% of the urine samples, negative in 91% and 84% and positive in 9 and 10% respectively. TERTp and FGFR3 mutations were detected in 3.9% and 7.2% and 6.1% and 3.6% of Uromonitor® in phase I and II respectively, with no KRAS mutations found. Uromonitor® results and grade distribution of the recurrences in the two phases are reflected in the Table 1.

CONCLUSIONS: The high NPV of the test in HG would justify further investigation with a molecular technique that improved S in the follow-up of NMIBT. KRAS adds nothing to its positivity. There are no relevant differences in positivity of the Uromonitor® results depending on whether it is carried out in the central laboratory or in that of each Center, although the detected mutation pattern shows some differences.

	S	Sp	PPV	NPV	Accuracy	AUC
Low grade	0.32	0.93	0.26	0.95	0.89	0.63
High grade	0.44	0.93	0.18	0.98	0.91	0.69
Phase I	0.35	0.93	0.35	0.93	0.88	0.64
Phase II	0.39	0.92	0.38	0.93	0.87	0.66
Whole series	0.36	0.93	0.36	0.93	0.87	0.65

Source of Funding: Pharmalink and Stella Maris

MP08-11
CLINICAL UTILITY OF A URINARY BIOMARKER (CXBLADDER TRIAGE) COMPARED TO STANDARD OF CARE FOR MICROSCOPIC HEMATURIA EVALUATIONS IN A LARGE INDEPENDENT DELIVERY NETWORK

Ronald K. Loo, Huntington Beach, CA; Christopher Filson, Los Angeles, CA; Jeffrey M. Slezak, Pasadena, CA; Jay Jhaveri, Tamer Aboushwareb, Hummelstown, PA*

INTRODUCTION AND OBJECTIVE: Hematuria evaluation remains a costly and burdensome process for both patients and physicians due to the sheer number of patients, invasive nature of examination (cystoscopy) and complex logistics. We sought to retrospectively evaluate real-world utility of adding the mRNA-based Cxbladder Triage urinary biomarker test to our microscopic hematuria diagnostic workflow.

METHODS: Between September 2021 – September 2024, of 5242 patients that received the Cxbladder triage test within Kaiser Permanente Southern California, a large integrated healthcare system, 2999 met study specific inclusion criteria and were included for analysis. A non-tested, 1:1-matched comparable cohort (based on date and hematuria risk index) of n = 2754 was identified for comparison. Patients were divided into low probability of urothelial cancer (UC), and physician directed protocol groups by the triage test. Patients were followed prospectively to measure the number of procedures in each arm (cystoscopy, and CT Urography) as well as number of UC diagnoses within 1 year after presentation.

RESULTS: Rates of cystoscopy were much lower (3.9%) in those with a low-risk Triage result compared to their matched comparison group without a test (42.5%), while those whose Triage result recommended physician-directed protocol had a much higher rate (72.9%) than their matched comparison group (41.4%).

CONCLUSIONS: These data reveal that CxBladder Triage testing resulted in significantly decreased cystoscopy and imaging utilization in those classified as low risk, while simultaneously demonstrating increases in the cystoscopy and bladder cancer detection rate in the physician directed protocol group. Further systematic evaluations of the Triage test are warranted to reduce the burden of unnecessary cystoscopy.

Low probability of UC n=2174			Physician directed protocol (higher probability) n=580		
Evaluated Parameters	*Triage*	*SOC*	*Evaluated Parameters*	*Triage*	*SOC*
Age (mean)	*59.8*	*61.7*	*Age (mean)*	*67.5*	*65.9*
Hematuria risk index (mean)	*3.8*	*3.9*	*Hematuria risk index (mean)*	*5.1*	*5.2*
Cystoscopy Rate (6 month)	84 (3.9%)	923 (42.5%)	*Cystoscopy Rate (6 month)*	423 (72.9%)	240 (41.4%)
CT Abdomen / Pelvis Rate	168 (7.7%)	340 (15.6%)	*CT Abdomen / Pelvis Rate*	115 (19.8%)	107 (18.5%)
Bladder Cancer Diagnosis Rate (1 year)	2 (0.1%)	29 (1.3%)	*Bladder Cancer Diagnosis Rate (1 year)*	16 (2.8%)	13 (2.2%)

Source of Funding: Ronald Loo and Christopher Filson are full time employees of KP - SCMPG with no other funding related to this study. Jeff Slezak reports funding paid to his institution by Pacific Edge related to the submitted work, and funding paid to his institution by Pfizer, Inc. and Dynavax Technologies, Inc., unrelated to the submitted work. Jay Jhaveri and Tamer Aboushwareb are full time employees of Pacific Edge Cancer Diagnostics

MP08-12
BLADDER EPICHECK TRIGGERED PHOTODYNAMIC DIAGNOSIS BIOPSIES DETECT HIGH GRADE RECURRENCES MISSED BY WHITE LIGHT CYSTOSCOPY

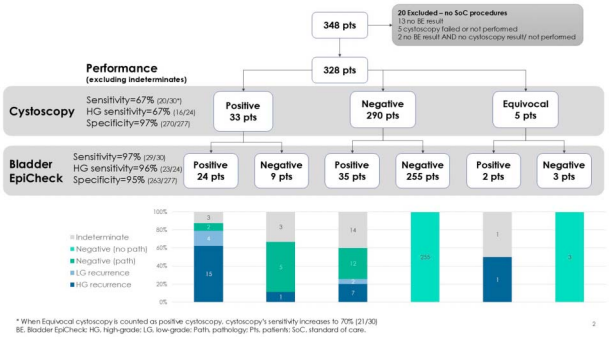
Paramananthan Mariappan, Suriaraj Karppaya, Tanya Lord-Mckenzie, Edinburgh, United Kingdom*

INTRODUCTION AND OBJECTIVE: White-light cystoscopy (WLC) has moderate sensitivity (70%) and misses ~50% of CIS recurrences. PhotoDynamic Diagnosis (PDD) guided biopsies detects some of these missed tumours. Bladder EpiCheck (BE) is a urine marker with high HG sensitivity (91%) and specificity (84%) and a strong anticipatory positive signal. We introduced BE into our routine high-risk NMIBC (HR-NMIBC) surveillance and analysed the performance of BE and WLC against PDD-guided biopsy as the diagnostic benchmark.

METHODS: As part of an NHS Quality Improvement project within our Effectiveness and Efficiency programme, all HR-NMIBC patients on surveillance in a tertiary centre underwent both WLC and BE from Jul'23, to explore an opportunity to replace WLC with BE. Voided urine was collected immediately prior to WLC and sent to a certified lab for BE testing. Cystoscopists were blinded to the BE result. Positive WLC triggered a resection/biopsy under general anesthesia (GA). Negative WLC but positive BE triggered PDD-guided biopsy under GA.

RESULTS: From 348 HR-NMIBC patients undergoing surveillance Jul'23-Aug'24, 328 were suitable for analysis (see Figure 1). 30(9.1%) patients had a pathologically proven recurrence: 24(7.3%) HG (10 CIS alone) and 6(1.8%) LG, 277(84.5%) were recurrence free (NEG) and 21(6.4%) are still indeterminate (9 pending GA, the rest unable/unwilling to undergo GA). BE detected 23(95.8%) HG, 9(90%) CIS and 6(100.0%) LG tumours. WLC detected 16 (66.7%) HG, 4 (40%) CIS and 4(66.7%) LG and was equivocal in 1(4.2%) HG. Of the 277 NEG patients, BE and WLC were negative in 263 (94.9%) and 270 (97.5%), respectively. BE PPV was 67.4%, both NPV and HG NPV were 99.6%. WLC PPV was 74.1%, NPV was 96.3% and HG NPV was 98.5%.

CONCLUSIONS: BE demonstrated superior sensitivity over WLC with 44% more HG recurrences detected, without compromising specificity. PDD-guided biopsies of BE positive patients who are WLC negative, unveil the true performance of BE and WLC, particularly in CIS. Performing BE alone at surveillance visits, followed by PDD-guided biopsy in BE positives, could reduce 95% of unnecessary cystoscopies while detecting 96% of HG disease. This evidence suggests significant utility in surveillance programmes by effectively and efficiently detecting high grade recurrence earlier, particularly CIS, which also translates to opportunities for bladder preservation.



Source of Funding: None