

1 **Large-Scale Longitudinal Comparison of Urine Cytological Classification Systems Reveals**
2 **Potential Early Adoption of The Paris System Criteria**

3
4 Joshua J. Levy PhD^{1,2,3,4,*}, Xiaoying Liu MD^{1,6}, Jonathan D. Marotti MD^{1,6}, Darcy A. Kerr
5 MD^{1,6}, Edward J. Gutmann MD, AM^{1,6}, Ryan E. Glass MD⁵, Caroline P. Dodge⁶, Louis J.
6 Vaickus MD, PhD^{1,6}

- 7 1. Emerging Diagnostic and Investigative Technologies, Department of Pathology and
8 Laboratory Medicine, Dartmouth Hitchcock Medical Center, Lebanon, NH, 03766
- 9 2. Department of Dermatology, Dartmouth Hitchcock Medical Center, Lebanon, NH, 03766
- 10 3. Department of Epidemiology, Dartmouth College Geisel School of Medicine, Hanover,
11 NH, 03756
- 12 4. Program in Quantitative Biomedical Sciences, Dartmouth College Geisel School of
13 Medicine, Hanover, NH, 03756
- 14 5. UPMC East, Pittsburg, PA, 15146
- 15 6. Dartmouth College Geisel School of Medicine, Hanover, NH, 03756

16
17 * To whom correspondence should be addressed: joshua.j.levy@dartmouth.edu

18
19 **Author Contributions**

20 JL and LV: conceptualization, formal analysis, funding acquisition, investigation, methodology,
21 project administration, resources, software, supervision, validation, visualization, writing -
22 original draft; XL, JM, DK, EG, RG, CD, LV: data curation; all authors: writing - review and
23 editing

24 **Abstract**

25 Urine cytology is used to screen for urothelial carcinoma in patients with hematuria or risk
26 factors (e.g., smoking, industrial dye exposure) and is an essential clinical triage and longitudinal
27 monitoring tool for patients with known bladder cancer. However, urine cytology is semi-
28 subjective and thus susceptible to issues including specimen quality, inter-observer variability,
29 and “hedging” towards equivocal (“atypical”) diagnoses. These factors limit the predictive value
30 of urine cytology and increase reliance on invasive procedures (cystoscopy). The Paris System
31 for Reporting Urine Cytology (TPS) was formulated to provide more quantitative/reproducible
32 endpoints with well-defined criteria for urothelial atypia. TPS is often compared to other
33 assessment techniques to justify its adoption. TPS results in decreased use of the atypical
34 category and better reproducibility. Previous manuscripts comparing diagnoses pre- and post-
35 TPS lacked a longitudinal component and thus failed to consider temporal differences between
36 diagnoses made under prior systems and TPS. By aggregating across time, studies may
37 underestimate the magnitude of differences between assessment methods. We conducted a large-
38 scale longitudinal reassessment of urine cytology using TPS criteria from specimens collected
39 from 2008 to 2018, prior to the mid-2018 adoption of TPS at an academic medical center.
40 Findings indicate that differences in atypical assignment were largest at the start of the period
41 and these differences progressively decreased towards insignificance just prior to TPS
42 implementation. This finding suggests that cytopathologists had begun to utilize the quantitative
43 TPS criteria prior to official adoption, which may more broadly inform adoption strategies,
44 communication, and understanding for evolving classification systems in cytology.

45

46 **Introduction**

47 Urothelial carcinoma is a cancer that demands attention because it is both common (7th leading
48 cause of malignancy) and has the highest recurrence rate of any cancer (70%)^{1,2}. Diagnostic
49 protocols and clinical triage are based on a combination of urine cytology, cystoscopy +/- biopsy,
50 and immunocytochemical and molecular studies, with resultant longitudinal follow-up for
51 negative and atypical findings. While cystoscopy followed by biopsy with histological
52 examination is the accepted standard for diagnosis, urine cytology is the preferred mode for
53 detection and screening because: it is inexpensive; has a rapid turn-around time; and specimens
54 are relatively easy to obtain. Abnormal cytology findings usually trigger more invasive and
55 expensive assessment methods (e.g., cystoscopy, FISH)^{4,5}, though further investigation is
56 important because these assessments determine treatment options, ranging from follow-up urine
57 cytology to chemotherapy or surgery. This sequence is not entirely foolproof since abnormal
58 cytological findings are subject to a high “false positive” rate (i.e., urine specimen is perceived to
59 be abnormal while cystoscopy/biopsy leads to negative finding). Cytology-based examination is
60 the currently accepted method for the detection of high-grade urothelial carcinoma (HGUC)
61 while assessment of specimens for low-grade urothelial carcinoma (LGUC) is subject to variable
62 cytomorphological features, which limits confidence in the findings. High inter-rater variability
63 and equivocal findings are factors which still adversely impact urine cytology for screening^{6,7}.
64 Surgical biopsy, while often considered the gold standard for bladder cancer is also fraught with
65 sampling issues that may result in a negative biopsy on an unequivocally positive cytology result
66 ⁸.
67

68 Several reporting and classification systems have been developed for assessing Urine Cytology
69 (UC), yet many such systems failed to gain widespread adoption due to inconsistent terminology
70 and grading criteria or overly complex scoring metrics ⁹. As assessment perspectives have shifted
71 toward quantitative endpoints of cytological atypia, The Paris System for Reporting Urinary
72 Cytology (TPS), conceived of in 2013 and officially published in 2016, has emerged as the
73 widely agreed upon classification system for urine cytology bladder cancer screening ².
74 Cytologists using TPS typically assess the individual urothelial cells of a cytology specimen for
75 features of atypia (nuclear hyperchromasia, high nuclear to cytoplasm area ratio (NC ratio),
76 nuclear membrane irregularities, chromatin abnormalities), culminating in the assignment of UC
77 diagnostic categories (four ordered classes: negative, atypical, suspicious, positive). Assignment
78 to diagnostic categories typically leads to the following triaged responses: 1) interval follow-up
79 if negative, 2) follow-up and potentially ancillary tests if atypical, and 3) cystoscopy and biopsy
80 if suspicious or positive. Diagnosis of high-grade urothelial carcinoma (positive) includes
81 following criteria: 1) at least 5 malignant urothelial cells, 2) NC ratio at or above 0.7, 3)
82 markedly irregular nuclear membrane, 4) coarse/clumped chromatin and 5) nuclear
83 hyperchromasia. A suspicious UC finding denotes uncertainty for a positive diagnosis but
84 assignment into either class (suspicious/positive) often leads to the same disease management as
85 aforementioned. At times, not all diagnostic criteria for malignancy may be satisfied due to a
86 variety of confounding reasons (e.g., variable/confluent cellularity, unreliable NC ratio estimates,
87 degenerated cells, specimen preparation), which may preclude ordering of a cystoscopy or
88 biopsy. Hypothetically, the cystoscopy or biopsy could yield positive results if the tests were
89 ordered ¹⁰⁻¹⁵. While high inter-observer variability and lower levels of diagnostic precision are

90 still reported for TPS for some of the diagnostic categories, TPS is still preferred to the prior
91 mode of cytological assessment (assigned acronym: PMC) ¹⁶.

92

93 Though TPS aimed to reduce this tendency, a remaining impediment to reliable urine cytology
94 assessment is the assignment of “atypia”, which serves as a “catch-all” term often overused by
95 pathologists to hedge against a negative, suspicious or positive UC diagnosis ¹⁷⁻¹⁹. The atypical
96 category is especially problematic because many clinicians interpret it as “effectively negative”.

97 TPS introduced diagnostic criteria designed to define “atypical urothelial cells” with emphasis on
98 reporting major criterion of an NC ratio of at least 0.5, while relaxing assessment of other
99 criteria (minor criteria): 1) nuclear hyperchromasia, 2) nuclear membrane irregularity and 3)

100 chromatin quality. The criteria for atypia, in comparison to suspicious or positive, only requires

101 identification of 2 (1 major, 1 minor criteria) of these features. Furthermore, atypia in TPS

102 should exclude morphologic alterations attributable to a lithiasis or infections. Introduction of

103 these diagnostic criteria under TPS serve to standardize assessment by encouraging assignment

104 of a definitive diagnosis (negative, positive) or a slightly hedged diagnosis (suspicious) to avoid

105 conflating loose criteria under PMC. However, reclassification under TPS has been shown to

106 have relatively modest impact on the predictive value for UC despite avoidance of atypia

107 classification, in part due to high interobserver variability and further complicated by the

108 presence of other potential confounding benign cell types (e.g., seminal vesicle cells, glandular

109 cells, inflammatory cells, etc.). In addition, urothelial cell clusters can be challenging to assess

110 due to ambiguous assessment criteria and blended cell borders.

111

112 Previous studies have sought to compare TPS to PMC to determine if the incidence of atypical
113 diagnoses had changed using the standardized criteria^{17,18,20-27}. The prior studies have been
114 cross-sectional in nature and were of limited sample size. Furthermore, few studies have assessed
115 the impact of architectural barriers (e.g., excessive blood, confluent cellularity,) and human
116 barriers (e.g., cytopathologist bias and human factors, i.e., fatigue, training, workload) for
117 diagnosis²⁸⁻³¹.

118
119 In this study, we performed a longitudinal assessment of diagnostic differences between TPS and
120 PMC in a retrospective cohort of nearly 1,300 specimens from 2008-2018, collected prior to the
121 May 2018 implementation of TPS guidelines at an academic medical center, to assess how
122 implementation of the TPS system influenced diagnoses. We hypothesized that, in agreement
123 with previous studies, TPS assessments would yield fewer atypical assignments than PMC, but,
124 uniquely, we were also interested in assessing and accounting for the time frame over which the
125 adoption occurred. Longitudinal changes in PMC assessments may carry far-reaching
126 implications for studies evaluating the implementation of cytology reporting systems, in that
127 early, informal dissemination and discussion of materials might influence adoption. If such
128 factors are identified, they could shed new light on optimal adoption strategies for new reporting
129 criteria and technologies. A longitudinal assessment is important for determining the scope of
130 such investigations.

131

132 **Methods**

133 **Data Collection**

134 After Institutional Review Board approval, we collected 1,277 urine samples across 141 patients
135 (8 (Q1=5; Q3=12; IQR=7) repeat measurements per patient based on longitudinal follow-up)
136 prior to incorporation of the TPS system at our institution, Dartmouth-Hitchcock Medical Center
137 (DHMC). Sample collection did not reflect true incidence of UC classifications in the
138 population, but were chosen to ensure a mix of different classifications for comparison between
139 the two assessment methods. Urine samples, most of them voided, were prepared using
140 ThinPrep® and Pap-stained prior to visual microscopic examination of the cytology slide.
141 Pathologic examination was done using Olympus microscopes of various makes. Primary UC
142 diagnoses were assigned at the time of collection by 10 different cytopathologists under the PMC
143 system. The PMC system consisted of four diagnostic options ("Negative", "Atypical",
144 "Suspicious", and "Malignant") without standardized criteria being applied.

145
146 After collecting PMC assessments from the original cytology report, all specimens were re-
147 assessed by five cytopathologists explicitly considering TPS criteria; cytopathologists were not
148 aware of the original diagnosis. Formal implementation of TPS occurred at DHMC in May 2018
149 and included the creation of specific TPS diagnostic categories within our Cerner Millennium
150 system ("Negative for High Grade Urothelial Carcinoma", "Atypical Urothelial Cells",
151 "Suspicious for High Grade Urothelial Carcinoma", and "Positive for High Grade Urothelial
152 Carcinoma"). Immediately prior to implementation, the cytopathologists met as a group to
153 review TPS criteria, and the TPS criteria were presented at a genitourinary tumor board.

154
155 Four of the five pathologists who reassessed with TPS had also made previous assessments under
156 PMC, but slides were selected to ensure the same pathologist did not assess the same slide twice.

157 We also recorded whether the slide contained any significant cytological artifacts/confounders
158 (e.g., blood, neobladder, confluent cellularity, etc.) that could impact assignment of high/low
159 risk. Nondiagnostic and undetermined samples were removed from the analysis. We additionally
160 recorded demographic information and collection date (**Table 1**). We also collected all PubMed
161 articles directly referencing TPS across the study period by querying PubMed with the search
162 term “The Paris System” and removing articles unrelated to urine cytology. For each article, we
163 recorded the publication year and number of articles citing these publications by the writing of
164 this work as a measure of lasting influence.

165
166 **Table 1:** Demographic characteristics of study participants; specimen source, history and
167 whether an artifact (e.g., blood, confluent cellularity, neobladders and viewing artifacts, anything
168 that could impact assessment quality) was present was registered on the specimen level, while
169 patient demographics (age, sex) were registered on the patient level; for these tables information
170 that could not be determined was removed

Source	N (%)	History	N (%)
Barbotage	6 (0.6)	Bladder Cancer	943 (86.6)
Catheterized	10 (0.9)	Hematuria	110 (10.1)
Cystoscopic	36 (3.3)	Other	36 (3.3)
Ileal	18 (1.7)		
Neobladder	7 (0.6)	Artifact Present N (%)	236 (21.7)
Nephrostomy	2 (0.2)		
Voided	986 (90.5)	Patient Demographics	
Wash	14 (1.3)	Age (mean (SD))	71.87 (10.94)
Other	10 (0.9)	Sex = M (%)	106 (75.7)

171
172 **Diagnostic Differences between PMC and TPS**

173 Assignment of UC categories with PMC and TPS were compared by cross-tabulating diagnoses
174 into a table form to visualize changes between the two systems. Several Bayesian hierarchical
175 logistic regression models were fit to the data in order to ascertain: 1) were cytopathologists
176 operating under TPS less likely to assign atypical class, and 2) whether the PMC assignments
177 began to resemble TPS nearing the end of the collection period (Appendix “Estimating
178 Diagnostic Changes between PMC and TPS across Time”). Separately, overall agreement
179 between the two systems was approximated through calculation of the intra-class correlation

180 coefficient (*ICC*; Appendix “Overall Differences between PMC and TPS”). Pathologists who
 181 assessed fewer than ten slides were removed from the statistical analyses. Collection dates were
 182 mean-centered and units were scaled to years. Odds ratios (OR) were reported to communicate
 183 effect size. Statistical significance was assessed using calculation of the 95% high density
 184 posterior credible interval (CI), analogous to the confidence interval, and the probability of
 185 direction, *pd*, related to the p-value ($p \approx 2 * (1 - pd)$; “p-value”). An OR and corresponding CI
 186 less than one indicates a negative association with the outcome or decreased likelihood, while
 187 greater than one indicates a positive association or increased likelihood given one unit increase in
 188 the predictor. Model fitting was accomplished using the R v3.6 statistical language via the *brms*
 189 package, with post hoc comparisons via the *emmeans* package^{32–36}.

190

191 Results

192 **Table 2:** Cross-tabulation of UC class assignments by cytopathologists operating under PMC
 193 and TPS across the entire study period (2008-2018)

UC Class		TPS				Total
		Negative	Atypical	Suspicious	Positive	
PMC	Negative	601	44	0	0	645
	Atypical	120	183	33	4	340
	Suspicious	1	24	40	18	83
	Positive	0	3	10	21	34
	Total	722	254	83	43	1102

194

195 **Table 3:** Estimated intra-class correlation coefficients for each UC class as a measure of
 196 agreement between PMC/TPS

UC Class	ICC	2.5% CI	97.5% CI
Negative	0.908	0.876	0.937
Atypical	0.744	0.671	0.804
Suspicious	0.775	0.689	0.852
Positive	0.903	0.834	0.957
Overall	0.907	0.883	0.929

197

198 **Table 4:** Reported effect estimates and “p-values” which compare odds of obtaining UC class
 199 given TPS versus odds of obtaining UC class given PMC; odds ratio greater than one indicates
 200 that assignment of UC was associated with TPS over PMC, whereas odds ratio less than one
 201 indicates UC assignment was associated with PMC over TPS; odds ratios were calculated: a)
 202 across all timepoints (*Overall*), and separately using the conditional effects from the interaction
 203 term– b) first date of collection in 2008 ($T=2008$) and c) last date of collection in 2018 prior to
 204 implementation of TPS ($T=2018$)

UC Class	Versus	Date	Odds Ratio	2.5% CI	97.5% CI	p
Negative	Other	Overall	3.56	2.18	6.02	<0.001
		T=2008	9.05	3.47	23.94	<0.001
		T=2018	1.22	0.45	3.18	0.69
Atypical	Other	Overall	0.37	0.25	0.54	<0.001
		T=2008	0.18	0.09	0.36	<0.001
		T=2018	0.83	0.36	1.73	0.66
Suspicious	Other	Overall	1.03	0.65	1.67	0.89
		T=2008	1.11	0.42	3.07	0.82
		T=2018	0.94	0.30	3.22	0.91
Positive	Other	Overall	1.56	0.59	4.30	0.38
		T=2008	1.19	0.14	9.57	0.86
		T=2018	2.20	0.18	26.99	0.51
Atypical	Negative	Overall	0.25	0.15	0.42	<0.001
		T=2008	0.09	0.04	0.25	<0.001
		T=2018	0.79	0.29	2.15	0.65
Atypical	Suspicious	Overall	0.66	0.40	1.04	0.07
		T=2008	0.47	0.17	1.24	0.13
		T=2018	0.96	0.33	3.15	0.95
Atypical	Positive	Overall	0.44	0.18	1.12	0.08
		T=2008	0.38	0.05	2.84	0.35
		T=2018	0.52	0.05	5.46	0.59

205

206 Comparing UC Assignments between TPS and PMC

207 Diagnoses (645 negative, 340 atypical, 83 suspicious, and 34 positive) were assigned by 10

208 different pathologists under the PMC system. Five pathologists provided UC assignments (722

209 negative, 254 atypical, 83 suspicious, and 43 positive) on the same set of cases from 2008-2018

210 using the TPS system (**Table 2**).

211

212 The cross-tabulated cytological findings in Table 2 suggest moderately high overall agreement

213 between PMC and TPS ($ICC=0.907$, 95% CI: 0.883-0.929). Pathologists operating under PMC

214 and TPS tended to agree on negative ($ICC=0.908$) and positive assignments ($ICC=0.903$), while

215 generally in lesser agreement over assignments of atypical ($ICC=0.744$) and suspicious

216 ($ICC=0.775$) UC classes (**Table 3**).

217

218 Notably, there were a few important findings that drove disagreement between the two systems.

219 There were significantly fewer atypical assignments in TPS (n=254) as compared to PMC

220 (n=340) (OR=0.37, 95% CI: [0.25-0.54], $p<0.001$; **Table 4**). A majority of the atypical cases that

221 were reassigned to other UC classes under TPS system were downgraded to negative (n=120),
 222 compared to PMC (n=44) (OR=0.25, 95% CI: [0.15-0.42], p<0.001; **Table 4**), and as compared
 223 to other UC classes (suspicious, positive), which were reassigned to less frequently. A few
 224 atypical findings were reassigned to suspicious (n=33), though reassignments were marginally
 225 insignificant (OR=0.66, 95% CI: [0.40-1.04], p=0.07; **Table 4**).

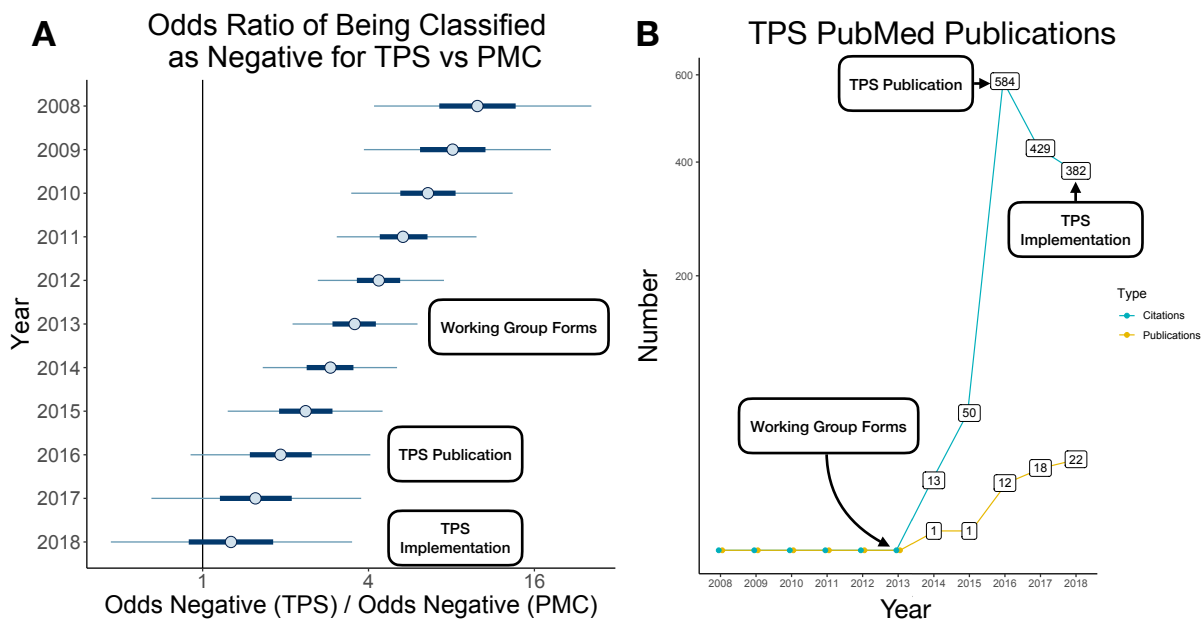
226

227 Interestingly, 71 specimens were reassigned from other UC classes under PMC to atypical under
 228 TPS. There were more negative and positive findings with TPS (n=765=722+43) as compared to
 229 PMC (n=679=645+34), while the number of suspicious findings did not change although 43
 230 specimens deemed suspicious under PMC were reassigned under TPS (**Table 2**).

231

232 There appeared to be no impact of specimen artifacts on grading (Supplementary Table 1).

233



234 **Figure 1: Longitudinal differences in negative UC class assignment under each system: A)**
 235 Odds ratios (OR) and 95% CI (thin line) for assignment of negative UC class between TPS and
 236 PMC systems reported yearly in the month of May using the multinomial statistical model; OR
 237 of 1 (vertical line) indicates no difference in assignment of negatives between TPS and PMC;
 238

239 OR greater than one indicates that, for the same cases, more negative diagnoses were assigned
240 using TPS as compared to PMC, with number of atypical UC diagnoses assigned as a reference
241 (proxy for how PMC atypical diagnoses were reassigned as negative under TPS); results indicate
242 that differences between the two systems disappeared by 2016, where the credible interval
243 intersects OR=1; **B**) Number of yearly TPS-related publications across time and number of
244 articles citing these publications by the writing of this article in 2022 as a proxy for article
245 influence

246

247 **Longitudinal Changes in UC Diagnostic Findings under PMC**

248 Interestingly, assignment patterns of UC atypical and negative categories under PMC began to
249 resemble TPS by the end of the collection period and the formal implementation of TPS. At the
250 beginning of the study period, 2008, pathologists operating under TPS were far less likely to
251 assign an atypical category as compared to PMC (OR=0.18, 95% CI: [0.09-0.36], $p<0.001$;
252 **Table 4**), even more so than reported overall across the collection period (OR=0.37, 95% CI:
253 [0.25-0.54], $p<0.001$; **Table 4**). However, at the end of the collection period, 2018, the
254 differences between PMC and TPS for atypical assignments had entirely disappeared (OR=0.83,
255 95% CI: [0.36-1.73], $p=0.66$; **Table 4**). Conversely, the opposite effect can be seen for the
256 negative UC class– larger differences between PMC and TPS at the start of the collection period
257 (OR=9.05, 95% CI: [3.47-23.94], $p<0.001$; **Table 4**), even more so than reported across the
258 collection period (OR=3.56, 95% CI: [2.18-6.02], $p<0.001$; **Table 4**), and then disappearance of
259 the association near the formal implementation of TPS (OR=1.22, 95% CI: [0.45-3.18], $p=0.69$;
260 **Table 4**). It should be noted that the time of assessment via PMC and re-assessment via TPS for
261 2008 cases is separated by more than a decade (2008 versus 2019), while the assessment time
262 difference has diminished by 2018, hence suggesting potential similarities in reporting of
263 negative cases towards the conclusion of the collection period. These associations were driven by
264 reductions in atypical and increases in negative assignments for PMC, relative to TPS, such that
265 assignment patterns of atypical and negative categories were similar between the two systems by

266 the end of the collection period (2008 OR=0.09, 95% CI: [0.04-0.25], $p<0.001$; 2018 OR=0.79,
267 95% CI: [0.29-2.15], $p=0.65$). Through longitudinal modeling, the differences between these two
268 systems may have disappeared as soon as 2016, when TPS was first officially published (**Figure**
269 **1A**). These findings appear to coincide with the rise in influential TPS-related publications
270 (**Figure 1B**).

271

272 **Discussion**

273 Assessment of urine cytology is currently used for prospective screening and longitudinal
274 monitoring, while cystoscopy and histological examination of a tissue biopsy provide the
275 definitive diagnoses for urothelial carcinoma and dictate oncology care³⁷⁻³⁹. Thus, a hybrid
276 system has evolved in which cytology serves to triage patients to more or less invasive follow up
277 procedures in a paradigm similar to cervical cancer screening. Reporting standardization enabled
278 through adoption of well-defined quantitative and qualitative criteria (e.g., TPS) may soon make
279 cytological analysis a viable alternative to a significant subset of cystoscopies and biopsies by
280 reducing equivocal diagnoses or reports of “atypia”⁴². This study sought to compare urine
281 cytology assessment findings between the prior mode of cytologic assessment (PMC) and the
282 Paris system (TPS) as well as to understand how urine screening assessment changed over the
283 prior decade given informal adoption of standard TPS criteria.

284

285 Findings from this study corroborate previously published evidence that diagnoses under TPS
286 discourage “atypia” classifications. In our study, we found that a vast majority of atypical cases
287 under PMC were instead reclassified as negative under TPS. Many PMC atypical specimens
288 were also relabeled as suspicious and positive, which suggests that the more definitive TPS

289 criteria may have led to fewer equivocal diagnoses. However, these differences were not uniform
290 over time, as differences between the PMC and TPS systems were greatest at the beginning of
291 the study period but negligible immediately before formal implementation of TPS. This suggests
292 several conclusions: 1) given that TPS was first discussed in 2013 and officially published in
293 2016, it is possible that cytopathologists may have been influenced by the discussion and
294 publication of the guidelines, even though official adoption of the guidelines at our institution
295 did not occur until mid 2018 (e.g., **Figure 1B** uses yearly publications and downstream citations
296 to discuss the burgeoning influence of the guideline’s discussions and publication), and 2)
297 previous studies failed to include longitudinal findings for the PMC/TPS comparison. By failing
298 to stratify or adjust for modification across time, prior studies may have underestimated the
299 degree to which these two systems initially diverged, as the atypical category was assigned in far
300 greater numbers by PMC as compared to reassessment by TPS for 2008. The gap between PMC
301 and TPS had disappeared before 2018, just prior to the implementation of TPS. These findings
302 have potential implications across other pathology subdisciplines, where timely adoption of new
303 criteria is of interest as evolving staging/grading guidelines feature more quantitative endpoints.
304 Efforts to establish rigorous quantitative endpoints for disease pathology may be further
305 accelerated using computer-assisted diagnostic decision aids (e.g., machine learning)^{1,43,44}.
306
307 Because several large studies have already examined the differences between TPS and PMC in
308 their ability to predict definitive diagnoses from cystoscopy and/or biopsy, we opted to not assess
309 concordance of TPS/PMC with these independent measures. Neither did we seek to compare
310 differences in assignment patterns between and “within” pathologists via estimates of inter- and
311 intra-observer reliability, which has been well characterized in the literature and could have

312 potentially impacted final effect estimates (e.g., changes in clinical staff). We did not assess
313 changes in diagnostic assignment per pathologists across time (i.e., years of experience,
314 training), which presents an interesting follow-up direction. Other limitations include differences
315 between real clinical sign-out (as was seen with PMC) and academic/research sign out (seen with
316 TPS) (e.g., TPS raters did not have access to clinical information or the cytotechnologist's
317 preliminary diagnoses, both of which can influence final diagnosis in practice). The lack of
318 clinical information may hamper the ability to apply TPS criteria for the final UC assessment, as
319 cytologists regularly look for and consider factors such as stones that can cause reactive atypia.
320 While associations were found that were longitudinal in nature, we are unable to definitely
321 conclude that publication of TPS influenced assessment prior to adoption of the system in our
322 institution due to the potential presence of temporal unmeasured confounders. Some potential
323 unmeasured confounders are known (e.g., attendance at conferences, interactions with other
324 practicing cytologists, professional social media exchanges as means to get information—e.g.,
325 Twitter, #CytoChat) while others may be unknown and unanticipated. Nonetheless, this study
326 evaluated agreement amongst two cytology assessment systems longitudinally over a range not
327 previously covered by any similar study.

328
329 The study findings point to the potential role and impact of widespread discussions and
330 communication on early adoption, dissemination, and implementation of evolving clinical
331 guidelines. While TPS is now commonplace, reasons for longitudinal differences in PMC prior
332 to TPS include early publications, presentation of TPS criteria at conferences, interactions and
333 communication among pathologists across various clinical and virtual settings, and differences in
334 training/experience. However, reasons for change are nearly impossible to evaluate as such data

335 is not routinely collected, which precludes the application of statistical methodologies (e.g.,
336 social network analysis, difference-in-differences design, causal inference) that could elucidate
337 such findings. As similar cytology reporting systems are rolled out or refined (e.g., The Bethesda
338 systems for thyroid and cervical cytopathology, Milan system for salivary gland cytopathology,
339 and other standardized reporting systems, such as fluid cytology, breast cytology, etc.), there
340 exist opportunities to develop and apply innovative data collection and analytical methods to
341 build a greater understanding of successful guideline adoption strategies. At a minimum,
342 retrospective comparisons to previous assessment methods prior to official adoption of new
343 criteria should be performed over a broad time horizon (e.g., 4-5 years as suggested by the
344 significant study findings) sufficient to characterize changing practices.

345

346 **Conclusion**

347 TPS purports to improve reporting of UC by defining clear criteria for diagnostic categories. In
348 our retrospective study, we provide confirming evidence of the achievement of intended changes
349 in diagnostic assignment. A unique longitudinal assessment suggested that early adoption of
350 standard reporting criteria may have occurred prior to the implementation of TPS in our
351 institution. This will hardly be the last change in reporting practices as increasingly quantitative
352 and reproducible reporting criteria are adopted. As these practices shift towards quantitative
353 endpoints, expert time and specialization required for nuanced decision making may increase,
354 though partial automation of expert decision making may ameliorate its most challenging
355 aspects. In the future, development and adoption of semi-autonomous classification systems may
356 present the next step forward in the development of standard reporting criterion, as these
357 computerized systems can augment the laborious process of cytological inspection and

358 potentially pave the way for adoption of urine cytology as an alternative for cystoscopy for high-
359 grade urothelial carcinoma.

360

361

362 **References**

363

- 364 1. Vaickus, L. J., Suriawinata, A. A., Wei, J. W. & Liu, X. Automating the Paris System for
365 urine cytopathology—A hybrid deep-learning and morphometric approach. *Cancer*
366 *Cytopathology* **127**, 98–115 (2019).
- 367 2. Barkan, G. A. *et al.* The Paris System for Reporting Urinary Cytology: The Quest to Develop
368 a Standardized Terminology. *ACY* **60**, 185–197 (2016).
- 369 3. Bostwick, D. G. 7 - Urine Cytology. in *Urologic Surgical Pathology (Fourth Edition)* (eds.
370 Cheng, L., MacLennan, G. T. & Bostwick, D. G.) 322-357.e7 (Elsevier, 2020).
371 doi:10.1016/B978-0-323-54941-7.00007-4.
- 372 4. Rosolen, D. C. B., Faria, D. K., Faria, C. S. & Antonangelo, L. Performance of the
373 UroVysion® FISH assay for the diagnosis of malignant effusions using two cutoff
374 strategies. *Cancer Med* **7**, 1967–1977 (2018).
- 375 5. Sciarra, A. *et al.* Comparison of the clinical usefulness of different urinary tests for the initial
376 detection of bladder cancer: a systematic review. *Curr Urol* **15**, 22–32 (2021).
- 377 6. Koss, L. G. Errors and pitfalls in cytology of the lower urinary tract. *Monogr Pathol* 60–74
378 (1997).
- 379 7. Ooms, E. C. & Veldhuizen, R. W. Cytological criteria and diagnostic terminology in urinary
380 cytology. *Cytopathology* **4**, 51–54 (1993).
- 381 8. Schwalb, D. M., Herr, H. W. & Fair, W. R. The management of clinically unconfirmed
382 positive urinary cytology. *J Urol* **150**, 1751–1756 (1993).
- 383 9. Owens, C. L., Vandenbussche, C. J., Burroughs, F. H. & Rosenthal, D. L. A review of
384 reporting systems and terminology for urine cytology. *Cancer Cytopathol* **121**, 9–14 (2013).
- 385 10. Layfield, L. J. *et al.* Accuracy and Reproducibility of Nuclear/Cytoplasmic Ratio
386 Assessments in Urinary Cytology Specimens. *Diagnostic Cytopathology* **45**, 107–112
387 (2017).
- 388 11. Vaickus, L. J. & Tambouret, R. H. Young investigator challenge: The accuracy of the
389 nuclear-to-cytoplasmic ratio estimation among trained morphologists. *Cancer Cytopathol*
390 **123**, 524–530 (2015).
- 391 12. Voss, J. S. *et al.* Changes in specimen preparation method may impact urine cytologic
392 evaluation. *Am J Clin Pathol* **130**, 428–433 (2008).
- 393 13. Barkan, G. A. Enough is enough: adequacy of voided urine cytology. *Cancer Cytopathology*
394 vol. 124 163–166 (2016).
- 395 14. Koss, L. G., Deitch, D., Ramanathan, R. & Sherman, A. B. Diagnostic value of cytology of
396 voided urine. *Acta cytologica* **29**, 810–816 (1985).
- 397 15. Leyh, H. *et al.* Comparison of the BTA statTM test with voided urine cytology and bladder
398 wash cytology in the diagnosis and monitoring of bladder cancer. *European urology* **35**, 52–
399 56 (1999).
- 400 16. Cowan, M. L., Rosenthal, D. L. & VandenBussche, C. J. Improved risk stratification for
401 patients with high-grade urothelial carcinoma following application of the Paris System for
402 Reporting Urinary Cytology. *Cancer Cytopathol* **125**, 427–434 (2017).

- 403 17. Deshpande, V. & McKee, G. T. Analysis of atypical urine cytology in a tertiary care center.
404 *Cancer* **105**, 468–475 (2005).
- 405 18. Piaton, E. *et al.* Diagnostic terminology for urinary cytology reports including the new
406 subcategories ‘atypical urothelial cells of undetermined significance’ (AUC-US) and ‘cannot
407 exclude high grade’ (AUC-H). *Cytopathology* **25**, 27–38 (2014).
- 408 19. Mokhtar, G. A., Al-Dousari, M. & Al-Ghamedi, D. Diagnostic significance of atypical
409 category in the voided urine samples: A retrospective study in a tertiary care center. *Urol*
410 *Ann* **2**, 100–106 (2010).
- 411 20. Raab, S. S., Grzybicki, D. M., Vrbin, C. M. & Geisinger, K. R. Urine cytology discrepancies:
412 frequency, causes, and outcomes. *Am J Clin Pathol* **127**, 946–953 (2007).
- 413 21. Granados, R., Duarte, J. A., Corrales, T., Camarmo, E. & Bajo, P. Applying the Paris System
414 for Reporting Urine Cytology Increases the Rate of Atypical Urothelial Cells in Benign
415 Cases: A Need for Patient Management Recommendations. *Acta Cytol* **61**, 71–76 (2017).
- 416 22. Chau, K. *et al.* Accuracy and risk of malignancy for diagnostic categories in urine cytology
417 at a large tertiary institution. *Cancer Cytopathol* **123**, 10–18 (2015).
- 418 23. Brimo, F. *et al.* Accuracy of urine cytology and the significance of an atypical category. *Am*
419 *J Clin Pathol* **132**, 785–793 (2009).
- 420 24. Simon, C. T. *et al.* The utility of upper urinary tract urine cytology before and after
421 application of the Paris system. *Diagnostic Cytopathology* **47**, 421–427 (2019).
- 422 25. Roy, M. *et al.* An institutional experience with The Paris System: A paradigm shift from
423 ambiguous terminology to more objective criteria for reporting urine cytology.
424 *Cytopathology* **28**, 509–515 (2017).
- 425 26. Mikou, P. *et al.* Evaluation of the Paris System in atypical urinary cytology. *Cytopathology*
426 **29**, 545–549 (2018).
- 427 27. Glass, R. *et al.* Analysis of the cytomorphological features in atypical urine specimens
428 following application of the Paris system for reporting urinary cytology. *Acta Cytologica* **62**,
429 54–61 (2018).
- 430 28. Cimino-Mathews, A. & Ali, S. Z. The clinicopathologic correlates of cellular atypia in
431 urinary cytology of ileal neobladders. *Acta cytologica* **55**, 449–454 (2011).
- 432 29. Ide, H., Kikuchi, E., Shinoda, K., Mukai, M. & Murai, M. Carcinoma in situ developing in
433 an ileal neobladder. *Urology* **69**, 576-e9 (2007).
- 434 30. Yamashita, S. *et al.* Urethral recurrence following neobladder in bladder cancer patients. *The*
435 *Tohoku Journal of Experimental Medicine* **199**, 197–203 (2003).
- 436 31. Hofland, C. A. & Mariani, A. J. Is cytology required for a hematuria evaluation? *The Journal*
437 *of urology* **171**, 324–326 (2004).
- 438 32. Bürkner, P.-C. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R*
439 *Journal* **10**, 395–411 (2018).
- 440 33. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of*
441 *Statistical Software* **80**, 1–28 (2017).

- 442 34. Makowski, D., Ben-Shachar, M. S., Chen, S. H. A. & Lüdecke, D. Indices of Effect
443 Existence and Significance in the Bayesian Framework. *Front. Psychol.* **10**, 2767 (2019).
- 444 35. Carpenter, B. *et al.* *Stan: A Probabilistic Programming Language. Grantee Submission* vol.
445 76 1–32 (2017).
- 446 36. Wickham, H. & Grolemund, G. *R for Data Science: Import, Tidy, Transform, Visualize, and*
447 *Model Data.* (O'Reilly Media, 2017).
- 448 37. Hassan, M. *et al.* Impact of Implementing the Paris System for Reporting Urine Cytology in
449 the Performance of Urine Cytology: A Correlative Study of 124 Cases. *American Journal of*
450 *Clinical Pathology* **146**, 384–390 (2016).
- 451 38. Rohilla, M. *et al.* Cytohistological correlation of urine cytology in a tertiary centre with
452 application of the Paris system. *Cytopathology* **29**, 436–443 (2018).
- 453 39. Stanzione, N. *et al.* The continual impact of the Paris System on urine cytology, a 3-year
454 experience. *Cytopathology* **31**, 35–40 (2020).
- 455 40. Kurtycz, D. F. I., Sundling, K. E. & Barkan, G. A. The Paris system of Reporting Urinary
456 Cytology: Strengths and opportunities. *Diagnostic Cytopathology* **48**, 890–895 (2020).
- 457 41. Rohra, P. *et al.* Effect of the Paris system for reporting urinary cytology with histologic
458 follow-up. *Diagnostic Cytopathology* **49**, 691–699 (2021).
- 459 42. Torous, V. F., Brancely, D. & VanderLaan, P. A. Implementation of the Paris System for
460 Reporting Urinary Cytology results in lower atypical diagnostic rates. *Journal of the*
461 *American Society of Cytopathology* **6**, 205–210 (2017).
- 462 43. Vaickus, L. J. & Tambouret, R. A simplified method for assessing cytotechnologist
463 workload. *Cancer Cytopathol* **122**, 15–22 (2014).
- 464 44. Sanghvi, A. B., Allen, E. Z., Callenberg, K. M. & Pantanowitz, L. Performance of an
465 artificial intelligence algorithm for reporting urine cytopathology. *Cancer Cytopathology*
466 **127**, 658–666 (2019).
- 467
468
469