

1 **Risk Factors for *Klebsiella* infections among hospitalized patients with Pre-Existing**
2 **Colonization**

3

4 **Running Title: Risk factors for *Klebsiella* infections**

5

6 Krishna Rao¹, Alieysa Patel², Yuang Sun², Jay Vornhagen², Jonathan Motyka¹, Abigail
7 Collingwood², Alexandra Teodorescu², Lili Zhao³, Keith S. Kaye¹, Michael Bachman^{2,4#}

8

9 ¹ Infectious Diseases Division, Department of Internal Medicine, Michigan Medicine; ²
10 Department of Pathology, Michigan Medicine; ³ Department of Biostatistics, School of Public
11 Health; ⁴ Department of Microbiology and Immunology, Michigan Medicine, Ann Arbor, MI

12

13 **# Corresponding Author:**

14 **7510E MSRB 1**

15 **1150 W Medical Center Drive**

16 **Ann Arbor, MI 48109**

17

18 **Abstract**

19 Background: *Klebsiella* commonly colonizes the intestinal tract of hospitalized
20 patients and is a leading cause of healthcare-associated infections. Colonization is
21 associated with subsequent infection, but the factors determining this progression are unclear.

22 Methods: Intensive care and hematology/oncology patients were screened for *Klebsiella*
23 colonization by rectal swab culture and monitored for infection for 90 days after a positive swab.
24 Electronic medical records were analyzed for patient factors associated with subsequent
25 infection, and variables of potential significance in bivariable analysis were used to build a final
26 multivariable model. Concordance between colonizing and infecting isolates was assessed by *wzi*
27 capsular gene sequencing.

28 Results: Among 2087 hospital encounters from 1978 colonized patients, 90 cases of infection
29 (4.3%) were identified. Mean time to infection was 20.6 ± 24.69 (range 0-91, median 11.5) days.
30 Of 86 typed cases, 68 unique *wzi* types were identified and 69 cases (80.2%) were colonized with
31 an isolate of the same type prior to infection. Based on multivariable modeling, overall
32 comorbidities, depression and low albumin level at the time of rectal swab were independently
33 associated with subsequent *Klebsiella* infection.

34 Conclusions: Despite the high diversity of colonizing strains of *Klebsiella*, there is high
35 concordance with subsequent infecting isolates and progression to infection is relatively quick.
36 Readily accessible data from the medical record could be used by clinicians to identify colonized
37 patients at increased risk of subsequent *Klebsiella* infection.

38 **Importance**

39 *Klebsiella* is a leading cause of healthcare-associated infections. Patients who are intestinally
40 colonized with *Klebsiella* are at significantly increased risk of subsequent infection, but only a
41 subset of colonized patients progress to disease. Colonization offers a potential window of
42 opportunity to intervene and prevent these infections, if the patients at greatest risk could be
43 identified. To identify patient factors associated with infection in colonized patients, we studied
44 1978 colonized patients. We found that patients with a higher burden of underlying disease in
45 general, depression in particular, and low albumin in a blood test were more likely to be a case of
46 infection. However, these variables did not completely predict infection, suggesting that other
47 host and microbial factors may also be important. The average time to infection was 3 weeks,
48 suggesting that there is time to intervene and prevent *Klebsiella* infections in hospitalized
49 patients.

50

51 **Introduction**

52 *Klebsiella pneumoniae* is a leading cause of healthcare-associated infections, commonly causing
53 pneumonia, bacteremia, and urinary tract infections (1). With advancements in genomics, it is
54 clear that infections attributed to *K. pneumoniae* may be caused by related members of a multi-
55 species complex (henceforth *Klebsiella*). Clinical laboratories can now identify *K. variicola*,
56 which may have clinical characteristics distinct from *K. pneumoniae*, but not other complex
57 members such as *K. quasipneumoniae* (2, 3) . *Klebsiella* is also a common colonizer of the
58 gastrointestinal tract (4), creating a reservoir of potential disease-causing bacteria. In fact, among
59 colonized patients who develop *Klebsiella* infections, the majority of isolates match those found
60 in the gastrointestinal tract (5, 6). This concordance between colonizing and invasive isolate can
61 be measured by sequencing of the *wzi* capsular gene with the same discriminatory power as 7-
62 gene multi-locus sequence typing (6). This tight link between colonization and infection suggests
63 an opportunity for prevention of infection in hospitalized patients, where detection of
64 colonization could trigger an intervention.

65
66 Despite the strong link between colonization and infection, the risk factors for infection in
67 colonized patients are unclear. A combination of patient and bacterial variables likely determines
68 if a patient will develop a *Klebsiella* infection. At a population level without assessment of
69 colonization, increased age, male sex, dialysis, chronic liver disease, solid organ transplant, and
70 cancer were associated with *Klebsiella* bacteremia (7). *Klebsiella* colonization itself is
71 associated with advanced age as well as vancomycin-resistant enterococci (VRE) colonization
72 (8). We previously found certain *Klebsiella* genes and patient factors associated with *Klebsiella*

73 infection compared to asymptomatic colonization (9). However, this study was limited by a
74 small sample size and unknown colonization status in some infected patients.

75

76 The goal of this study was to identify risk factors for *Klebsiella* infection among colonized
77 patients at the time of detectable colonization by a rectal swab. In a cohort study of over 2000
78 hospital encounters with colonized patients, the bacterial species and concordance with
79 subsequent infecting strains was determined. Patient variables were assessed in bivariable
80 analysis and used to create an explanatory model of *Klebsiella* infection. Our findings indicate
81 that readily accessible chart variables can be used to stratify risk of *Klebsiella* infection in
82 colonized patients.

83 **Results**

84 Swab collection and *Klebsiella* colonization

85 From 5/1/2017 to 12/13/2018, rectal swabs collected for VRE screening were cultured for
86 *Klebsiella*, identifying 2098 positive swabs from 1978 patients (recruitment goal=2000).

87 Eliminating duplicate swabs from the same patient within 90 days of the enrollment swab, the
88 cutoff for a subsequent infection to meet our case definition,, resulted in a final cohort consisting
89 of 2087 inpatient admissions with baseline *Klebsiella* colonization among 1978 patients. There
90 was a noticeable seasonal variation in rates of colonization (**Figure 1**) that appeared to co-vary
91 with the mean daily temperature with increased colonization rates occurring in warmer months.

92 Up to 3 isolates were archived from each colonized patient, yielding 5463 *Klebsiella* isolates. Of
93 2087 patient encounters, 1712 admissions (82.03%) were colonized with *K. pneumoniae* only,
94 308 (14.76%) were colonized with *K. variicola* only, and 67 (3.21%) admissions were colonized
95 with both *K. pneumoniae* and *K. variicola*.

96 97 Clinical cultures and case determinations

98 There were 402 clinical cultures positive for *Klebsiella* from 262 patients, of which 386 (95.8%)
99 were *K. pneumoniae* and 16 (3.97%) were *K. variicola*. These cultures were assessed for cases of
100 *Klebsiella* infection based on chart review. 52 blood cultures were reviewed, and 51 (98.1%)
101 were considered unique cases of infection (one subject had two positive blood cultures and only
102 the first was counted), of which 37 met the 90-day cutoff and were not duplicates (Table 1). 199
103 urine cultures were reviewed and 46 (23.1%) were considered cases of infection, 22 of which
104 met the 90-day cutoff and were not duplicates. 72 respiratory cultures were reviewed and 40
105 (55.6%) were considered cases of infection, of which 27 met the 90-day cutoff and were not

106 duplicates. 79 cultures from other body sites were reviewed and 15 (18.8%) were considered
107 cases of infection, of which 4 met the 90-day cutoff and were not duplicates (1 abscess, 2 bile,
108 and 1 tissue). In total, there were 90 infections meeting case definitions for a unique episode
109 within 90 days of the enrollment swab (1 infection occurred 91 days from the index swab but
110 was still included in the analysis). This represents a rate of infection of 4.3% out of 2087
111 hospital encounters from 1978 *Klebsiella* colonized patients. Bloodstream infections were most
112 frequent (37, 41.1%) followed by respiratory (27, 30.0%), urine (22, 24.4%) and other sites (4,
113 4.4%).

114

115 Most infections were caused by *K. pneumoniae* (86.7%). However, there was high strain
116 diversity with 68 unique *wzi* types observed across 90 infections. Despite this diversity, the
117 majority of these infections were concordant with a *Klebsiella* isolate from the enrollment swab
118 (69/86, 80.2%, 4 cases had missing *wzi* data). Urinary tract infections (UTI) had the highest rate
119 of concordance (86.3%), bacteremia had a similar rate (81.1%) and pneumonia was lowest (63%;
120 $P=0.11$) The mean time from swab to infection was 20.6 ± 24.69 (median 11.5, range 0-91) for
121 all infections and $17.7 \text{ days} \pm 21.98$ (median 10, range 0-90) for concordant cases (**Table 1**). The
122 time to concordant infection (**Figure 2**) differed by site of infection overall ($P = .001$) and was
123 significantly longer for UTI (32.8 ± 27.56) than for bacteremia (12.0 ± 17.88 ; $P < 0.001$) or
124 pneumonia (13.4 ± 14.93 ; $P = 0.006$).

125

126 Among all cases of infection, 19 (21.1%), 13 were found to be extended spectrum beta-lactamase
127 positive (7 bloodstream infections, 4 respiratory, and 2 UTI), and all 13 infecting strains were *K.*
128 *pneumoniae*. In addition, 6 cases were carbapenem resistant, including 3 bloodstream and 3

129 respiratory infections. 5 of these strains were *K. pneumoniae* and the other was *K. variicola* .

130 Overall, 17 of these antibiotic-resistant cases were concordant with a colonizing rectal swab.

131

132 Clinical variables and invasive infection with *Klebsiella*

133 The initial unadjusted analyses compared cases to controls and numerous variables showed

134 statistically significant differences (**Table 2**). There were no significant differences in age, sex,

135 or race between cases and controls. The season that *Klebsiella* colonization was detected on the

136 baseline rectal swab did not associate with increased risk of infection ($P = 0.358$). Comorbid

137 illness, as assessed by a Weighted Elixhauser score, and high-risk antibiotic exposure were both

138 enriched in cases compared to controls (**Table 2**). Notably, depression has a neutral weight for

139 predicting in-hospital mortality in the Elixhauser score, but it was significantly associated with

140 *Klebsiella* infection and thus considered separately for the multivariable modeling described

141 below. Invasive devices including urinary catheters and feeding tubes actually had an inverse

142 association with infection and ventilator at baseline trended towards an inverse association

143 ($P = 0.1$). A number of medications, including diuretics, antidepressants, Vitamin D,

144 pressors/inotropes, were also associated with *Klebsiella* infection.

145

146 Multivariable Model of *Klebsiella* infection

147 The purposeful selection process resulted in the explanatory model for *Klebsiella* infection

148 shown in **Table 3**. Here the baseline features of weighted Elixhauser score, depression, and

149 serum albumin < 2.5 g/dL associated with an increased risk of subsequent *Klebsiella* infection,

150 and use of pressors/inotropes showed a borderline association with increased risk. Prior diuretic,

151 vitamin D, and high-risk antibiotic use did not independently associate with the outcome, but

152 were retained in the model as confounders. Although depression was selected, anti-depressants
153 were not. The overall fit of the model by area under the receiver operator characteristic curve
154 (AuROC) was 0.74. The mean predicted risk of infection in cases was 8.61% vs. 4.35% in
155 controls, for a discrimination slope of 4.26% (95% confidence interval [CI] 2.59%–5.94%).
156
157 For the 69 cases of infection that were concordant by *wzi* type with the colonizing isolate, we
158 built a separate multivariable model, shifting the cases without concordance to the control group
159 (**Table 4**). Here the weighted Elixhauser score, depression, and serum albumin <2.5 g/dL were
160 again associated with an increased risk of infection, while baseline serum hemoglobin showed a
161 borderline association such that reduced hemoglobin increased risk. Other potential confounders
162 were not retained. The model AuROC was 0.73, and the mean predicted risk of infection in cases
163 was 5.6% vs. 3.44% in controls for a discrimination slope of 2.16% (95% CI 1.31%–3%). We
164 also conducted a subset analysis by species, and on unadjusted analysis, infection with *K.*
165 *pneumoniae* vs. *K. variicola* was associated with an increased risk of infection that did not reach
166 statistical significance (OR 1.81 [0.92, 4.1], $P = .114$).

167 **Discussion**

168 This study of over 2000 hospital encounters provides a comprehensive analysis of the dynamics
169 of *Klebsiella* infection in colonized patients and identifies patient factors that are present at
170 admission and associated with infection. Overall, a small but significant percentage of colonized
171 patients (4.3%) progress to infection. These infections are caused by diverse strains of *Klebsiella*
172 and are predominantly caused by strains present in a rectal swab (80.2%) at the time of
173 admission. When infection occurs in colonized patients, it is within days to weeks (median 11.5
174 days) from the first positive rectal swab. The risk of infection is influenced, at least in part, by
175 characteristics of the colonized patients themselves. In particular, the overall burden of
176 comorbidities, pre-existing diagnosis of depression, and a low albumin at baseline are significant
177 and independently associated with infection. Identification of colonized patients at high-risk of
178 *Klebsiella* infection at the time of admission could be used to direct interventions to prevent
179 these infections in our most vulnerable patients.

180

181 The dynamics of colonization confirm and expand on previous studies showing a strong link
182 between *Klebsiella* colonization and infection. Indeed, in a smaller pilot study we demonstrated
183 that intensive care patients were frequently colonized (23%), colonization was associated with
184 infection (5.2% vs 1.3% uncolonized patients; OR 4.01), the colonizing isolates were diverse,
185 and infections were caused often by colonizing strains (13/16 paired isolates matched) (6).
186 Similarly, Gorrie et al found colonization rates of 5.9-19% among intensive care patients, a
187 strong association between colonization and infection (OR 6.9; 13 infections among 27
188 colonized patients) and high diversity of colonizing and infecting isolates (5). The current study
189 focusing on colonized patients refines the results of these smaller studies, demonstrating a

190 similar rate of infections (4.3%), high diversity of infecting strains and a more accurate estimate
191 of concordant infections (80.2%). It also demonstrates that detectable colonization does not
192 capture all of the infection risk, and that patient variables modify this risk.

193
194 The predominance of concordant infections among colonized patients, and the fact that the
195 infecting strain is often present at admission, indicates that infection prevention for *Klebsiella*
196 should focus on endogenous, patient risk factors. To that end, our study determined that
197 comorbidities, medication history, and routine laboratory results at admission are informative
198 regarding infection risk, with cases having twice the predicted risk compared to controls (8.61%
199 vs. 4.35%). Importantly, the model for concordant infection (**Table 4**) was more parsimonious
200 than the model for all infections (**Table 3**). Infection risk increased with the Elixhauser
201 comorbidity score, and depression contributed substantially (OR 2.04, CI 1.2-3.41; $P=0.007$).
202 While it is not clear in our study if the association is causal, exposure to antidepressant
203 medications is known to significantly alter the gut microbiome (10). Additionally, prior research
204 has identified depression and antidepressant medication use as risk factors *C. difficile* infection,
205 an infection that is mediated through disruption of the gut microbiome (11). Although
206 antidepressant use was associated with *Klebsiella* infection on unadjusted analysis, it was
207 confounded by other variables. When we added antidepressant use to the final model, it was not
208 statistically significant ($P=.66$), nor was it when we replaced depression with antidepressant use
209 in the model ($P=0.3$). It is plausible that the gut microbiome mediates the association between
210 depression and subsequent *Klebsiella* infection that we observed in colonized patients, but this
211 requires further study.

212

213 The association between low albumin and infection (OR 2.0, CI 1.15-3.48, P=0.013) was also
214 previously observed for *Klebsiella* infections in the ICU including both colonized and non-
215 colonized patients (6). Low albumin can be caused by decreased synthesis due to malabsorption,
216 malnutrition or hepatic dysfunction, or due to losses from ascites, nephropathy, or enteropathy
217 (12). Although the pathophysiologic link between low albumin and *Klebsiella* infection is
218 unclear, this is an informative and readily widely available predictor, ordered by the physician in
219 > 90% of the cohort.

220
221 In this multi-year study, the rates of colonization wax and wane by month with the highest rates
222 in the summer. This seasonal pattern was previously demonstrated for *Klebsiella* bacteremia
223 across 4 continents (13), with both higher temperature and humidity being associated with
224 increased rates. The observed seasonality of *Klebsiella* colonization may in turn cause a
225 seasonality of infection. Indeed, the proportion of cases by season varied from a low of 8.9% in
226 Winter to a high of 38.9% in Summer (**Table 2**), but tracked closely with the overall proportion
227 of colonized patients by season and the risk of infection in a given colonized patient did not vary
228 by season ($P = 0.358$). The seasonality of *Klebsiella* bacteremia is not seen for the related
229 pathogens *Escherichia coli* or *Serratia* (13). Therefore, prevention efforts that focus on
230 identifying environmental reservoirs of *Klebsiella* that fluctuate with temperature could have a
231 significant impact on overall rates of infection through reduction of the initial colonization event
232 rate.

233
234 Our study has several limitations including its retrospective design focused on a convenience
235 sample of ICU/oncology ward patients with post-hoc data extraction. We undertook several

236 strategies to mitigate this central limitation. We only included variables that we felt sure were
237 baseline features present at colonization and not a reflection of emerging or already present
238 infection (**Supplementary Methods**), and for variables such as vitals, laboratory results, and
239 medication/device exposures we either excluded them from modeling, captured them only if
240 present prior to swab collection, and/or manually confirmed their baseline status by chart review.
241 We also used matching and multivariable modeling to reduce bias/confounding, and we
242 conducted several sensitivity analyses to confirm our main results. We also could not confirm
243 associations for variables that in prior research were found to be associated with *Klebsiella*
244 infection. A population-based study of *Klebsiella* bacteremia found male sex, advanced age,
245 dialysis, solid-organ transplantation, chronic liver disease, and cancer were associated with
246 infection (7). We previously found that advanced age was associated with colonization (8). Like
247 with temperature and seasonality, some infection risks may actually represent colonization risks.
248 Although the comorbidities were not categorized in the same way across studies, none of the
249 other risk factors were significantly associated with infection in colonized patients
250 (**Supplementary Results**). Finally, technical limitations arising from the sensitivity of our
251 plating method and picking 3 colonies per plate for characterization could have led us to miss
252 some patients who were colonized at baseline and to miss a low abundance, concordant strain in
253 infected patients with detectable colonization.
254
255 Although the models of *Klebsiella* infection had significant explanatory power, there are likely to
256 be additional factors that modulate the risk of infection in colonized patients. Relative abundance
257 of *Klebsiella* colonization is also associated with infection, even after controlling for the
258 variables identified here (co-submitted manuscript), and it could further refine infection risk in

259 colonized patients. In addition, the complex genomics of *Klebsiella* is likely to be an important
260 predictor of infection, and we have previously identified several putative pathogenicity loci (9).
261 The present study can further inform such future work by identifying the most important clinical
262 variables that should be considered as confounders when assessing the independent contribution
263 of bacterial genetic loci, colonization density and other features of the gut microbial community
264 on risk of subsequent *Klebsiella* infection. Such a holistic approach to modeling the risk of
265 infection, from host characteristics to the gut community to the genetics of the colonizing strain,
266 may be what is needed to develop the most accurate and useful models of invasive *Klebsiella*
267 infection.
268

269 **Methods**

270 Cohort identification, sample collection, and strain typing

271 This study was approved by the Institutional Review Boards of the University of Michigan
272 Medical School (IRBMED). We sequentially enrolled subjects to establish a cohort of patients
273 with *Klebsiella* (*K. pneumoniae* or *K. variicola*) rectal colonization. Inpatients admitted to the
274 intensive care units [ICUs] and the oncology wards of the University of Michigan hospitals in
275 Ann Arbor, Michigan, from May 2017 to September 2018, were eligible for inclusion if
276 discarded vancomycin resistant *Enterococcus* [VRE] screening rectal swab media was available
277 for culture. All patients admitted to these wards undergo screening for rectal VRE carriage per
278 hospital policy, using a flocced swab placed into Amies transport media (E-swab, Becton
279 Dickinson, Franklin Lakes, NJ). Subjects were subsequently enrolled into our study if *K.*
280 *pneumoniae* was isolated from the swab media through plating on MacConkey agar followed by
281 taxonomic identification using MALDI TOF mass spectrometry. Up to 3 isolates per rectal swab
282 culture were archived along with all clinical culture isolates positive for *K. pneumoniae*. Wzi
283 typing was done using PCR and Sanger sequencing (14). Wzi type was assigned by uploading
284 the consensus gene sequence to the BigsDB database (<https://bigsdb.pasteur.fr>).

285

286 Colonization rates by month

287 The percentage of rectal swabs containing *Klebsiella* was calculated for each month from records
288 of MacConkey agar growth and MALDI-TOF results for species identification. Average daily
289 temperature for the month was obtained for Ann Arbor from www.wunderground.com.

290

291 Clinical data extraction and case definitions

292 Structured queries were used to abstract clinical data from the electronic medical record
293 including demographics, pre-existing comorbidities, vitals, laboratory results, imaging and
294 procedure reports. These clinical data were used to construct variables comprised of baseline
295 features present at or before the time of swab collection. As described further in our results,
296 certain variables were collapsed into colligative variables for the unadjusted and adjusted
297 analyses. Positive clinical cultures for *Klebsiella* species from the day of, to 90 days after, rectal
298 swab collection were independently reviewed for alignment with our case definitions by two
299 clinician study team members (J.B. and M.B), and disagreements were adjudicated by a third
300 clinician team member (K.R.). All patients with a blood culture growing *Klebsiella* were
301 considered to have an infection. Manual chart review was conducted by the study team to decide
302 if potential cases met published infection criteria from professional societies / governmental
303 organizations for pneumonia or urinary tract infection (15-19), and cultures from other sites,
304 such as wounds, were considered cases based on the clinical judgment of reviewers. For those
305 meeting clinical case definitions of infection, the clinical isolate and preceding rectal swab
306 isolates were evaluated for concordance of *Klebsiella* isolates by *wzi* gene sequencing as
307 previously described (6, 14). We have previously demonstrated that *wzi* sequencing has similar
308 discriminatory power to 7-gene multi-locus sequence typing (6).

309

310 Statistical Methods

311 Initial descriptive statistics, visualizations such as histograms, and bivariable relationships
312 between predictors and the primary outcome of invasive infection were calculated and used to
313 inform variable construction and inclusion for multivariable modeling. Bivariable relationships
314 were assessed with Student's t-tests (or ANOVA first if multiple categories) and Chi-squared

315 tests. Given the primary outcome occurring at low frequency and the resulting statistical power
316 considerations, we sought to collapse variables into colligative metrics where applicable. For
317 example, a weighted Elixhauser score (20) was included for modeling in lieu of most individual
318 comorbidities, save when the weight was set to zero or if the weight differed notably in the
319 direction/magnitude from what was observed on bivariable associations with the primary
320 outcome. High-risk antibiotics were defined as third- or fourth-generation cephalosporins,
321 fluoroquinolones, lincosamides, β -lactam/ β -lactamase inhibitor combinations, oral vancomycin,
322 and carbapenems, based on their association with disruption of the intestinal microbiome (21).
323 We first constructed an explanatory model to identify independent predictors of the primary
324 outcome. We did this using a modified purposeful selection approach (22) to allow for inclusion
325 of variables in the final model that adjust for confounding in other covariates, even if not
326 independently associated with the primary outcome themselves (see **Supplementary Methods**).
327 Following this, we constructed additional models to test hypotheses about specific risk factors
328 that were not selected by the above procedure and to conduct sensitivity analyses using the
329 primary model from the above procedure. The sensitivity analyses included limiting the case
330 definition to only apply when isolates from clinical infections were concordant with the
331 colonizing strain by *wzi* sequencing; and assessing how colonizing species influences risk of
332 infection. All analyses were conducted in R version 4.0.3 (R Foundation for Statistical
333 Computing, Vienna, Austria). A cutoff *P*-value of $<.05$ was used for statistical inference. Model
334 performance was further assessed using the area under the receiver operator characteristic curve
335 (AuROC) and the discrimination slope using the R package *pROC* (23).

336

337 **Data Availability**

338 Deidentified data from human subjects may be made available upon request, pending approval
339 from the University of Michigan Institutional Review Board.

340 **Acknowledgements**

341 Research reported in this publication was supported by the National Institute Of Allergy And

342 Infectious Diseases of the National Institutes of Health under Award Number R01AI125307. The

343 content is solely the responsibility of the authors and does not necessarily represent the official

344 views of the National Institutes of Health.

345 **References**

- 346 1. Magill SS, Edwards JR, Bamberg W, Beldavs ZG, Dumyati G, Kainer MA, Lynfield R,
347 Maloney M, McAllister-Hollod L, Nadle J, Ray SM, Thompson DL, Wilson LE, Fridkin
348 SK. 2014. Multistate point-prevalence survey of health care-associated infections. *N Engl*
349 *J Med* 370:1198-208.
- 350 2. Wyres KL, Lam MMC, Holt KE. 2020. Population genomics of *Klebsiella pneumoniae*.
351 *Nat Rev Microbiol* 18:344-359.
- 352 3. Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Naucler P, Brisse S, Giske
353 CG. 2014. *Klebsiella variicola* is a frequent cause of bloodstream infection in the
354 stockholm area, and associated with higher mortality compared to *K. pneumoniae*. *PLoS*
355 *One* 9:e113539.
- 356 4. Selden R, Lee S, Wang WL, Bennett JV, Eickhoff TC. 1971. Nosocomial *Klebsiella*
357 infections: intestinal colonization as a reservoir. *Ann Intern Med* 74:657-64.
- 358 5. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, Pratt NF,
359 Garlick JS, Watson KM, Pilcher DV, McGloughlin SA, Spelman DW, Jenney AWJ, Holt
360 KE. 2017. Gastrointestinal Carriage Is a Major Reservoir of *Klebsiella pneumoniae*
361 Infection in Intensive Care Patients. *Clin Infect Dis* 65:208-215.
- 362 6. Martin RM, Cao J, Brisse S, Passet V, Wu W, Zhao L, Malani PN, Rao K, Bachman MA.
363 2016. Molecular Epidemiology of Colonizing and Infecting Isolates of *Klebsiella*
364 *pneumoniae*. *mSphere* 1:e00261-16.
- 365 7. Meatherall BL, Gregson D, Ross T, Pitout JD, Laupland KB. 2009. Incidence, risk
366 factors, and outcomes of *Klebsiella pneumoniae* bacteremia. *Am J Med* 122:866-73.

- 367 8. Collingwood A, Blostein F, Seekatz AM, Wobus CE, Woods RJ, Foxman B, Bachman
368 MA. 2020. Epidemiological and Microbiome Associations Between *Klebsiella*
369 *pneumoniae* and Vancomycin-Resistant Enterococcus Colonization in Intensive Care
370 Unit Patients. *Open Forum Infect Dis* 7:ofaa012.
- 371 9. Martin RM, Cao J, Wu W, Zhao L, Manthei DM, Pirani A, Snitkin E, Malani PN, Rao K,
372 Bachman MA. 2018. Identification of Pathogenicity-Associated Loci in *Klebsiella*
373 *pneumoniae* from Hospitalized Patients. *mSystems* 3:mSystems00015-18.
- 374 10. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR,
375 Fernandez KC, Dose H, Mori H, Patil KR, Bork P, Typas A. 2018. Extensive impact of
376 non-antibiotic drugs on human gut bacteria. *Nature* 555:623-628.
- 377 11. Rogers MA, Greene MT, Young VB, Saint S, Langa KM, Kao JY, Aronoff DM. 2013.
378 Depression, antidepressant medications, and risk of *Clostridium difficile* infection. *BMC*
379 *Med* 11:121.
- 380 12. Mc Pherson RA. 2007. Specific Proteins, p 236. *In* McPherson RA, Pincus MR (ed),
381 Henry's Clinical Diagnosis and Management by Laboratory Methods. Saunders Elsevier,
382 Philadelphia, PA.
- 383 13. Anderson DJ, Richet H, Chen LF, Spelman DW, Hung YJ, Huang AT, Sexton DJ, Raoult
384 D. 2008. Seasonal variation in *Klebsiella pneumoniae* bloodstream infection on 4
385 continents. *J Infect Dis* 197:752-6.
- 386 14. Brisse S, Passet V, Haugaard AB, Babosan A, Kassis-Chikhani N, Struve C, Decre D.
387 2013. *wzi* Gene sequencing, a rapid method for determination of capsular type for
388 *Klebsiella* strains. *J Clin Microbiol* 51:4073-8.

- 389 15. American Thoracic S, Infectious Diseases Society of A. 2005. Guidelines for the
390 management of adults with hospital-acquired, ventilator-associated, and healthcare-
391 associated pneumonia. *Am J Respir Crit Care Med* 171:388-416.
- 392 16. CDC. 2016. Urinary tract infection (catheter-associated urinary tract infection [CAUTI]
393 and non-catheter-associated urinary tract infection [UTI]) and other urinary system
394 infection [USI]) events. Centers for Disease Control and Prevention, Atlanta, GA.
- 395 17. Hooton TM, Bradley SF, Cardenas DD, Colgan R, Geerlings SE, Rice JC, Saint S,
396 Schaeffer AJ, Tambayh PA, Tenke P, Nicolle LE, Infectious Diseases Society of A.
397 2010. Diagnosis, prevention, and treatment of catheter-associated urinary tract infection
398 in adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases
399 Society of America. *Clin Infect Dis* 50:625-63.
- 400 18. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, Dowell
401 SF, File TM, Jr., Musher DM, Niederman MS, Torres A, Whitney CG, Infectious
402 Diseases Society of A, American Thoracic S. 2007. Infectious Diseases Society of
403 America/American Thoracic Society consensus guidelines on the management of
404 community-acquired pneumonia in adults. *Clin Infect Dis* 44 Suppl 2:S27-72.
- 405 19. Centers for Disease Control and Prevention. . . Centers for Disease Control and
406 Prevention A, GA., 16. 2014. National Healthcare Safety Network device-associated
407 module: ventilator-associated events. Centers for Disease Control and Prevention. Centers
408 for Disease Control and Prevention, Atlanta, GA.
- 409 20. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. 2009. A modification of
410 the Elixhauser comorbidity measures into a point system for hospital death using
411 administrative data. *Med Care* 47:626-33.

- 412 21. Baggs J, Jernigan JA, Halpin AL, Epstein L, Hatfield KM, McDonald LC. 2018. Risk of
413 Subsequent Sepsis Within 90 Days After a Hospital Stay by Type of Antibiotic Exposure.
414 Clin Infect Dis 66:1004-1012.
- 415 22. Bursac Z, Gauss CH, Williams DK, Hosmer DW. 2008. Purposeful selection of variables
416 in logistic regression. Source Code Biol Med 3:17.
- 417 23. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. 2011. pROC:
418 an open-source package for R and S+ to analyze and compare ROC curves. BMC
419 Bioinformatics 12:77.
- 420

421 **Tables**
422

Table 1. Summary of *Klebsiella* infections by species and strain type

	Bacteremia	Pneumonia	Urinary tract infection	Other	Total
Cases of infection (% of total)	51 (33.6)	40 (26.3)	46 (30.3)	15 (9.9)	152
Unique cases within 90 days of enrollment swab (% of total)	37 (41.1)	27 (30.0)	22 (24.4)	4 (4.4)	90
Species causing infection (n,% of unique cases by site)					
<i>K. pneumoniae</i>	33 (89.2)	24 (88.9)	18 (81.8)	3 (75.0)	78 (86.7)
<i>K. variicola</i>	4 (10.8)	3 (11.1)	4 (18.2)	1 (25.0)	12 (13.3)
Total unique wzi types	29	25	21	3	68
Unknown wzi type	1	0	0	1	2
Cases with concordant rectal swab within 90 days prior to infection (n,% of unique cases by site)	30 (85.7)	17 (65.4%)	19 (86.3)	3 (100.0)	69 (80.2)
mean +/- SD (median, range) days between concordant	12.0 ± 17.88 (5, 0-	13.4 ± 14.93 (11, 0-42)	32.8 ± 27.56	2.7 ± 3.79 (1,	17.7 ± 21.98 (10,

colonization and infection	67)		(24, 1- 90)	0-7)	0-90)
----------------------------	-----	--	----------------	------	-------

423

424

Table 2. Demographics and selected baseline¹ characteristics

	Colonized (N=1997)	Case (N=90)	<i>P</i>
Age			
Mean (SD)	60.4 (15.9)	60.9 (12.9)	0.701
Median [Min, Max]	63.0 [1.00, 100]	62.0 [25.0, 86.0]	
Gender			
Female	883 (44.2%)	41 (45.6%)	0.802
Male	1114 (55.8%)	49 (54.4%)	
Race			
Non-white	335 (16.8%)	12 (13.3%)	0.391
White	1662 (83.2%)	78 (86.7%)	
Weighted Elixhauser Score			
Mean (SD)	17.1 (11.9)	22.6 (11.1)	<0.00 1
Median [Min, Max]	16.0 [-14.0, 70.0]	22.5 [1.00, 51.0]	
Depression			
No	1558 (78.0%)	59 (65.6%)	0.006
Yes	439 (22.0%)	31 (34.4%)	
Uncomplicated Diabetes			
No	1529 (76.6%)	62 (68.9%)	0.094
Yes	468 (23.4%)	28 (31.1%)	
Complicated Hypertension			

Table 2. Demographics and selected baseline¹ characteristics

	Colonized (N=1997)	Case (N=90)	<i>P</i>
No	1375 (68.9%)	54 (60.0%)	0.077
Yes	622 (31.1%)	36 (40.0%)	
Uncomplicated Hypertension			
No	947 (47.4%)	52 (57.8%)	0.054
Yes	1050 (52.6%)	38 (42.2%)	
Urinary catheter at baseline			
Yes	1162 (58.2%)	62 (68.9%)	0.044
No	835 (41.8%)	28 (31.1%)	
Feeding tube at baseline			
Yes	73 (3.7%)	7 (7.8%)	0.046
No	1924 (96.3%)	83 (92.2%)	
Ventilator at baseline			
Yes	760 (38.1%)	42 (46.7%)	0.1
No	1237 (61.9%)	48 (53.3%)	
Central venous catheter at baseline			
Yes	878 (44.0%)	37 (41.1%)	0.593
No	1119 (56.0%)	53 (58.9%)	
Prior diuretic use			
No	1684 (84.3%)	58 (64.4%)	<0.001
Yes	313 (15.7%)	32 (35.6%)	

Table 2. Demographics and selected baseline¹ characteristics

	Colonized (N=1997)	Case (N=90)	P
Prior PPI use			
No	1637 (82.0%)	58 (64.4%)	<0.00 1
Yes	360 (18.0%)	32 (35.6%)	
Prior vitamin D use			
No	1832 (91.7%)	69 (76.7%)	<0.00 1
Yes	165 (8.3%)	21 (23.3%)	
Prior use of pressors/inotropes			
No	1854 (92.8%)	66 (73.3%)	<0.00 1
Yes	143 (7.2%)	24 (26.7%)	
Prior use of antidepressants/antipsychotics			
No	1738 (87.0%)	66 (73.3%)	<0.00 1
Yes	259 (13.0%)	24 (26.7%)	
Prior use of histamine antagonists			
No	1753 (87.8%)	68 (75.6%)	<0.00 1
Yes	244 (12.2%)	22 (24.4%)	
Prior high risk antibiotic use²			
No	1699 (85.1%)	56 (62.2%)	<0.00 1
Yes	298 (14.9%)	34 (37.8%)	
Baseline circulating WBC (thousands of cells/microliter)			
Mean (SD)	6.48 (5.64)	5.60 (3.78)	0.037

Table 2. Demographics and selected baseline¹ characteristics

	Colonized (N=1997)	Case (N=90)	P
Median [Min, Max]	6.00 [0.0900, 160]	5.50 [0.0900, 15.5]	
Missing	14 (0.7%)	0 (0%)	
Baseline serum hemoglobin (g/dL)			
Mean (SD)	8.74 (2.30)	7.58 (1.79)	<0.00 1
Median [Min, Max]	8.30 [3.60, 17.4]	7.00 [4.70, 13.4]	
Missing	14 (0.7%)	0 (0%)	
Baseline circulating platelets (thousands of cells/microliter)			
Mean (SD)	145 (97.9)	122 (94.0)	0.023
Median [Min, Max]	134 [0, 627]	107 [1.00, 380]	
Missing	14 (0.7%)	0 (0%)	
Baseline serum creatinine (mg/dL)			
Mean (SD)	0.870 (0.781)	0.772 (0.495)	0.078
Median [Min, Max]	0.700 [0.0900, 15.8]	0.660 [0.160, 3.61]	
Missing	19 (1.0%)	0 (0%)	
Serum albumin			
<2.5 g/dL	353 (17.7%)	39 (43.3%)	<0.00 1
>=2.5 g/dL	1456 (72.9%)	47 (52.2%)	
Missing	188 (9.4%)	4 (4.4%)	
Site of clinical infection			
other	0 (0%)	4 (4.4%)	-

Table 2. Demographics and selected baseline¹ characteristics

	Colonized (N=1997)	Case (N=90)	<i>P</i>
blood	0 (0%)	37 (41.1%)	-
respiratory	0 (0%)	27 (30.0%)	-
Urine	0 (0%)	22 (24.4%)	-
none	1997		-
Season of baseline rectal swab colonization			0.38
Winter (January-March)	248 (12.4%)	8 (8.9% of cases, 3.1% of Winter swabs)	
Spring (April-June)	503 (25.2%)	28 (31.1% of cases, 5.3% of Spring swabs)	
Summer (July-September)	717 (35.9%)	35 (38.9% of cases, 4.7% of Summer swabs)	
Fall (October-December)	529 (26.5%)	19 (21.1% of cases, 3.5% of Fall swabs)	

¹Features considered to be baseline if present >48 hours but <90 days prior to rectal swab collection.

²High-risk antibiotics were defined as third- or fourth-generation cephalosporins, fluoroquinolones, lincosamides, β -lactam/ β -lactamase inhibitor combinations, oral vancomycin, and carbapenems.

428

Table 3. Model of *Klebsiella* infection in colonized patients.

Variable	OR [95% CI]	P
Elixhauser Score (weighted)	1.02 [1.00, 1.04]	.034
Depression	1.80 [1.12, 2.88]	.014
Prior ¹ diuretic use	1.28 [0.70, 2.29]	.407
Prior ¹ vitamin D use	1.36 [0.71, 2.57]	.343
Prior ¹ use of pressors/inotropes	1.9 [0.96, 3.71]	.062
Prior ¹ use of high-risk antibiotics	1.38 [0.74, 2.52]	.298
Albumin < 2.5 g/dL	2.13 [1.3, 3.48]	.003

¹Features considered to be baseline if present >48 hours but <90 days prior to rectal swab collection.

429

430

431

Table 4. Model for concordant *Klebsiella* infections (retaining cases without concordance)

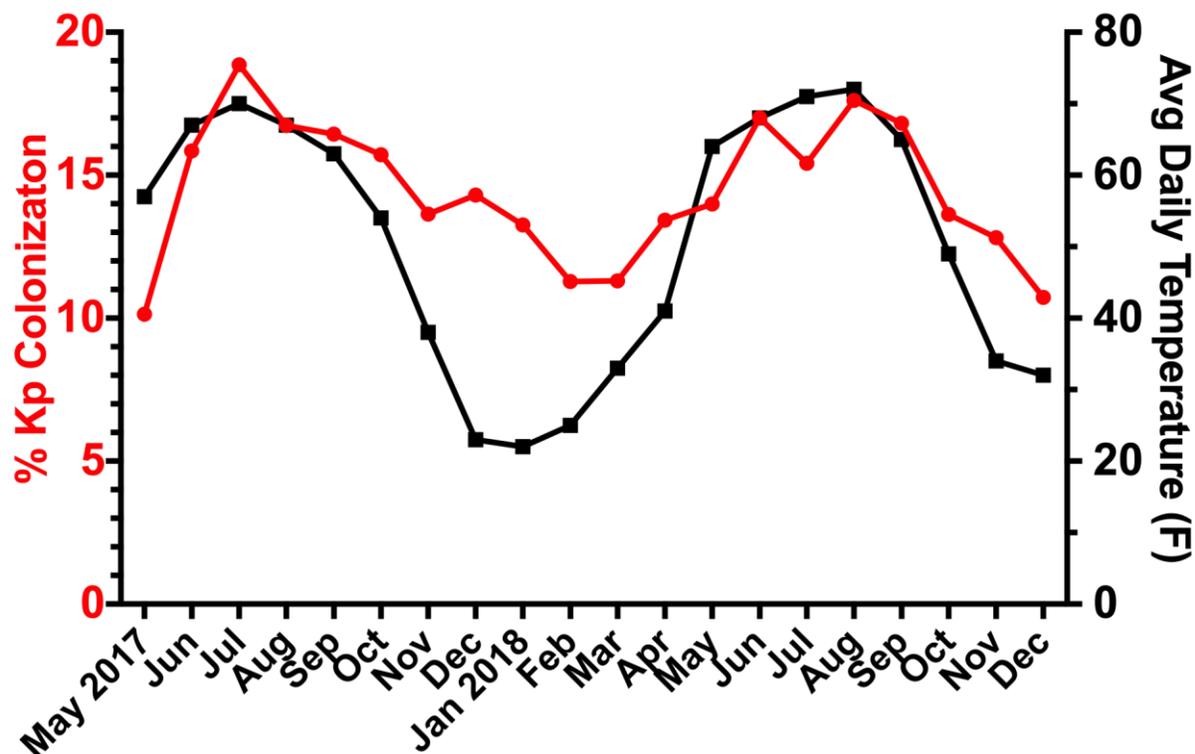
Variable	OR [95% CI]	P
Weighted Elixhauser	1.03 [1.01, 1.05]	.01
Depression	2.04 [1.2, 3.41]	.007
Baseline ¹ Hemoglobin	0.88 [0.75, 1.01]	.073
Albumin <2.5 g/dL	2 [1.15, 3.48]	.013

¹Features considered to be baseline if present >48 hours but <90 days prior to rectal swab collection.

432

433

434 **Figures**
435



436

437

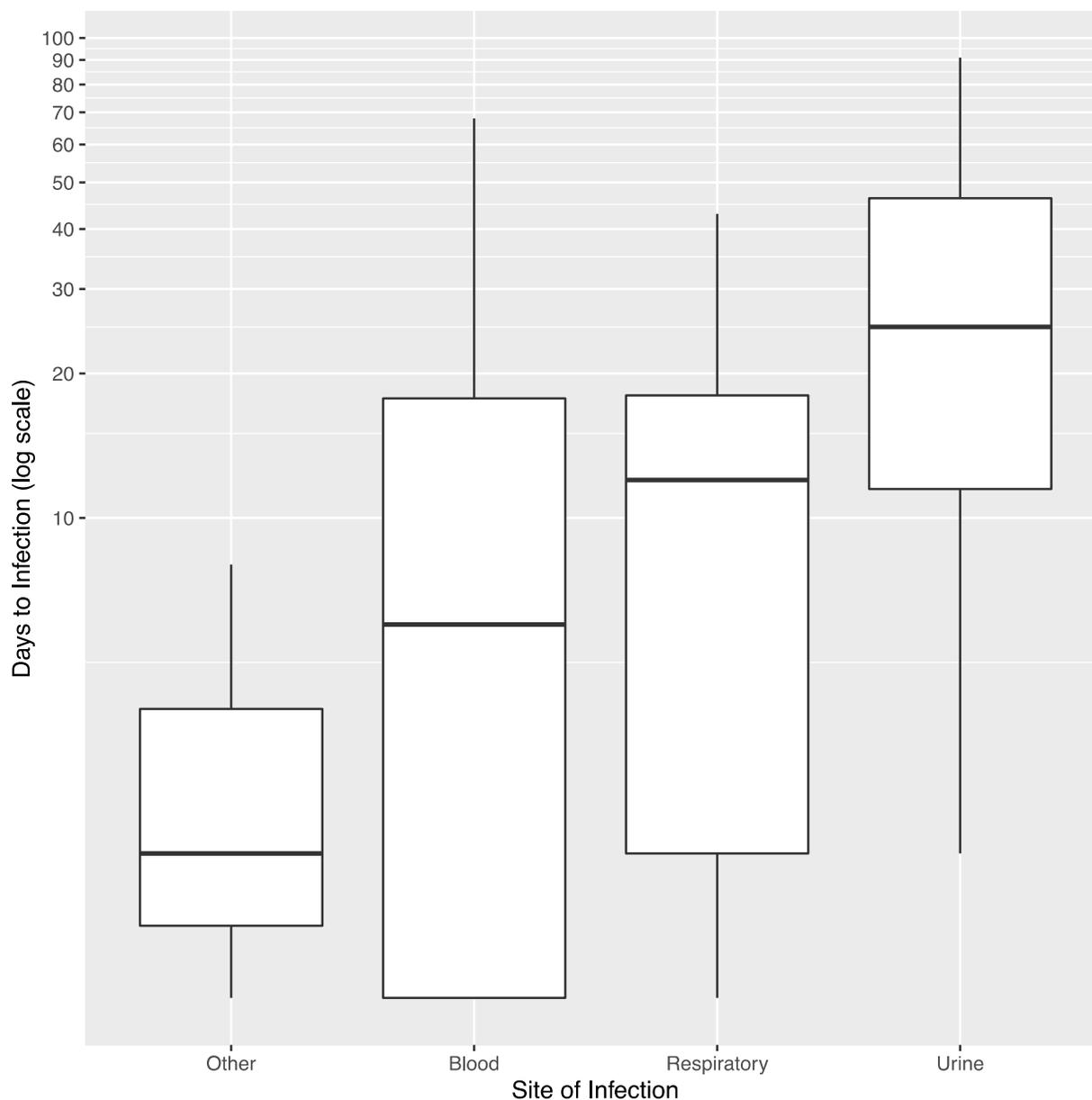
438 **Figure 1: Colonization rate and average daily temperature by month of enrollment.** For
439 each month of the study, the percentage of rectal swabs positive for *Klebsiella* is shown, along
440 with the average daily temperature for that month.

441

442

443

444 **Figure 2. Boxplots of time to concordant infection vs. site of infection.** Given the significant
445 right skew of the data and for better visibility, the number of days is plotted on a log scale. The
446 boxes show the medians and interquartile ranges (IQRs), while the whiskers show 1.5x the IQRs.



447
448