

1 **TITLE**

2 Reassessing the management of uncomplicated urinary tract infection: A retrospective analysis
3 using machine learning causal inference

4
5 **AUTHORS**

6 Noah C Jones¹, Ming-Chieh Shih², Elizabeth Healey³, Chen Wen Zhai⁴, Sonali D Advani⁵, Aaron
7 Smith-McLallen⁶, David Sontag^{*7}, Sanjat Kanjilal^{*8}

8 * Co-last author

9
10 **AFFILIATIONS**

11 ¹ MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA

12 ² College of Life Sciences and Medicine, National Tsing Hua University, Hsinchu, Taiwan

13 ³ Program in Health Sciences and Technology, Massachusetts Institute of Technology,
14 Cambridge, MA

15 ⁴ Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA

16 ⁵ Division of Infectious Diseases, Department of Medicine, Duke University School of Medicine,
17 Durham, NC

18 ⁶ Independence Blue Cross, Philadelphia, PA

19 ⁷ Institute for Medical Engineering and Sciences, Massachusetts Institute of Technology,
20 Cambridge, MA

21 ⁸ Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare
22 Institute, Boston, MA

23
24 **CORRESPONDING AUTHOR**

25 Sanjat Kanjilal MD, MPH

26 Department of Population Medicine

27 Harvard Medical School & Harvard Pilgrim Healthcare Institute

28 401 Park Drive, Suite 401 East

29 Boston, MA 02215

30 skanjilal@bwh.harvard.edu

31 (617) 867-4211

32
33 **WORD COUNT**

34 3103

35 **ABSTRACT**

36 Background

37 Uncomplicated urinary tract infection (UTI) is a common indication for outpatient antimicrobial
38 therapy. National guidelines for the management of uncomplicated UTI were published by the
39 Infectious Diseases Society of America in 2011, however it is not fully known the extent to which
40 they align with current practices, patient diversity, and pathogen biology, all of which have evolved
41 significantly in the time since their publication.

42 Objective

43 We aimed to re-evaluate efficacy and adverse events for first-line antibiotics (nitrofurantoin, and
44 trimethoprim-sulfamethoxazole), versus second-line antibiotics (fluoroquinolones) and versus
45 alternative agents (oral β -lactams) for uncomplicated UTI in contemporary clinical practice by
46 applying machine learning algorithms to a large claims database formatted into the Observational
47 Medical Outcomes Partnership (OMOP) common data model.

48 Outcomes

49 Our primary outcome was a composite endpoint for treatment failure, defined as outpatient or
50 inpatient re-visit within 30 days for UTI, pyelonephritis or sepsis. Secondary outcomes were the
51 risk of 4 common antibiotic-associated adverse events: gastrointestinal symptoms, rash, kidney
52 injury and *C. difficile* infection.

53 Statistical methods

54 We adjusted for covariate-dependent censoring and treatment indication using a broad set of
55 domain-expert derived features. Sensitivity analyses were conducted using [OMOP-learn](#), an
56 automated feature engineering package for OMOP datasets.

57 Results

58 Our study included 57,585 episodes of UTI from 49,037 patients. First-line antibiotics were
59 prescribed in 35,018 (61%) episodes, second-line antibiotics were prescribed in 21,140 (37%)
60 episodes and alternative antibiotics were prescribed in 1,427 (2%) episodes. After adjustment,
61 patients receiving first-line therapies had an absolute risk difference of -2.1% [95% CI -2.9% to
62 -1.6%] for having a revisit for UTI within 30 days of diagnosis relative to second-line antibiotics.
63 First-line therapies had an absolute risk difference of -6.6% [95% CI -9.4% to -3.8%] for 30-day
64 revisit compared to alternative β -lactam antibiotics. Differences in adverse events were clinically
65 similar between first and second line agents, but lower for first-line agents relative to alternative
66 antibiotics (-3.5% [95% CI -5.9% to -1.2%]). Results were similar for models built with [OMOP-](#)
67 [learn](#).

68 Conclusion

69 Our study provides support for the continued use of first-line antibiotics for the management of
70 uncomplicated UTI. Our results also provide proof-of-principle that automated feature extraction
71 methods for OMOP formatted data can emulate manually curated models, thereby promoting
72 reproducibility and generalizability.

73 INTRODUCTION

74 Up to 50% of women will experience a urinary tract infection (UTI) in their lifetime¹, making it the
75 third most common indication for antibiotic treatment in the United States after respiratory tract
76 infection and skin and soft tissue infections². Treatment guidelines published by the Infectious
77 Diseases Society of America (IDSA) encourage the use of nitrofurantoin, trimethoprim-
78 sulfamethoxazole and fosfomycin as first-line treatments for uncomplicated UTI based on their
79 efficacy and relatively limited side effect profile^{3,4}. Fluoroquinolones are listed as a second-line
80 option due to their predilection for selecting for multidrug resistant organisms⁵ and their
81 association with serious adverse events including *C. difficile* colitis⁶. Despite this, ciprofloxacin
82 and levofloxacin are still the most commonly used antibiotics in the treatment of UTI, which may
83 reflect the real or perceived threat of antibiotic resistance to the first line agents². The guidelines
84 list beta-lactams as alternative treatments as they are associated with reduced treatment efficacy⁷.

85 The evidence base supporting the IDSA treatment guidelines are based on a small number of
86 randomized controlled trials and observational studies⁷⁻¹¹, many of which were completed several
87 decades ago. While these provide important information for policy making, they were limited in
88 the diversity of patients they recruited and were performed at a time when standards of care, and
89 health-seeking behavior differed significantly from current practice. Furthermore, the pathogen
90 strains in circulation at the time of these studies have likely been replaced by new strains that
91 may have a differential response to drug therapies regardless of their susceptibility phenotype.
92 Therefore, a re-evaluation of management strategies for uncomplicated UTI could provide useful
93 information for treating clinicians and policy makers. In this study, we sought to estimate treatment
94 efficacy and adverse events for guideline-concordant and discordant treatments for UTI using
95 causal inference supported by machine learning applied to a large contemporary claims dataset.

96 METHODS

97 Study design and data

98 We conducted a retrospective cohort study using the claims database from Independence Blue
99 Cross, which contains health-related information for 3 million people living primarily in a 5 county
100 area surrounding Philadelphia, PA. The dataset contains inpatient, outpatient, laboratory and
101 pharmacy claims made between 2012 and 2021. The database is formatted in the Observational
102 Medical Outcomes Partnership (OMOP) common data model (version 5), developed by the
103 Observational Health Data Sciences and Informatics (OHDSI) initiative¹². Reporting of this study
104 follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)
105 statement¹³. This study was deemed exempt by the Institutional Review Board of the
106 Massachusetts Institute of Technology (protocol E-3970).

107 Study population

108 The analysis cohort consisted of non-pregnant females aged 18 and older with a diagnosis of
109 uncomplicated, non-recurrent UTI at an outpatient setting. The list of diagnosis codes associated
110 with a diagnosis of UTI is provided in Supplementary Table 1. Patient included in the analysis
111 must also have been treated with one of the following three classes antimicrobials within a 7-day
112 period after the diagnosis: a) nitrofurantoin, and trimethoprim-sulfamethoxazole (first line
113 treatments), b) ofloxacin, ciprofloxacin and levofloxacin (second-line treatments), or c) specific

114 oral β -lactam drugs used for UTI such as amoxicillin-clavulanate, cefadroxil, and cefpodoxime
115 (alternative treatments). Fosfomycin was excluded due to the low number of treatment events.

116 We excluded individuals with UTI who received treatment outside of the three classes above, e.g.
117 fluconazole, and individuals treated with more than one antibiotic within a 7-day period. In
118 addition, to avoid contamination of previous antibiotic exposures, we excluded patients that had
119 antibiotic exposure within 7 days before the date of UTI diagnosis. We also excluded those with
120 recurrent UTI, defined as ≥ 2 episodes in a 180 day period and ≥ 3 episodes in a 365 day period,
121 and people with complicated UTI, defined as any males with a UTI diagnosis or females with a
122 predefined list of procedures and diagnoses associated with complicated UTI within 180 days of
123 the diagnosis, or any histories of complicating long-term comorbidities such as neurogenic
124 bladder, spina bifida, or cancers of the genitourinary tract prior to the UTI diagnosis. A full list of
125 comorbidities flagged for exclusion can be found in Supplementary Table 2.

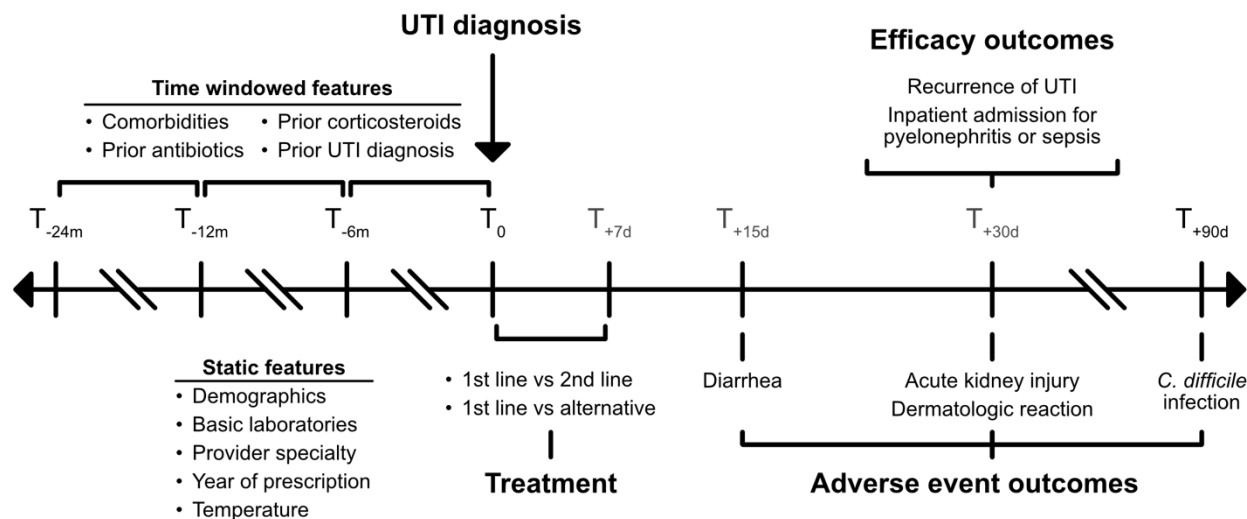
126 Outcomes and Censoring

127 We defined two primary endpoints for the analysis. The first was a composite endpoint for
128 treatment failure, defined as outpatient or inpatient re-visit within 30 days for UTI, pyelonephritis
129 or sepsis. The second set of endpoints involved adverse events, defined as the presence of
130 diarrhea within 15 days of the UTI event, acute kidney injury (AKI) or a dermatologic adverse
131 event within 30 days of a UTI event or a diagnosis of *C. difficile* infection within 90 days of the UTI
132 event. The conditions and the corresponding codes included in each adverse event category are
133 listed in **Error! Reference source not found.** Individuals were right-censored from the analysis
134 if they left the health plan before the observational period of the outcome of interest.

135 Confounder generation

136 We derived 2 sets of baseline covariates, which served as potential confounders. The first utilized
137 domain expert knowledge from two practicing infectious disease physicians (SA and SK), and the
138 second was derived from the [OMOP-learn](#) coding package. [OMOP-learn](#) is a data-driven feature
139 extractor developed in prior work and is specifically designed for use with claims datasets
140 formatted in the OMOP common data model¹⁴. The package serves to automatically generate
141 time-windowed covariates.

142 Domain expert-derived features were classified into demographics, medical conditions, drug
143 prescriptions, prior UTI history, prior antibiotic exposures, laboratory measurements, and provider
144 specialty and year of prescription to account for secular trends in prescribing behavior. A list of
145 medical conditions and drug prescriptions included in the features is shown in Supplementary
146 Table 4. Domain expert-derived features related to medical histories were binned into non-
147 overlapping time windows of 0 to ≤ 6 months, >6 to ≤ 12 months and >12 to ≤ 24 months relative to
148 the date of UTI diagnosis. Laboratories (urinalysis and blood tests) were counted if they were
149 drawn at the time of the UTI diagnosis. The final expert-derived model consisted of 245 features.
150 [OMOP-learn](#) features were derived from diagnoses, procedures, medications, and provider
151 specialties. The total OMOP-derived model consisted of 143,830 features for the comparison
152 between first and second-line antibiotics, and 131,035 features for the comparison between first-
153 line and alternative antibiotics. Figure 1 depicts the cohort, outcome and confounder definitions.



154

155 **Figure 1: Cohort inclusion criteria and definitions for outcomes and feature.** Static features
 156 were evaluated at T₀. Abbreviations, T, time.

157 Statistical analysis

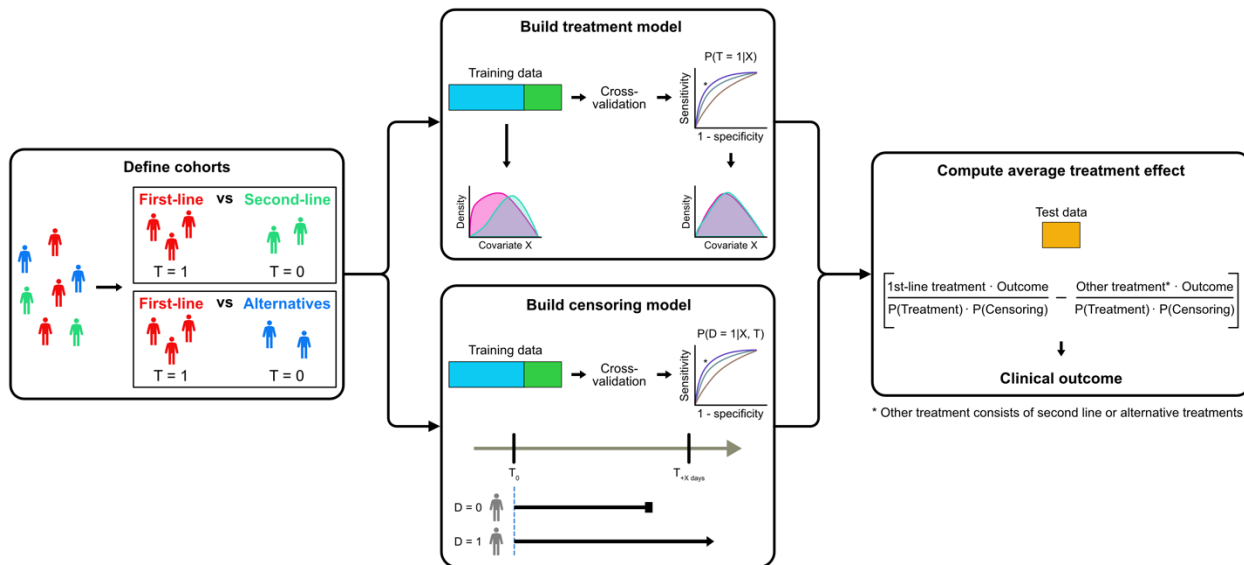
158 We estimated the absolute risk difference of first line therapies versus second line or alternative
 159 therapies for patients with uncomplicated UTI on 30-day recurrence and adverse effects. To
 160 account for possible covariate-dependent censoring, we used inverse probability of censoring
 161 weighting (IPCW) to reweight individuals that were observed or not censored¹⁵. In addition, the
 162 central problem in estimating antibiotic treatment on outcomes is confounding by indication,
 163 therefore we utilized inverse probability weighted propensity scores to adjust for the likelihood of
 164 receipt of each treatment class given an individual's confounders.

165 For both the observation probability model and treatment propensity score model, the dataset
 166 was split 80/20 into a training and test set and the training set was further split 75/25 into a
 167 development and validation set to search optimal hyperparameters. Hyperparameters were
 168 selected using a grid search across three model types, logistic regression, random forests and
 169 light gradient boosted machine models. The model with the highest area-under-the ROC curve
 170 (AUROC) after 3-fold cross-validation was chosen to generate the probabilities of being observed
 171 and propensity scores for the entire dataset. To avoid the undue influence of extreme propensity
 172 scores, we applied symmetric trimming and only included patients with propensity scores between
 173 0.05 and 0.95. We additionally only included patients with follow-up time for the treatment
 174 outcome under consideration. After adjusting for the observation probability and propensity for
 175 treatment, the average treatment effect (ATE) was estimated as follows,

176
$$ATE = E \left[\frac{T_i Y_i}{P(T = 1 | X_i) * P(D_Y = 1 | T = 1, X_i)} - \frac{(1 - T_i) Y_i}{P(T = 0 | X_i) * P(D_Y = 1 | T = 0, X_i)} \right]$$

177 where i is the participant, X are the covariates, $T = 1$ if given a first-line treatment, $D_Y = 1$ if the
 178 patient was followed for at least the outcome variable's follow up period and Y is the outcome,
 179 which is treated as missing when $D_Y = 0$. Confidence intervals for the propensity scores and ATE

180 were generated using bootstrapping¹⁶. Feature importance for both models was determined using
 181 Shapley Additive Explanation values (SHAP values)¹⁷. Figure 2 represents the analytic pipeline.



182
 183 **Figure 2: Analytic pipeline.** We built and separately analyzed cohorts for first-line versus
 184 second-line and first-line versus alternative treatment. Eighty percent of the total data was set
 185 aside for training and this was further split 75/25 into development (blue) and validation (green)
 186 datasets. Two models were then run to estimate the probability of treatment and of being observed
 187 through the outcome period post-diagnosis. We used a 3-fold cross-validation to select the model
 188 with the highest AUROC, indicated by the asterisk. Average treatment effect for a given outcome
 189 was estimated on test data (yellow) by the risk difference between those receiving first-line
 190 treatment or another treatment (second-line or alternative) after normalizing for the probability of
 191 receiving a treatment and of being observed at the end of the outcome's follow-up period (e.g 30
 192 days). Abbreviations, T, treatment, X, covariates, D, observed.

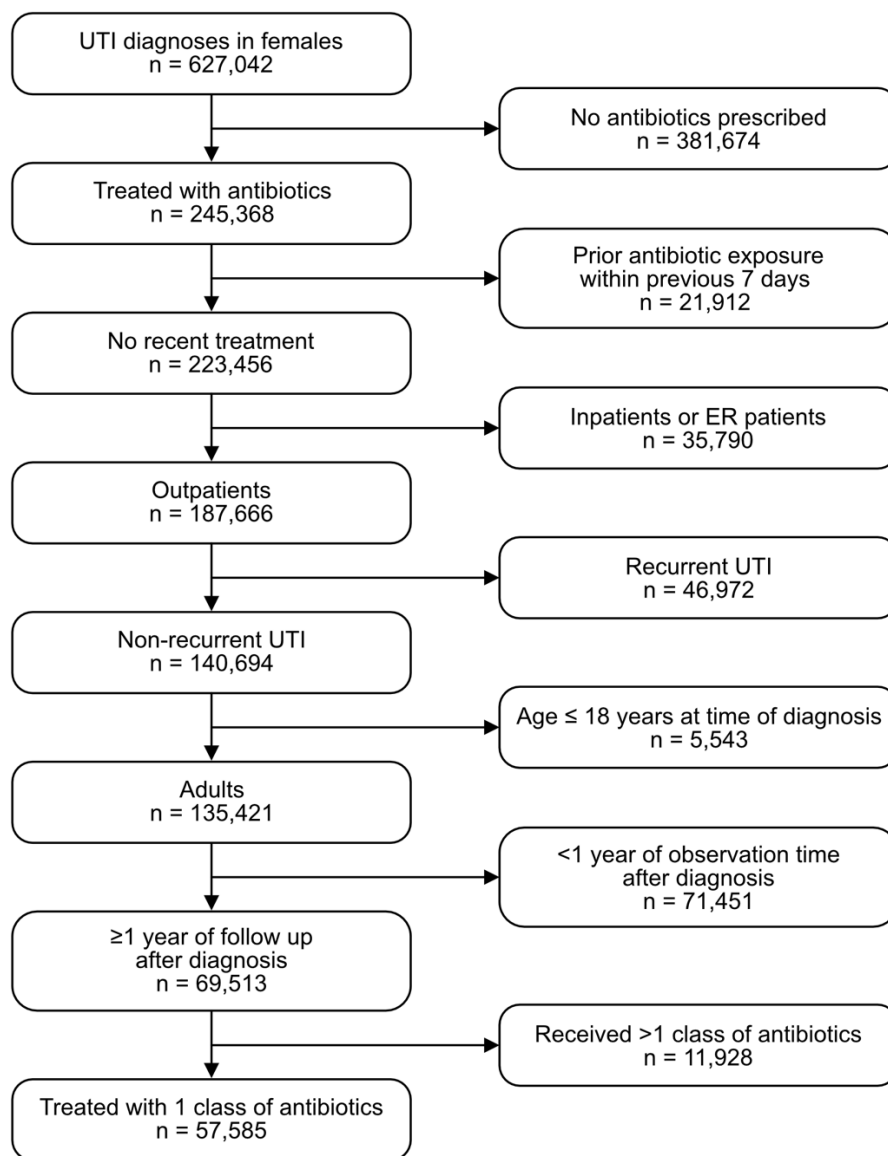
193 We assessed for residual confounding by assessing treatment effect on three negative control
 194 outcomes, fibrocystic disease of the breast, hernia and fracture^{4,18}. These were selected based
 195 on domain knowledge and a comprehensive literature search that found no evidence of a causal
 196 association with exposure to our antibiotics of interest. Treatment effect was calculated by
 197 estimating the prevalence of each negative control outcome at 1 month and 3 months after
 198 exposure.

199 The primary analysis used the model specified by domain expert knowledge. Sensitivity analyses
 200 included subgroup analysis in patients who were admitted within a 30-day period after their initial
 201 diagnosis of UTI and with the model specified by **OMOP-learn**, using the same analysis pipeline.
 202 All analyses were run in Python v3.85 and source code to reproduce all analyses is available at
 203 GitHub (<https://github.com/clinicalml/uti-causal-inference/>).

204 RESULTS

205 Baseline cohort description

206 The study flow diagram is shown in Figure 3 and baseline cohort characteristics are summarized
 207 in Table 1.



208

209 **Figure 3: Study flow diagram.** Sample sizes indicate UTI diagnoses.

Baseline characteristics	Full cohort	First line	Second line	Alternative
Sample size (UTI diagnoses)	57,585	35,018	21,140	1,427
Age (median, IQR)	52 (29)	49 (30)	57 (29)*	59 (35)*
Fever at presentation	649 (1.1)	207 (0.6)	390 (1.8)*	52 (3.6)*
Urinalysis ordered at presentation	13,713 (23.8)	7,365 (21.0)	5,937 (28.1)*	411 (28.8)*
Blood test ordered at presentation	2,248 (3.9)	1,040 (3.0)	1,057 (5.0)*	151 (10.6)*
Menopause	3,852 (6.7)	2,320 (6.6)	1,424 (6.7)	108 (7.6)
UTI in past year	5,863 (10.2)	3,418 (9.8)	2,257 (10.7)*	188 (13.2)*
Underlying conditions				
Hypertension	21,026 (36.5)	10,742 (30.7)	9,525 (45.1)*	759 (53.2)*
Diabetes Mellitus	8,298 (14.4)	4,056 (11.6)	3,910 (18.5)*	332 (23.3)*
Arthritis	11,220 (19.5)	6,122 (17.5)	4,698 (22.2)*	400 (28.0)*
Cancer	5,684 (9.9)	2,847 (8.1)	2,639 (12.5)*	198 (13.9)*
Chronic kidney disease	2,992 (5.2)	1,364 (3.9)	1,458 (6.9)*	170 (11.9)*
Autoimmune	2,956 (5.1)	1,639 (4.7)	1,207 (5.7)*	110 (7.7)*
Thyroid Disorder	292 (0.5)	148 (0.4)	136 (0.6)	8 (0.6)
Year of UTI episode				
2012-2014	7,376 (12.8)	3,435 (9.8)	3,792 (17.9)*	149 (10.4)
2015-2017	26,770 (46.5)	14,757 (42.1)	11,420 (54.0)*	593 (41.6)
2018-2021	23,439 (40.7)	16,826 (48.1)	5,928 (28.0)*	685 (48.0)
Provider specialties				
Family medicine	17,523 (30.4)	9,710 (27.7)	7,426 (35.1)*	387 (27.1)
Internal medicine	8,025 (13.9)	3,858 (11.0)	3,930 (18.6)*	237 (16.6)*
Emergency care	4,757 (8.3)	2,968 (8.5)	1,689 (8.0)*	100 (7.0)
Obstetrics / Gynecologist	2,839 (4.9)	2,124 (6.1)	684 (3.2)*	31 (2.2)*
Other non-urology specialist	6,729 (11.7)	5,119 (14.6)	1,480 (7.0)*	130 (9.1)*
Urology	1,498 (2.6)	907 (2.6)	522 (2.5)	69 (4.8)*
Others	2,385 (4.1)	1,433 (4.1)	818 (3.9)	134 (9.4)*

210 **Table 1. Baseline characteristics of cohort.** Unless otherwise indicated, values represent
 211 sample size and column percentage. * p <0.05 compared to first-line cohort

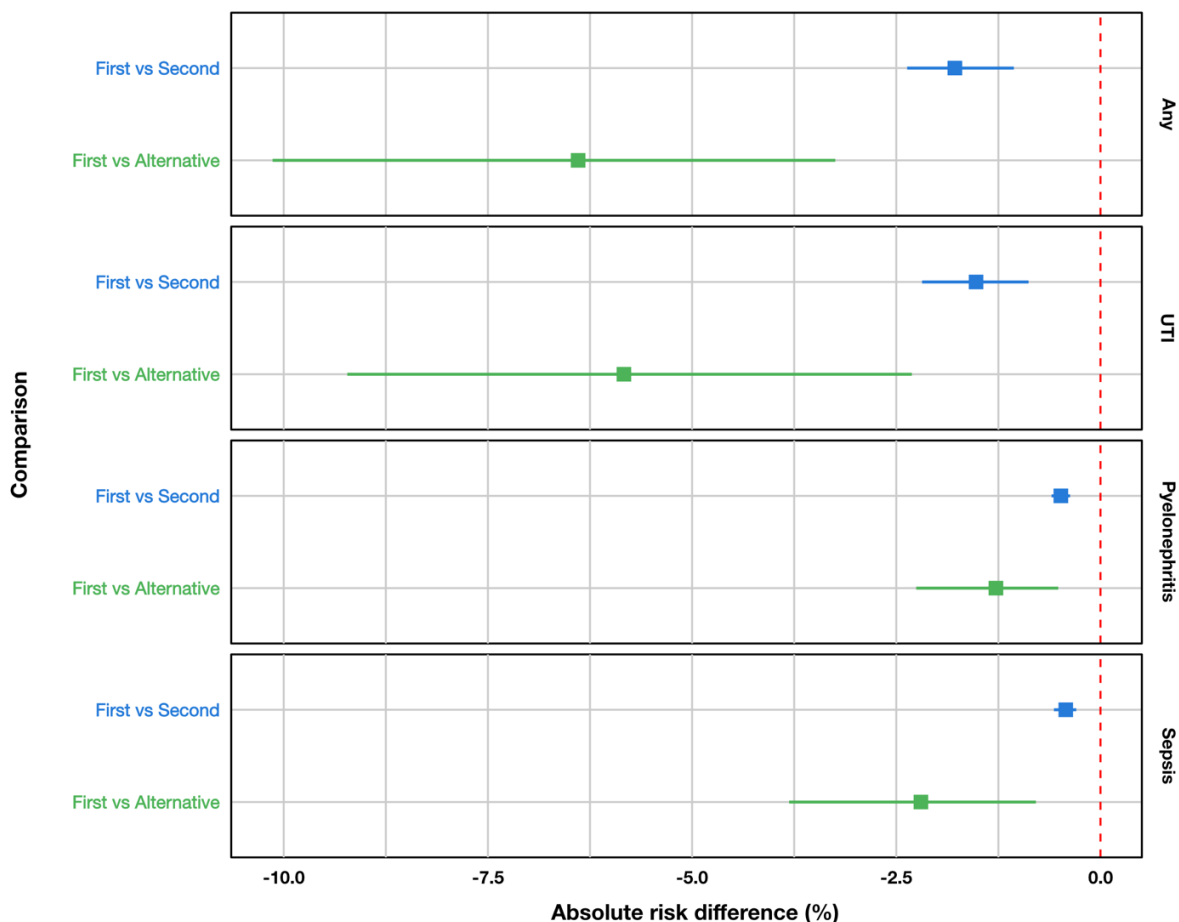
212 The final analysis cohort consisted of 57,585 episodes of UTI occurring in 49,037 patients. Of
 213 these, first-line antibiotics were prescribed in 35,018 (61%) episodes, second-line antibiotics were
 214 prescribed in 21,140 (37%) episodes and alternative antibiotics were prescribed in 1,427 (2%)
 215 episodes. Compared to those prescribed with first-line antibiotics, patients prescribed second-line
 216 antibiotics were older, more likely to present with fever, more likely to have laboratory tests
 217 ordered, and had a higher comorbidity burden. Those prescribed alternative treatments were
 218 similarly older, more likely to be febrile and had a higher comorbidity burden. They were also more
 219 likely to be seen in the emergency room.

220

221 Primary outcomes

222 For the domain expert-derived features, a light gradient boosting machine was the best model for
223 predicting censorship as well as the likelihood of treatment (details of performance in
224 Supplementary Results and Supplementary Figures). The top 5 covariates predicting first-line
225 versus second-line therapy were year at UTI diagnosis, patient age, whether the provider was an
226 advanced specialist, internal medicine or family medicine doctor. The top 5 covariates predicting
227 first-line versus alternative therapy were age, whether the provider was an obstetrics /
228 gynecologist, and whether the individual had received beta-lactams in the previous 2 years.

229 Patients with UTI who were prescribed first-line antibiotics had a lower probability of an inpatient
230 and outpatient revisit within 30 days compared to those who received a second-line antibiotic
231 (adjusted risk difference = -1.8% [95% CI -2.4% to -1.1%]). Relative to alternative beta-lactam
232 treatments, patients prescribed first-line antibiotics for UTI had a 6.4% [95% CI -10.1% to -3.2%]
233 lower probability of inpatient or outpatient revisit at 30 days (Figure 4). For both comparisons,
234 these findings were driven largely by individuals with uncomplicated UTI, but were also observed
235 in those with pyelonephritis or sepsis to a lesser extent.



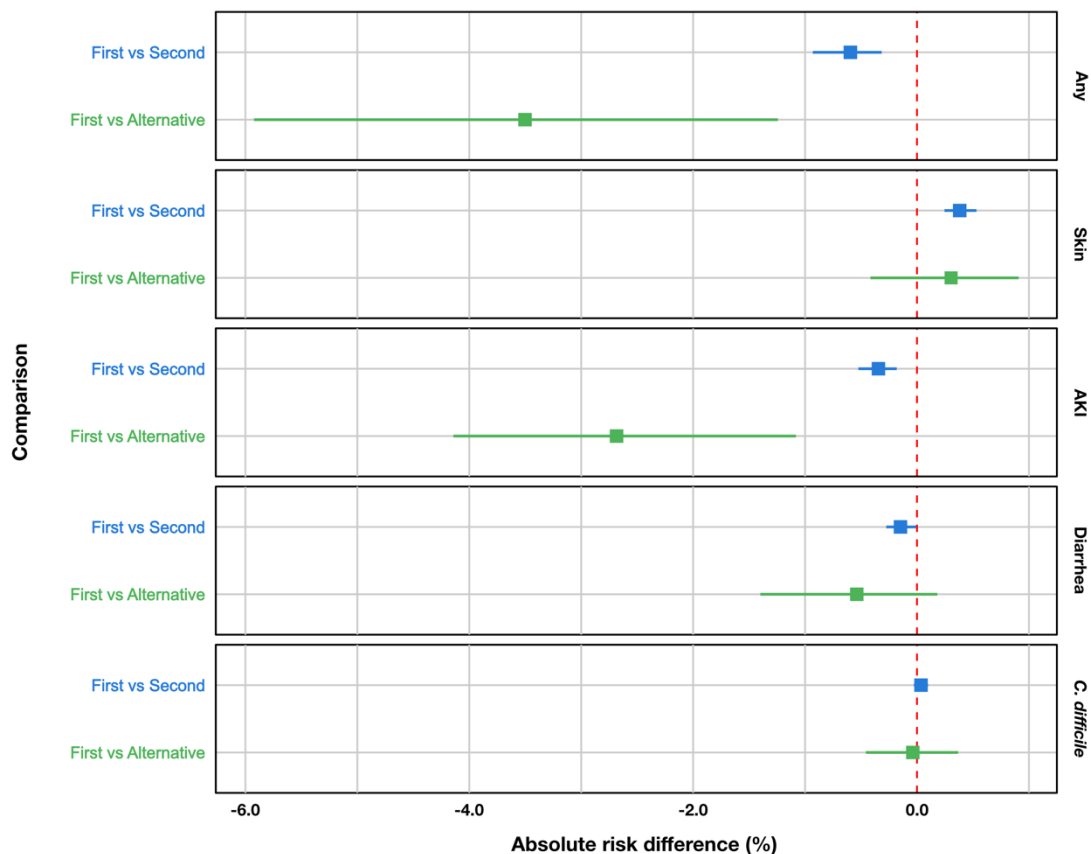
236

237 **Figure 4.** Adjusted rate difference for revisits for patients receiving first-line versus second-line
238 antibiotics, and first-line versus alternative treatments, after adjusting for potential confounding
239 factors and censoring.

240 Secondary outcomes

241 In terms of adverse events, receipt of first-line antibiotics was associated with a slightly increased
242 risk for skin-related adverse events (adjusted risk difference +0.4% [95% CI: +0.2% to +0.5%])
243 compared to second-line antibiotics and a decreased risk of acute kidney injury within 30 days of
244 treatment (adjusted risk difference -0.3% [95% CI: -0.5% to -0.2%]) (Figure 5). There was no
245 difference in the risk for *C. difficile* infection between the treatment groups.

246 Receipt of first-line antibiotics for UTI was associated with a lower risk of acute kidney injury at 30
247 days (adjusted risk difference -2.7% [95% CI: -4.1% to -1.1%]), relative to receiving an alternative
248 treatment (Figure 5). There was no difference in the risk of skin-related adverse events, diarrhea,
249 and *C. difficile* infection at 90 days.



250
251 **Figure 5.** Adjusted rate difference for treatment-related adverse effects for patients receiving first-
252 line versus second-line antibiotics, and first-line versus alternative treatments, after adjusting for
253 potential confounding factors and censoring.

254 Negative Control Outcomes and Sensitivity Analyses

255 There were no differences in the 1-month and 3-month risk for the three negative control
256 outcomes regardless of whether the patient received a first-line, second-line or alternative
257 treatment (Figure S3). We observed little to no difference between treatment arms in the
258 sensitivity analysis for patients who had an inpatient revisit within 30 days (1st line versus 2nd line
259 -0.1%, [95% CI: -0.3% to 0.0%]; 1st line versus alternative 0.3%, [95% CI -0.7% to 1.1%], Figures
260 S4). Additionally, results obtained from the model specified by [OMOP-learn](#) were similar across

261 all comparators and outcomes to the domain expert specified model (Supplementary Results and
262 Figures S5, S6 and S7). For the **OMOP-learn** model, first-line antibiotics had better efficacy than
263 second-line antibiotics as measured by lower risk of medical revisits (-2.1% [95% CI: -2.9% to -
264 1.6%]) overall, and in those with inpatient revisits (-0.2% [95% CI: -0.3% to 0.0%]). They also had
265 a lower overall risk of any adverse events (-0.7% [95% CI: -1.0% to -0.4%]) but a higher rate of
266 skin-related adverse events (adjusted risk difference: 0.3% [95% CI: 0.2% to 0.4%]). Similar
267 results were observed in the comparison between first-line antibiotics and alternative antibiotics
268 using the **OMOP-learn** derived model.

269 **DISCUSSION**

270 Using a large contemporary real-world dataset, we demonstrate that IDSA guidelines for
271 treatment of uncomplicated UTI remain robust in terms of both efficacy and adverse events,
272 despite major changes in the epidemiology of antibiotic resistance^{21,22}. Unless a patient has a
273 history of drug resistance, or intolerance or lives in a region where local rates of resistance are
274 high, nitrofurantoin and trimethoprim-sulfamethoxazole remain the treatments of choice. We
275 replicated our domain-expert derived results with an automated feature building package applied
276 to a common data model, thereby supporting the hypothesis that complex causal inference
277 analyses combined with careful cohort selection can be semi-automatable. This will help promote
278 reproducibility of our findings in other health systems and opens inquiry into other important
279 clinical questions.

280 We observed a small increase in rates of revisits for patients receiving second-line therapy relative
281 to those receiving first-line antibiotics. This result is surprising as fluoroquinolones are thought to
282 be equivalent or superior to nitrofurantoin and TMP-SMX in terms of clinical efficacy²³. The
283 differences were limited to outpatients with a diagnosis of lower urinary tract infection and were
284 much less pronounced for inpatients, suggesting the benefit of first-line treatments is restricted to
285 classic presentations of uncomplicated UTI. Follow up visits soon after treatment may be driven
286 by drug intolerance, toxicity or by selection of a drug to which an organism is resistant. The latter
287 may be a possible explanation for why people treated with nitrofurantoin and TMP-SMX had fewer
288 revisits. Recent work has suggested that rates of resistance to nitrofurantoin remain low despite
289 its widespread use and may be due to a high barrier to resistance²⁴. While resistance to TMP-
290 SMX is more common, clinicians are less likely to use this drug based on IDSA guidance that
291 recommends avoiding it when rates of local resistance exceed 20%, which is a common scenario
292 throughout the United States. In contrast, resistance to fluoroquinolones is most often mediated
293 by the accumulation of mutations in a single gene often in response to antibiotic exposure. Given
294 the high rate of fluoroquinolone prescription in the community, this may increase the risk for
295 prescribing an agent to which the agent is resistant. This is further complicated by the fact that
296 uncomplicated UTI, is often managed over telephone and without culture data. Lastly, given that
297 prescribers are prone to prescribe the same antibiotic²⁵⁻²⁷, the impact of prior exposure may be
298 more likely to lead to selection of resistance if that antibiotic is a fluoroquinolone and the patients
299 are otherwise healthy outpatients with a low risk for colonization by drug-resistant organisms.

300 We applied two approaches to feature construction to correct for confounding. Domain expert-
301 derived features are derived from expert knowledge on the biologic mechanisms of disease and
302 real-world experience with managing uncomplicated UTI. These features have the advantage of

303 theoretical backup from established pathophysiology and clinical data, but suffer from the
304 possibility of missing potential confounders, especially when the disease has diverse mechanistic
305 pathways or is not well-understood. In contrast, [OMOP-learn](#)¹⁴, which captures all information
306 available in the data without prior knowledge of its relationship with the disease, lowers the
307 probability of missing confounders but comes at the expense of including a large number of non-
308 relevant covariates. Our study provides an empirical demonstration that extracting features under
309 the [OMOP-learn](#) framework can yield conclusions comparable to that domain expert-derived
310 features, which supports application of causal inference methods using automatic feature
311 generation in the medical context.

312 Recent work has shown that carefully constructed retrospective cohorts with proper statistical
313 adjustment can provide robust results that complement findings from prospective randomized
314 controlled trials²⁸. However, as with all observational studies, there is a possibility that our results
315 may be biased due to residual confounding. We believe the degree of confounding is small as we
316 adjusted for both covariate-dependent censoring and treatment indication, which are the major
317 forms of confounding we expect to impact our results. This is further supported by the results of
318 the negative control outcome analysis, which shows an equal distribution of control outcomes
319 between treatment arms. The consistency in the strength and direction of our outcomes between
320 domain-expert derived and [OMOP-learn](#) derived features lends further strength to the validity of
321 our findings. The major strength of this study is the inclusion of a real-world dataset with a
322 comprehensive collection of covariates translated into a common data model. The rich set of
323 features permits construction of models that better specify causal mechanisms and the use of a
324 common data model enhances the study's reproducibility for other patient populations. Lastly,
325 large observational datasets offer the opportunity to gain real-world insight that is both up to date
326 and representative of the patients presenting with disease in practice today.

327 Other limitations of our study are that the prevalence of certain comorbidities is lower in our cohort
328 than in the general population. This may partly reflect the limited scope of our data, which comes
329 from a single health insurer primarily based in Southeast Pennsylvania but may also reflect our
330 inclusion criteria, which intentionally restricted our analyses to people with uncomplicated UTI.
331 We also had limited data on patient race, ethnicity and socioeconomic status, which precluded
332 our ability to assess for fairness across diverse subpopulations. Future work should seek to
333 reproduce our analysis using larger datasets with more diverse populations to ensure equity. The
334 increase in prescription of first-line antibiotics over time, which likely reflects the effect of guideline
335 dissemination and promotion of antibiotic stewardship²⁹⁻³¹, should not by itself bias outcomes,
336 assuming care practices did not dramatically change over the study period.

337 In conclusion, our results provide reassurance that guideline-concordant therapy remains the
338 optimal treatment decision for uncomplicated UTI. The application of an automated feature
339 extraction package for datasets translated into a common data model, combined with a rigorous
340 analytic pipeline, is a promising approach to assess the impact of guideline-directed therapy in
341 real-world populations and over time.

342 **FUNDING SUPPORT**

343 EH was supported by the National Science Foundation Graduate Research Fellowship Program
344 (grant no. 2141064). SDA was supported by the National Institute of Diabetes and Digestive and
345 Kidney Diseases (grant no. K12DK100024). DS and MCS were supported in part by grants from
346 Independence Blue Cross and the Office of Naval Research (grant no. N00014-21-1-2807). SK
347 was supported by AHRQ (grant no. K08 HS027841-01A1).

348

349 **CONFLICTS OF INTEREST**

350 SDA reports support from Centers for Disease Control and Prevention SHEPheRD
351 75D30121D12733-D5-E003 (grant no. 5U54CK000616-02), the Society for Healthcare
352 Epidemiology of America, and the Duke Claude D. Pepper Older Americans Independence
353 Center (National Institute on Aging grant no. P30AG028716), as well as consulting fees from
354 Locus Biosciences, Sysmex America, GlaxoSmithKline, bioMérieux, and the Infectious Diseases
355 Society of America. SDA became an employee of GSK/ViiV Healthcare one year after her
356 contribution to this project.

357 REFERENCES

- 358 1. Foxman, B., Barlow, R., D'Arcy, H., Gillespie, B. & Sobel, J. D. Urinary Tract Infection Self-
359 Reported Incidence and Associated Costs. *Ann Epidemiol* **10**, 509–515 (2000).
- 360 2. Shapiro, D. J., Hicks, L. A., Pavia, A. T. & Hersh, A. L. Antibiotic prescribing for adults in
361 ambulatory care in the USA, 2007–09. *J Antimicrob Chemoth* **69**, 234–240 (2014).
- 362 3. Gupta, K. *et al.* International Clinical Practice Guidelines for the Treatment of Acute
363 Uncomplicated Cystitis and Pyelonephritis in Women: A 2010 Update by the Infectious
364 Diseases Society of America and the European Society for Microbiology and Infectious
365 Diseases. Preprint at (2011).
- 366 4. MacDougall, C. *et al.* Pseudomonas aeruginosa, Staphylococcus aureus, and
367 Fluoroquinolone Use. *Emerg Infect Dis* **11**, 1197–1210 (2005).
- 368 5. Nseir, S. *et al.* First-generation fluoroquinolone use and subsequent emergence of multiple
369 drug-resistant bacteria in the intensive care unit* *Crit Care Med* **33**, 283–289 (2005).
- 370 6. Deshpande, A. *et al.* Community-associated Clostridium difficile infection and antibiotics: a
371 meta-analysis. *J Antimicrob Chemoth* **68**, 1951–1961 (2013).
- 372 7. Hooton, T. M. *et al.* Amoxicillin-Clavulanate vs Ciprofloxacin for the Treatment of
373 Uncomplicated Cystitis in Women: A Randomized Trial. *Jama* **293**, 949–955 (2005).
- 374 8. Arredondo-García, J. L. *et al.* Comparison of short-term treatment regimen of ciprofloxacin
375 versus long-term treatment regimens of trimethoprim/sulfamethoxazole or norfloxacin for
376 uncomplicated lower urinary tract infections: a randomized, multicentre, open-label, prospective
377 study. *J Antimicrob Chemoth* **54**, 840–843 (2004).
- 378 9. Raz, R. *et al.* Empiric use of trimethoprim-sulfamethoxazole (TMP-SMX) in the treatment of
379 women with uncomplicated urinary tract infections, in a geographical area with a high
380 prevalence of TMP-SMX-resistant uropathogens. *Clinical Infectious Diseases* **34**, 1165–1169
381 (2002).
- 382 10. Stein, G. E. Comparison of single-dose fosfomycin and a 7-day course of nitrofurantoin in
383 female patients with uncomplicated urinary tract infection. *Clin. Ther.* **21**, 1864–1872 (1999).
- 384 11. Minassian, M. A. *et al.* A comparison between single-dose fosfomycin trometamol
385 (Monuril®) and a 5-day course of trimethoprim in the treatment of uncomplicated lower urinary
386 tract infection in women. *Int. J. Antimicrob. Agents* **10**, 39–47 (1998).
- 387 12. Stang, P. E. *et al.* Advancing the Science for Active Surveillance: Rationale and Design for
388 the Observational Medical Outcomes Partnership. *Ann Intern Med* **153**, 600 (2010).
- 389 13. Elm, E. von *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology
390 (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457
391 (2007).

- 392 14. Kodialam, R. S. *et al.* Deep Contextual Clinical Prediction with Reverse Distillation. *arXiv*
393 (2020) doi:10.48550/arxiv.2007.05611.
- 394 15. Robins, J. M. & Finkelstein, D. M. Correcting for Noncompliance and Dependent Censoring
395 in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank
396 Tests. *Biometrics* **56**, 779–788 (2004).
- 397 16. Austin, P. C. Variance estimation when using inverse probability of treatment weighting
398 (IPTW) with survival analysis. *Stat. Med.* **35**, 5642–5655 (2016).
- 399 17. Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* (2017)
400 doi:10.48550/arxiv.1705.07874.
- 401 18. Wilcox, M. A., Villasis-Keever, A., Sena, A. G., Knoll, C. & Fife, D. Evaluation of disability in
402 patients exposed to fluoroquinolones. *BMC Pharmacol. Toxicol.* **21**, 40 (2020).
- 403 19. Vihta, K.-D. *et al.* Trends over time in Escherichia coli bloodstream infections, urinary tract
404 infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998–2016: a study of electronic
405 health records. *Lancet Infect. Dis.* **18**, 1138–1149 (2018).
- 406 20. Ku, J. H. *et al.* Multidrug Resistance of Escherichia coli From Outpatient Uncomplicated
407 Urinary Tract Infections in a Large United States Integrated Healthcare Organization. *Open*
408 *Forum Infect. Dis.* **10**, ofad287 (2023).
- 409 21. Pedrotti, C. H. S. *et al.* Antibiotic stewardship in direct-to-consumer telemedicine
410 consultations leads to high adherence to best practice guidelines and a low prescription rate. *Int*
411 *J Infect Dis* **105**, 130–134 (2021).
- 412 22. Palms, D. L. *et al.* Comparison of Antibiotic Prescribing in Retail Clinics, Urgent Care
413 Centers, Emergency Departments, and Traditional Ambulatory Care Settings in the United
414 States. *Jama Intern Med* **178**, 1267 (2018).
- 415 23. Trestioreanu, A. Z., Green, H., Paul, M., Yaphe, J. & Leibovici, L. Antimicrobial agents for
416 treating uncomplicated urinary tract infection in women. *Cochrane Db Syst Rev* **34**, CD007182
417 (2010).
- 418 24. Vallée, M., Harding, C., Hall, J., Aldridge, P. D. & TAN, A. Exploring the in situ evolution of
419 nitrofurantoin resistance in clinically derived uropathogenic Escherichia coli isolates. *J*
420 *Antimicrob Chemoth* **78**, 373–379 (2022).
- 421 25. Rodrigues, A. T., Roque, F., Falcão, A., Figueiras, A. & Herdeiro, M. T. Understanding
422 physician antibiotic prescribing behaviour: a systematic review of qualitative studies. *Int J*
423 *Antimicrob Ag* **41**, 203–212 (2013).
- 424 26. Livorsi, D., Comer, A., Matthias, M. S., Perencevich, E. N. & Bair, M. J. Factors Influencing
425 Antibiotic-Prescribing Decisions Among Inpatient Physicians: A Qualitative Investigation. *Infect.*
426 *Control Hosp. Epidemiology* **36**, 1065–1072 (2015).

- 427 27. Krockow, E. M. *et al.* Balancing the risks to individual and society: a systematic review and
428 synthesis of qualitative research on antibiotic prescribing behaviour in hospitals. *J. Hosp. Infect.*
429 **101**, 428–439 (2019).
- 430 28. Wang, S. V. *et al.* Emulation of Randomized Clinical Trials With Nonrandomized Database
431 Analyses. *Jama* **329**, 1376–1385 (2023).
- 432 29. Goebel, M. C., Trautner, B. W. & Grigoryan, L. The Five Ds of Outpatient Antibiotic
433 Stewardship for Urinary Tract Infections. *Clin Microbiol Rev* **34**, e00003-20 (2021).
- 434 30. Marcelin, J. R., Chung, P. & Schooneveld, T. C. V. Antimicrobial stewardship in the
435 outpatient setting: A review and proposed framework. *Infect Control Hosp Epidemiology* **41**,
436 833–840 (2020).
- 437 31. Nys, C. L. *et al.* Impact of Education and Data Feedback on Antibiotic Prescribing for
438 Urinary Tract Infections in the Emergency Department: An Interrupted Time-Series Analysis.
439 *Clin Infect Dis* (2022) doi:10.1093/cid/ciac073