

1 Instrumental variable analysis: choice of control variables is
2 critical and can lead to biased results.

3

4 Fergus Hamilton^{1,2}, Todd C Lee³, Guillaume Butler-Laporte³

5

6 1. MRC Integrative Epidemiology Unit, University of Bristol

7 2. North Bristol NHS Trust

8 3. Division of Infectious Diseases, Department of Medicine, McGill University,

9 Montreal, Canada

10

11

12 Corresponding author:

13

14 Fergus Hamilton;

15 Fergus.hamilton@bristol.ac.uk

16

17 Abstract:

18

19 Instrumental variable (IV) analysis is a widely used technique in econometrics to
20 estimate causal effects in the presence of confounding. A recent application of this
21 technique was used in a high-profile analysis in *JAMA Internal Medicine* to estimate the
22 effect of cefepime, a broad-spectrum antibiotic, on mortality in severe infection. There
23 has been ongoing concern that piperacillin-tazobactam, another broad-spectrum
24 antibiotic with greater anaerobic activity might be inferior to cefepime, however this has
25 not been shown in randomized controlled trials. The authors used an international
26 shortage of piperacillin-tazobactam as an instrument, as during this shortage period,
27 cefepime was used as an alternative. The authors report a strong mortality effect (5%
28 absolute increase) with piperacillin-tazobactam. In this paper, we closely examine this
29 estimate and find it is likely conditional on inclusion of a control variable (metronidazole
30 usage). Inclusion of this variable is highly likely to lead to collider bias, which we show
31 via simulation. We then generate estimates unadjusted for metronidazole which are
32 much closer to the null and may represent residual confounding or confounding by
33 indication. We highlight the ongoing challenge of collider bias in empirical IV analyses
34 and the potential for large biases to occur. We finally suggest the authors consider
35 including these unadjusted estimates in their manuscript, as the large increase in
36 mortality reported with piperacillin-tazobactam is unlikely to be true.

37 Introduction:

38

39 In a recent issue of *JAMA Internal Medicine*, Chanderraj et al¹ provide an innovative
40 analysis attempting to use a shortage of a widely used broad-spectrum antimicrobial,
41 piperacillin-tazobactam (TZP), to conduct a comparative effectiveness study versus
42 cefepime (CPM), an agent used as a replacement during this period which has a similar
43 spectrum of activity except less anaerobic activity. Their instrumental variable (IV)
44 analysis¹ utilises the ‘natural experiment’ of substitution of TZP with CPM over this
45 period in a manner which is, in principle, unrelated to other confounders. They
46 hypothesize that the additional anti-anaerobic activity of TZP might lead to increased
47 mortality via effects on the microbiome or resistome. Their primary analysis using a
48 two-stage-least-squares (TSLS) IV analysis found that TZP use was associated with a 5%
49 absolute increase in mortality (95% CI 1.8-8.1%, $p = 0.002$). This was reported as:

50

51 *“administration of piperacillin-tazobactam was associated with higher mortality and*
52 *increased duration of organ dysfunction compared with cefepime”*

53

54 This is in stark contrast to recent randomized controlled trial data directly comparing
55 the two drugs which do not suggest such a benefit.² Chanderraj et al’s result was widely
56 discussed (Altmetric score 457 on 2nd July) and highlighted in *NEJM Journal Watch*³,
57 which concluded:

58

59 *“Therefore, we can infer that among severely ill patients with undifferentiated sepsis,*
60 *treatment with cefepime rather than PTZ [piperacillin-tazobactam] is associated with an*
61 *absolute mortality reduction at 90 days.”*

62

63 Interestingly, despite a marked reduction in piperacillin-tazobactam and increase in
64 cefepime usage during the shortage period, there was no independent effect of the
65 shortage period on mortality (20.8% outside shortage, 19.8% within shortage, $p = 0.31$).
66 This is unusual and raises a concern. In general, TSLS estimates are more imprecise and
67 have larger standard errors than the independent IV-outcome analysis reflecting the

68 uncertainty involved in the twostep process.⁴ Less formally, there are few plausible
69 explanations whereby all three of the below hold:

70

- 71 1. The instrument (shortage period) strongly influences the exposure (cefepime
72 usage)
- 73 2. The instrument (shortage period) has no influence on the outcome (mortality)
- 74 3. The exposure (cefepime usage) has a very strong effect on the outcome
75 (mortality)

76

77 In this short analysis, we explored potential reasons for the difference between the TSLS
78 estimate (suggesting strong evidence of benefit) and the raw analysis (suggesting no
79 effect) and postulate that this has occurred due to erroneous inclusion of control
80 variables in the analysis. Control variables are often included to increase precision of
81 TSLS estimation but should only be included under certain conditions and when they
82 are upstream of both the instrument and the outcome.⁵ For example, an analysis using
83 state or country level cigarette taxation as an instrument for the outcome of lung cancer
84 might include socio-economic status of the participant as a control variable to increase
85 precision. Or similarly, analyses using genetic variants as instruments might include
86 age of participants as control variables, as this cannot be plausible caused by genetics,
87 but may influence the outcome.

88

89 However, the inclusion of control variables that are downstream of the IV should
90 generally be avoided⁵, and can be considered as being similar to inappropriately
91 adjusting for post-randomisation variables in a randomised trial.⁶ In the presented
92 analyses, the authors adjust for a number of demographic variables but also control for
93 metronidazole usage in both stages of the regression. This is despite the authors
94 reporting that the shortage period strongly influences metronidazole prescription
95 (eFigure 2 in the original publication).

96

97 This is inappropriate and has the potential to lead to collider bias⁷ if unmeasured
98 confounding affects both MTZ usage and mortality which is probable, as metronidazole
99 usage is a marker of sickness (particularly when cefepime is used, as metronidazole is

100 often co-prescribed for intra-abdominal infection where mortality is high). This collider
101 bias induces an association between the IV and confounders (e.g. disease severity) of
102 the association between the exposure (cefepime usage) and outcome (mortality). This
103 directly contravenes a core assumption of IV analyses.

104

105 We first show via simulation why the addition of metronidazole into the IV analysis
106 could lead to bias. We then present our own analysis using summary data from the trial
107 to estimate the causal effect unadjusted for control variables.

108

109

110 Methods:

111

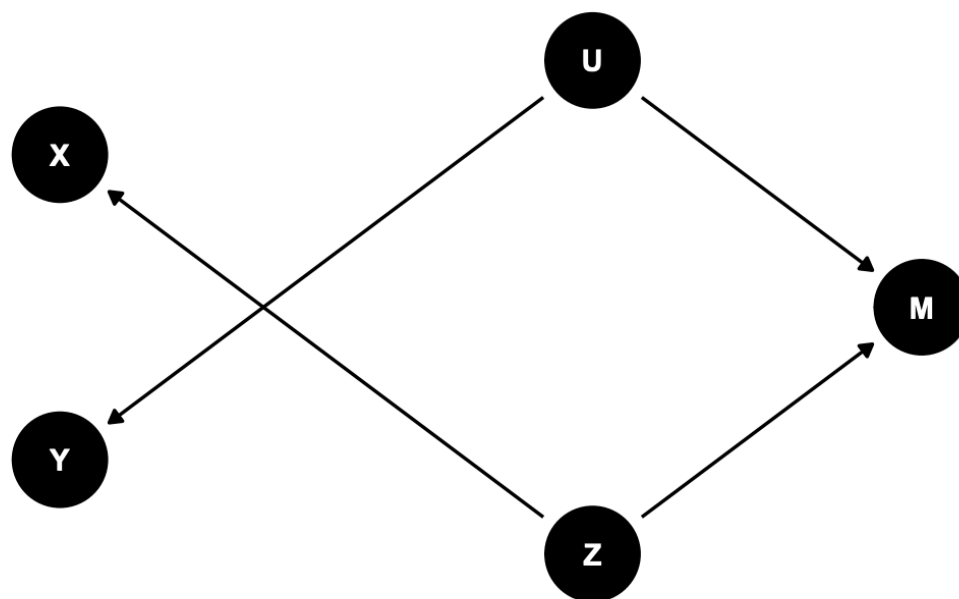
112 *Simulation study*

113

114 For our simulation study, we aimed to match the potential real-life causal data-
115 generating process, although we recognise such a simulation is an oversimplification,
116 and the major value is in demonstrating the potential collider bias. We generated 5
117 simulated variables, reflecting the key components: Z (the shortage period,
118 representing the instrument), X (the exposure, use of cefepime), M (metronidazole
119 usage), Y (the outcome, mortality), U (unmeasured confounding). We set up the
120 relationship between these variables as per **Figure 1**.

121

122 **Figure 1:** A directed acyclic graph (DAG) describing our simulation. Each point
123 represents a variable, while each arrow represents the simulated effect direction.



124

125

126 In this situation, we have a strong relationship between Z and X (as expected, the
127 shortage period led to a large increase in cefepime prescribing), but we also have a
128 strong relationship between Z and M (as was reported in their eFigure2). We also
129 simulated no causal effect of the exposure on the outcome (i.e. the prescription of
130 cefepime has no effect on mortality). This can be seen by the lack of an arrow between

131 X and Y. Finally, we simulated unmeasured confounding (U) which influences both
132 mortality and metronidazole usage. This variable represents the fact that metronidazole
133 usage is influenced by other factors that themselves associate with mortality (e.g. UTI
134 (no metronidazole, lower mortality) vs. intrabdominal or undifferentiated sepsis
135 (metronidazole, higher mortality)). We can strongly suspect this is the case, as
136 metronidazole users had a strong marginal association with mortality in the reported
137 logistic regression analysis (OR 1.54, 95% CI 1.32, 1.8, data from eTable 6 in the original
138 publication¹).

139
140 Simulated variables were generated in R for 5,000 participants, using the *tidyverse*
141 package, and IV analysis was performed using the AER package to generate estimates.
142 Further technical details and statistical code are provided in **Supplementary Note 1**. Of
143 note, we can also simulate the equally plausible situation where M is entirely
144 downstream of X, rather than Z (as the decision to give metronidazole is made after the
145 decision to give piperacillin-tazobactam, and there is no independent effect of Z onto M.
146 This does not fundamentally alter results but is reported in Supplementary Note 1 for
147 completeness.

148
149 The important detail in this simulation is that we should not have any evidence of an
150 effect in IV analyses; there is no relationship between the exposure and the outcome,
151 and evidence of a relationship is a sign of bias.

152
153 *Generation of summary level estimates from the original manuscript*

154
155 Although estimates in IV analyses are usually made using individual level data, it is
156 possible to generate estimates using summary level data. This is commonly performed
157 in Two-Sample Mendelian randomisation, where only the summary effects of a genetic
158 variant on an exposure and outcome are required.⁸

159
160 We can therefore estimate, accepting a degree of error, a TSLS estimate that is
161 unadjusted for the control variables in Chanderraj et al by simply calculating the IV-
162 outcome estimate, and dividing this by the IV-exposure estimate (the Wald ratio). The

163 standard error of this estimate can be approximated by dividing the standard error of the
164 IV-outcome estimate by the IV-exposure estimate. We calculate this (full details in the
165 **Supplementary Note 2**) and compare this with the presented analyses.

166

167

168 **Ethics**

169

170 This analysis is a re-analysis of presented work and requires no ethical approval

171

172 **Funding**

173

174 FH's time was funded by the NIHR Clinical Lectureship scheme.

175 Results:

176

177 *Simulation study*

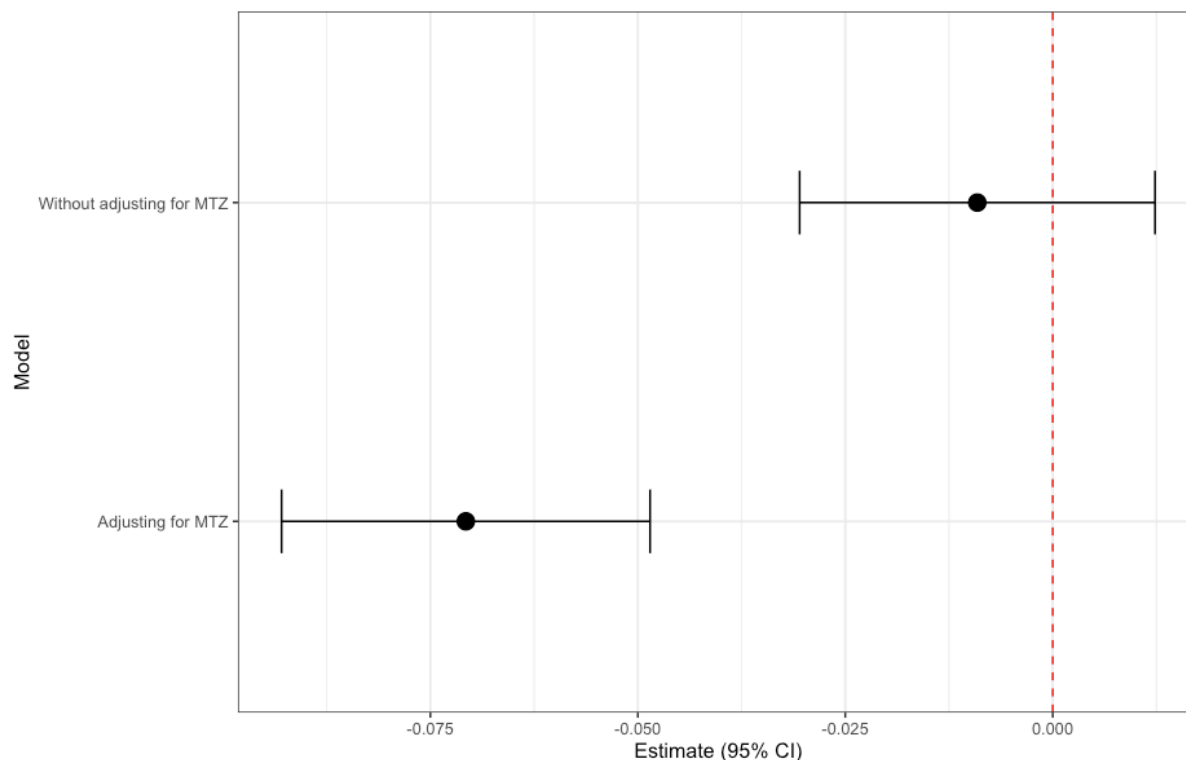
178

179 In our simulation study, we aimed to closely replicate the Chanderraj et al’s analysis.¹ In
180 the presented paper, they report that the IV strongly influences metronidazole (M)
181 usage, alongside having a large effect on the exposure (X). In our simulation, we have no
182 causal effect of the exposure on the outcome, so when analyses are run unadjusted for
183 metronidazole, IV estimates are null (**Figure 2**).

184

185

186 **Figure 2:** Simulated estimates of the effect of adjusting for metronidazole on an
187 outcome using the causal model above. In the unadjusted model, causal estimates are
188 close to the null, and are unbiased. In the adjusted model, estimates are biased and
189 show a causal effect of X (“cefepime”) on the outcome.



190

191

192 However, when we include metronidazole in the model we induce a bias, and we get
193 strong evidence of a protective effect of cefepime. This occurs despite there being no

194 causal effect of cefepime on the outcome. The reasons for this counterintuitive result
195 are worth discussing: the bias occurs because we have opened a path between the
196 unmeasured confounding (U) and the outcome. Essentially, we bring back most of the
197 bias that IV analyses are meant to protect against (unmeasured confounding).
198 Consider that the patient only gets metronidazole during the shortage if the clinician
199 thinks the patient is sick enough to require coverage for anaerobes which cefepime
200 itself does not provide. By adjusting for metronidazole, we induce a negative
201 relationship between patient sickness and the instrument that then biases the estimate
202 on mortality.

203
204 This bias occurs in the same way when performing analyses stratified by metronidazole
205 usage. Because metronidazole usage associates with mortality, in both metronidazole
206 users and non-users, there is a negative relationship induced between the instrument
207 and mortality which leads to biased estimates. Other examples of similar collider
208 biases have been widely described in COVID-19 and are likely to explain (some) of the
209 ‘obesity paradox’, the paradoxical relationship between increased BMI and better
210 critical disease outcomes in severe infection.⁹ It is important to note that this bias also
211 occurs if the causal path we present is incorrect, and metronidazole usage is entirely
212 downstream of piperacillin-tazobactam usage, and the IV has no direct effect on this
213 **(Supplementary Note 1)**.

214
215 In summary, our simulation study shows that in the presence of unmeasured
216 confounding of metronidazole usage and mortality, adjusting for metronidazole will
217 induce bias in estimates of the effect of cefepime on mortality.

218
219 *Generation of new estimates*

220
221 Although our simulation study provided support for potential bias, to test this we
222 needed to generate estimates unadjusted for metronidazole. We attempted to use the
223 reported summary level data and estimated the causal effect of cefepime on mortality.

224

225 Details of this calculation are provided in **Supplementary Note 2** but can be
226 summarised as calculating the effect of the shortage on piperacillin-tazobactam usage
227 (IV-exposure effect, -0.659) and the effect of the shortage on mortality (IV-outcome
228 effect , -0.0650) using logistic regression, and using the Wald ratio to generate the IV
229 estimate of -0.065/-0.659, which corresponds to a log odds ratio of 0.098, and an odds
230 ratio of 1.10, 95% CI 0.91-1.33. Formulas for standard error are in the **Supplementary**
231 **Note 2**. The absolute change in mortality is then calculated using the odds ratio and
232 baseline mortality.

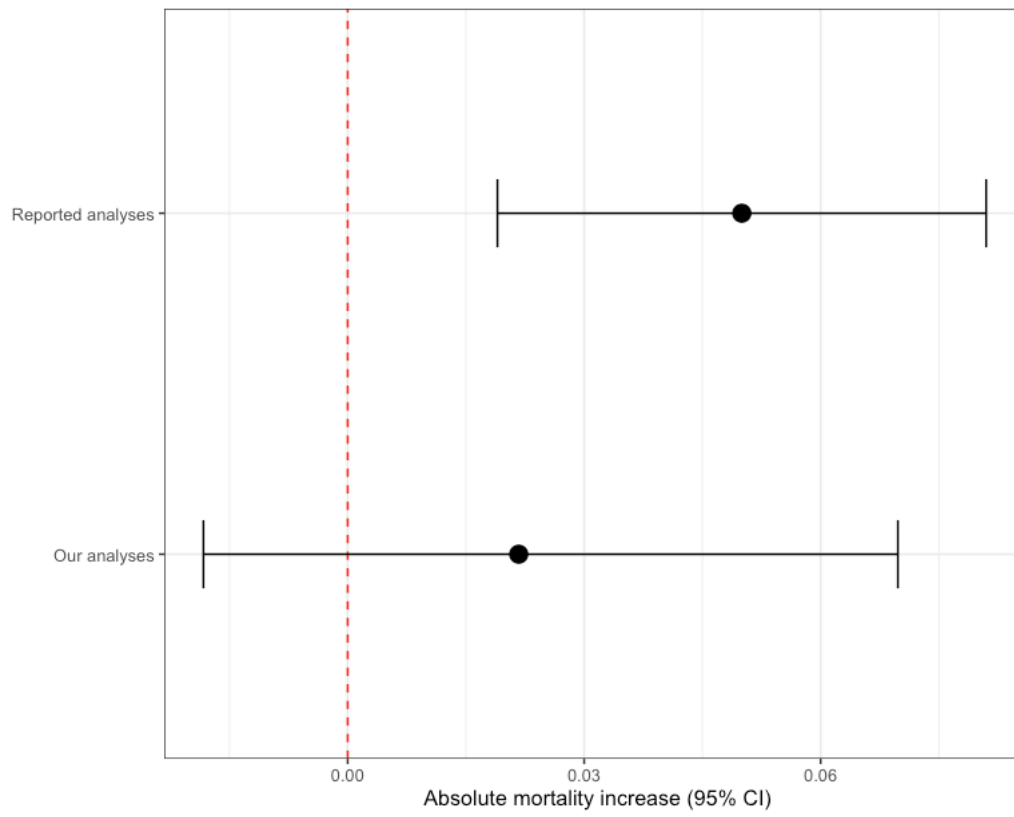
233

234 Our final estimate is an absolute risk increase of 2.2% with piperacillin-tazobactam
235 usage (95% CI -1.8% to 7.0%), weaker and more imprecisely estimated than their
236 primary estimate (5%; 95% CI 1.8-8.1%, **Figure 3**). This does not provide compelling
237 evidence of a strong effect of superiority of cefepime. We recognise the limitations of
238 this analysis, but are unable to do more detailed analysis without access to individual
239 level data.

240

241 **Figure 3:** Absolute risk increase associated with piperacillin-tazobactam usage
242 reported in the original manuscript, and with our estimates from summary level data.
243 We recognise these are likely to be imprecise than estimates using individual level data.

244



245

246

247

248

249 **Discussion**

250

251 Instrumental variable analyses can allow for causal inference when randomised trials
252 are lacking, unethical, or impractical. Although in principle simple analytically, it is clear
253 from our results above that they can be seriously biased when inappropriate control
254 variables are included.

255

256 In this particular case, we think it is clear that the addition of metronidazole into the IV
257 estimation is inappropriate, as it is a clinically relevant decision which is downstream of
258 the instrumental variable. This can be seen as adjusting for a post-randomisation
259 covariate in a randomised trial, which is explicitly warned against.

260

261 We recognise our simulations are based on one specified causal model which is an
262 oversimplification of the true model. We do not state with certainty that the model we
263 present represents truth, but it is certain that metronidazole usage is associated with
264 more severe disease (i.e. there is unmeasured confounding), and that metronidazole
265 usage is strongly associated with the IV. Using domain specific knowledge, our causal
266 model seems plausible but we would encourage the authors to present further data
267 supporting or refuting our analysis. We did ask the authors to provide analyses
268 unadjusted for metronidazole which would help confirm if collider bias was occurring
269 or whether our causal model was appropriate but they could not provide this data citing
270 institutional regulations, so we are limited to our summary analysis, which does provide
271 evidence that the unadjusted effect is much close to the null.

272

273 The authors do additionally present a secondary analysis not discussed above: whereby
274 the exposure is the receipt of *either* piperacillin-tazobactam or metronidazole, as both
275 of these have anaerobic activity. We note that the effect reported in this analysis is even
276 stronger (12% increase in absolute mortality), and highly implausible given the null
277 effect in the unadjusted analyses for piperacillin-tazobactam alone. This also raises
278 challenges for IV analyses: the exposure is not a single drug, so we cannot state
279 estimates represents the causal effect of increasing piperacillin-tazobactam or
280 metronidazole individually: rather it represents the causal effect of “increased

281 piperacillin-tazobactam and metronidazole usage” as a combined variable and cannot
282 be interpreted as saying that giving more of either drug individually would lead to
283 increased mortality. As with the primary analysis, it is likely that bias is generated via
284 unmeasured confounding in this combined variable. It is worth thinking of the
285 plausibility of situations whereby the causal effect would be predicted to almost double
286 by changing the exposure definition to include a composite exposure that is less
287 compellingly related to the IV (which was a shortage of piperacillin-tazobactam only).

288

289 The data reported from Chanderraj et al is therefore actually highly consistent with the
290 recent ACORN trial², and should be used as evidence that there is unlikely to be a large
291 difference in mortality between piperacillin-tazobactam and cefepime based on these
292 two studies. This is reassuring, and contrary to some of the reports surrounding this
293 paper,³ clinicians should not favour cefepime simply based on this evidence.

294

295 We strongly suggest that consideration should be given to reporting the unadjusted
296 results in the original manuscript.¹ Currently, readers of the manuscript might assume
297 that cefepime is a far superior agent to piperacillin-tazobactam which is not likely to be
298 the case.

299

300 More broadly, we make the point that IV analyses are challenging to perform, and
301 susceptible to biases like other observational studies. We recommend that authors,
302 editors, and readers, interpret IV analyses cautiously, particularly when results appear
303 to be contingent on a particular analytical set-up. The reason for our investigation of this
304 issue was the identification that there was no convincing relationship between the
305 shortage period and overall mortality but a 5% increase in mortality with piperacillin-
306 tazobactam reported in the IV analysis. Although it is technically possible for IV
307 estimates to be more precise and stronger than the raw outcome onto instrument
308 relationship, this is highly unusual. It does not seem plausible that during a period in
309 which there was almost complete changeover from piperacillin-tazobactam to
310 cefepime, and in which there was no change in overall mortality, that cefepime could be
311 associated with a 5% absolute reduction in mortality. The secondary analyses reporting
312 an even larger 12% increase in mortality are equally implausible.

313

314 This highlights the value of examining each step in any analysis, and examining it for
315 plausibility, and when implausible estimates are generated, examining the model inputs
316 and outputs for potential bias and its causes.

317

318 **Conclusion:**

319

320 The reported estimates of benefit of cefepime in Chanderraj et al are likely due to
321 collider bias due to conditioning on metronidazole. Corrected causal analyses identify a
322 much-reduced effect. The authors should consider correcting the original study to
323 include estimates unadjusted for metronidazole.

324

325

326

327

328

329

330 **REFERENCES:**

331

332

333 1. Chanderraj R, Admon AJ, He Y, et al. Mortality of Patients With Sepsis Administered
334 Piperacillin-Tazobactam vs Cefepime. *JAMA Intern Med* [Internet] 2024; Available
335 from: <http://dx.doi.org/10.1001/jamainternmed.2024.0581>

336 2. Qian ET, Casey JD, Wright A, et al. Cefepime vs Piperacillin-Tazobactam in Adults
337 Hospitalized With Acute Infection: The ACORN Randomized Clinical Trial. *JAMA*
338 [Internet] 2023 [cited 2024 Jun 11];330(16):1557–67. Available from:
339 <https://jamanetwork.com/journals/jama/fullarticle/2810592>

340 3. Dressler DD, MSc MD. Cefepime vs. Piperacillin-Tazobactam for Sepsis: The Debate
341 Continues [Internet]. 2024 [cited 2024 Jul 10]; Available from:
342 [https://www.jwatch.org/na57489/2024/05/21/cefepime-vs-piperacillin-](https://www.jwatch.org/na57489/2024/05/21/cefepime-vs-piperacillin-tazobactam-sepsis-debate)
343 [tazobactam-sepsis-debate](https://www.jwatch.org/na57489/2024/05/21/cefepime-vs-piperacillin-tazobactam-sepsis-debate)

- 344 4. Greene WH. Econometric Analysis [Internet]. Prentice Hall; 2003. Available from:
345 <https://play.google.com/store/books/details?id=JkWAQAAMAAJ>
- 346 5. Deuchert E, Huber M. A cautionary tale about control variables in IV estimation. Oxf
347 Bull Econ Stat [Internet] 2017;79(3):411–25. Available from:
348 <https://onlinelibrary.wiley.com/doi/10.1111/obes.12177>
- 349 6. ICH E9 statistical principles for clinical trials - Scientific guideline [Internet]. [cited
350 2024 Jun 3]; Available from: [https://www.ema.europa.eu/en/ich-e9-statistical-](https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline)
351 [principles-clinical-trials-scientific-guideline](https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline)
- 352 7. Griffith GJ, Morris TT, Tudball M, et al. Collider bias undermines our understanding
353 of COVID-19 disease risk and severity [Internet]. bioRxiv. 2020; Available from:
354 <http://dx.doi.org/10.1101/2020.05.04.20090506>
- 355 8. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. Nature
356 Reviews Methods Primers [Internet] 2022 [cited 2023 May 22];2(1):1–21. Available
357 from: <https://www.nature.com/articles/s43586-021-00092-5>
- 358 9. Dobner J, Kaser S. Body mass index and the risk of infection - from underweight to
359 obesity. Clin Microbiol Infect [Internet] 2018;24(1):24–8. Available from:
360 <http://dx.doi.org/10.1016/j.cmi.2017.02.013>