Variation in monitoring: Glucose measurement in the ICU as a case study to preempt spurious correlations

Running title: Glucose for checking potentially spurious correlations

Author List
Khushboo Teotia † – 1 – khushboo@berkeley.edu
Yueran Jia † – 1 – adsef47@gmail.com
Naira Link Woite, MD – 2 – nairalink@hsph.harvard.edu
Leo Anthony Celi, MD, MS, MPH – 1, 2, 3 – lceli@mit.edu
João Matos, MSc* – 1, 4, 5 – jcmatos@mit.edu
Tristan Struja, MD, MSc, MPH* – 1, 6 – tstruja@mit.edu

† shared first authors
* shared last authors

Author Affiliations
1. Laboratory for Computational Physiology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA
2. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
3. Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA
4. Faculty of Engineering, University of Porto (FEUP), Porto, Portugal
5. Institute for Systems and Computer Engineering, Technology and Science (INESCTEC), Porto, Portugal
6. Medical University Clinic, Kantonsspital Aarau, Aarau, Switzerland

Corresponding Author
Naira Link Woite, MD, ORCID 0000-0003-3145-7903
Cell +1 617-735-0555
Email: nairalink@hsph.harvard.edu
MIT - Institute for Medical Engineering and Science, Laboratory for Computational Physiology
MIT, E25-530- 53 Carleton St, Cambridge, MA 02139

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Conflicts of Interest
None of the authors have any conflicts of interest relevant to this work.

Manuscript Information
Abstract word count: 316
Manuscript word count: 2201
Number of figures: 2
Number of tables: 3

Key Words
Sepsis, Glucose, MIMIC-IV, Critical Care, Health Equity, ICU, EHR

Version
September 30, 2023

Author Contributions
All authors contributed to writing the manuscript.
KT and YJ performed the statistical analysis.
NLW wrote the first draft of the manuscript.
LAC, JM, and TS provided idea, concept, technical and administrative support.

Funding
The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

LAC is supported by the NIBIB, under R01 EB001659.
JM was supported by a Fulbright / FLAD Grant, Portugal, AY 2022/2023.
TS is supported by the Swiss National Science Foundation (P400PM_194497 / 1).
Data Availability

The data that support the findings of this study are available in MIMIC-IV with the identifier doi.org/10.1093/jamia/ocx084 publicly available on PhysioNet (https://physionet.org/).

Code Availability

The code that produces the results in this manuscript can be accessed at https://github.com/e754/Glucose-Research, which includes detailed instructions for running the code.
ABSTRACT

Objective: Health inequities can be influenced by demographic factors such as race and ethnicity, proficiency in English, and biological sex. Disparities may manifest as differential likelihood of testing which correlates directly with the likelihood of an intervention to address an abnormal finding. Our retrospective observational study evaluated the presence of variation in glucose measurements in the Intensive Care Unit (ICU).

Methods: Using the MIMIC-IV database (2008-2019), a single-center, academic referral hospital in Boston (USA), we identified adult patients meeting sepsis-3 criteria. Exclusion criteria were diabetic ketoacidosis, ICU length of stay under 1 day, and unknown race or ethnicity. We performed a logistic regression analysis to assess differential likelihoods of glucose measurements on day 1. A negative binomial regression was fitted to assess the frequency of subsequent glucose readings. Analyses were adjusted for relevant clinical confounders, and performed across three disparity proxy axes: race and ethnicity, sex, and English proficiency.

Results: We studied 24,927 patients, of which 19.5% represented racial and ethnic minority groups, 42.4% were female, and 9.8% had limited English proficiency. No significant differences were found for glucose measurement on day 1 in the ICU. This pattern was consistent irrespective of the axis of analysis, i.e. race and ethnicity, sex, or English proficiency. Conversely, subsequent measurement frequency revealed potential disparities. Specifically, males (incidence rate ratio (IRR) 1.06, 95% confidence interval (CI) 1.01 - 1.21), patients who identify themselves as Hispanic (IRR 1.11, 95% CI 1.01 - 1.21), or Black (IRR 1.06, 95% CI 1.01 - 1.12), and patients being English proficient (IRR 1.08, 95% CI 1.01 - 1.15) had higher chances of subsequent glucose readings.

Conclusion: We found disparities in ICU glucose measurements among patients with sepsis, albeit the magnitude was small. Variation in disease monitoring is a source of data bias that may lead to spurious correlations when modeling health data.
INTRODUCTION

Healthcare disparities, especially based on race and ethnicity, have been extensively documented across various medical conditions and stages of disease (1–3). Disparities manifest as unequal access to the healthcare system, differences in the quality of care, variations in medical device performance, and discrepancies in health outcomes not explained by clinical factors (4). They reflect systemic and societal biases that are easily incorporated into artificial intelligence (AI) models if researchers are not aware of them (5).

One of the main challenges in fair AI modeling is missingness handling (6,7). Notably, a 2017 study found that 49 out of 107 electronic health record (EHR)-based risk prediction approaches evaluated did not mention missing data at all (8). A common strategy is to replace missing values by physiologically-normal values. Similarly to clinical practice – where certain readings are only performed if there is a diagnostic suspicion – an “absent test” would be treated by the AI model as a “normal test” (9).

While this is considered a valid approach, it has important drawbacks. The likelihood of detecting an abnormal finding largely depends on how often a test is conducted. Thus, data that is not missing at random can drive spurious correlations, which are noncausal relationships between the input and the outcome which may shift in deployment (10). When the reason for missing data is a result of systemic social discrimination, these biases can be embedded in subsequent AI models, perpetuating and exacerbating existing disparities (11,12). Especially, as frequency of measurements has gained interest in the AI building community (13). As a case study of a potential source of such spurious correlations in medical AI models, we picked
glucose measurement frequency among patients with sepsis admitted to the Intensive Care Unit (ICU).

Sepsis is a severe life-threatening systemic infection that has a significant impact on the health systems (14–16). Up to 90% of ICU patients exhibit elevated glucose levels irrespective of pre-existing diabetes (17–19). While elevated glucose has been linked to severity of illness, and increased mortality rates (20–22). Numerous studies have shown conflicting results, partly attributed to the detrimental effects of hypoglycemia, a consequence of strict glucose control (22–25). Sepsis patients who are admitted to the ICU are particularly vulnerable to blood glucose fluctuations (23,26) due to the inflammatory response and various aspects of care, such as corticosteroid use (27).

This paper aims to investigate potential disparities in glucose monitoring practices. We advocate that this kind of analysis must be done prior to any AI modeling. The primary objective is to determine if racial and ethnic or demographic differences influence the frequency of glucose measurements during sepsis management. By shedding light on this aspect of care, we aim to contribute to a more comprehensive understanding of healthcare inequalities and to provide researchers developing AI models with a framework to prevent potential biases adversely influencing model fairness and equity.
METHODS

This observational retrospective study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (28). The health equity language, narrative and concepts of this paper follows the American Medical Association’s recommendations (29).

Data Extraction

Data was extracted from the publicly-available MIMIC-IV database using SQL via Google’s BigQuery (30). The MIMIC database is maintained by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology and shared via the PhysioNet platform (31). The dataset has been de-identified, and the institutional review boards of the Massachusetts Institute of Technology (No. 040300206) and Beth Israel Deaconess Medical Center (2001-P-001699/14) both approved the use of the database for research. The MIMIC-IV database includes physiologic data collected from bedside monitors, laboratory test results, medications, medical images and clinical progress notes captured in the electronic health record from patients admitted to the ICU between 2008-2019. Approximately 70,000 de-identified medical records are archived in the MIMIC-IV database.

Hypothesis

We hypothesized that the likelihood for a patient to receive measurement as well as the frequency of those measurements are not equally distributed across race and ethnicity, sex, or English proficiency.
Cohort Selection

The following cases are excluded to create a study cohort: patients under 18 years of age, those without sepsis as defined by the sepsis 3 criteria (32), those with length of ICU stay less than 1 day, those with a diagnosis of diabetic ketoacidosis, and with racial description that does not fit within White, Asian, Black, or Hispanic, excluding those of the heterogenous group “other”.

Covariates

We drew directed acyclic graphs to understand which variables to extract (see Supplemental Figure 1 and 2), as well as their interplay. A total of 13 patient-level variables were extracted, including non-time-varying variables such as demographics, comorbidities, and admission information and also time-varying variables including Sequential Organ Failure Assessment (SOFA) score (33), insulin treatment, glucocorticoid treatment, and instances of glucose measurements. Time-varying variables were aggregated in different ways. SOFA was calculated for the day of ICU admission, insulin treatment was used as a binary variable for whether or not it was received on day one. Glucose measurement was also used as a binary variable for whether or not it was measured on day one, in addition to taking the overall number of measurements for the whole ICU stay normalized by length of ICU stay. Glucocorticoid treatment was also normalized by ICU length of stay (see Supplemental Table 1).
Outcomes

We had two primary outcomes: The first was a binary variable predicting whether a patient received measurement on day 1, and the second being a prediction of how many glucose measurements a patient would receive per day throughout the length of their ICU stay.

Statistical Analysis

Statistical analysis was performed using Python 3.10. For summary statistics, we used the `tableone` Python package (34). For the outcome of whether or not a patient received a glucose measurement on day 1, we fitted a penalized, ridge logistic regression using the Python package `statsmodel` (35) adjusted for confounders to estimate the likelihood of receiving a glucose measurement. We report our findings as odds ratios (OR) with 95% confidence intervals (95% CI). White patients were considered as the reference group.

For the outcome of the number of glucose measurements during an ICU stay, we fitted a non-penalized, negative binomial regression (also with `statsmodel`) adjusted for confounders to estimate the number of glucose measurements for each patient each day during their respective ICU stay. We report our findings as incident rate ratios (IRR) with 95% CI. White patients were considered as the reference group.
RESULTS

Baseline Study Cohort
The MIMIC-IV database has 73,181 ICU stays, of which 24,927 (see Figure 1) made up our final cohort following the inclusion and exclusion criteria. The race and ethnicity distribution was 11.8% Black, 3.3% Asian, 3.3% Hispanic, 80.47% White and did not change relevantly after applying exclusion criteria.

In the resulting cohort, English proficiency was distributed unevenly across race and ethnicity with Asian (37.8%) and Hispanic (39.3%) patients being significantly less frequently English proficient than White (95.2%) and Black (89.6%) patients (see Table 1). Insurance coverage also differed between race and ethnicity with White patients having the highest Medicare coverage (52.3%), and Hispanic patients had the highest Medicaid coverage (20.2%), followed by Asian patients (17.5%). Of note, even though the percentage of patients with diabetes was higher for Black (18.6%) and Hispanic (17.9%) patients when compared to White (11.4%) and Asian (14.2%) patients, the probability for receiving a glucose measurement on day 1 was roughly even for all groups (see Supplementary Figure 3). Furthermore, Hispanic (56.1%) patients had the lowest probability of receiving insulin on day 1.

Model Results
We used logistic regression to predict whether a patient received a glucose measurement on day 1 with the null hypothesis that all patients are equally likely to receive a measurement. Asian patients (OR 0.80, 95% CI 0.62 - 1.03), Black patients (OR 1.00, 95% CI 0.87 - 1.16), and Hispanic patients (OR 1.05, 95% CI 0.82 - 1.35) patients were compared to White patients as a reference (see Figure 2A, Table 2,
and Supplementary Figure 4). In addition, the effect of being English proficient (OR 0.98, 95% CI 0.82 - 1.17) and being male (OR 0.94, 95% CI 0.82 - 1.17) were not statistically significant.

A negative binomial model was fitted to predict the total frequency of glucose measurements per patient and length of stay with the null hypothesis that all patients have the same frequency. Being male (IRR 1.06, 95% CI 1.01 - 1.21), Hispanic (IRR 1.11, 95% CI 1.01 - 1.21), Black (IRR 1.06, 95% CI 1.01 - 1.12), or English proficient (IRR 1.08, 95% CI 1.01 - 1.15) significantly increased the frequency of receiving a measurement (see Figure 2B, Table 3, and Supplementary Figure 5), while being Asian (IRR 1.04, 95% CI 0.94 - 1.15) did not.
DISCUSSION

Policymakers have increasingly recognized the importance of health equity to public health(36). In December of 2022, tying reimbursements to health equity outcomes was proposed by the Center for Medicare and Medicaid Services (CMS)(37). If these policies were to be implemented, they would also need to be considered in AI development as disparities could have unintended consequences on a model's prediction (38). There is growing awareness in the domain of medical AI development that potential advancements through AI could be hindered by such biases (39). The low accuracy of the EPIC sepsis model in external datasets is a good example of what can happen if temporal bias is neglected(40). Another study found increased illness burden in Black patients because the model wrongly included the post-treatment cost as a pre-treatment feature(41).

Aside from the risks of inputting biased data into a model, unmeasured data for a particular group reflect quality of care. Absence of data in medicine is not simply a void. In fact, caregivers often do not measure a variable deliberately; the reason to do so is usually highly confounded, affected by ongoing treatment, and reflected in the potential outcome. Models also interpret these voids as information influencing predictions, driving spurious correlations (10). Therefore, it is important to establish a thorough understanding of such associations before fitting any AI model.

To avoid encrypting biases into AI models, many advocate the use of causal inference frameworks to check for, understand, and control for potential biases (42,43). Performing similar analyses as we did – even before fitting a causal inference model
Glucose for checking potentially spurious correlations

should become a common practice, as the awareness and understanding of such incongruences can help to understand these models better.

Furthermore, one key assumption of causal inference models is positivity, i.e. subjects from the treatment and control group should have enough overlap in the distribution of their confounding variables(44). Usually researchers check this assumption by tabulating the data across the most pertinent variables which can become unwieldy in case of many strata. As an additional check, we propose fitting simple regression models as we did to ascertain that the positivity assumption is not being violated and adding frequency of measurements as a distinct feature to any model.

Besides all these implications for AI models, our findings are also relevant clinically. While our study found no differences on glucose measurements in the glucose measurements on day 1 among patients admitted to the ICU due to sepsis, there was a slightly higher frequency of glucose measurement among Black, Hispanic, male, and English proficient patients.

Glucose measurement plays a pivotal role in in-patient care, as glucose levels are linked to severity of the disease and patient outcomes (20,21). In the context of sepsis patients, hyperglycemia is a common bystander associated with severity of disease and mortality (23). Studies have shown that frequent glucose measurements are associated with less hypoglycemia and hyperglycemia episodes, as well as decreased glucose variability (45–48) which has been associated with decreased hospital mortality, decreased length of hospital stay, and composite hospital complications (48).
Reports of Black patients admitted to the ICU having poorer glycemic control are thus troublesome (49).

LIMITATIONS
While our study adds to the discussion of disparities and algorithmic bias in the critical care setting, we acknowledge that it has certain limitations. These include potential selection bias, as the database does not include patients who were admitted to the regular ward, and the data comes from only one academic tertiary care center in the U.S, where most of the population is White (80.47%). Also, our study design does not allow us to test whether the associations are from unmeasured confounding. Future research should strive to address these limitations for a more comprehensive understanding of the topic and should also cover other areas of care such as emergency departments, regular wards, or ambulatory care.

CONCLUSION
The implications of our study goes beyond glucose monitoring during sepsis management. Besides the challenge of achieving healthcare equity in a system marked by systemic biases, researchers need to be aware of such disparities before building any AI model. Such biases cannot only skew predictions, but may even generate further adverse effects when the predictions are being used for treatment or management decisions.
REFERENCES


36. Health Equity and Health Disparities Environmental Scan. FINAL Rep.


Figure 1. Consort diagram, depicting inclusion and exclusion criteria, as well as cohort composition at each step

**Legend:** DKA, Diabetic ketoacidosis; LOS, Length of stay
Figure 2. Main results, respective to race and ethnicity categorizations.

Panel A: Results from logistic regression on measurement on day 1;

Panel B: Results from negative binomial regression on frequency of measurements over whole ICU stay

**Abbreviations**: OR, odds ratio; CI, confidence interval; IRR, incidence rate ratio
Table 1. Baseline information on the study cohort, derived from MIMIC-IV

<table>
<thead>
<tr>
<th>Race and Ethnicity</th>
<th>N (%)</th>
<th>Missing</th>
<th>Overall</th>
<th>Asian</th>
<th>Black</th>
<th>Hispanic</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (%)</td>
<td></td>
<td>24,927 (100)</td>
<td>847 (3.4)</td>
<td>2,962 (11.9)</td>
<td>1,057 (4.2)</td>
<td>20,061 (80.5)</td>
<td></td>
</tr>
<tr>
<td>Age, median [Q1,Q3]</td>
<td>0</td>
<td>68.0 [57.0,78.0]</td>
<td>67.0 [55.0,79.0]</td>
<td>65.0 [54.0,76.0]</td>
<td>59.0 [46.0,70.0]</td>
<td>69.0 [58.0,79.0]</td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td>Male</td>
<td>14351 (57.6)</td>
<td>507 (59.9)</td>
<td>1344 (45.4)</td>
<td>645 (61.0)</td>
<td>11855 (59.1)</td>
<td></td>
</tr>
<tr>
<td>English proficient, n (%)</td>
<td>Yes</td>
<td>22486 (90.2)</td>
<td>320 (37.8)</td>
<td>2654 (89.6)</td>
<td>415 (39.3)</td>
<td>19097 (95.2)</td>
<td></td>
</tr>
<tr>
<td>Insurance, n (%)</td>
<td>Medicaid</td>
<td>0</td>
<td>1715 (6.9)</td>
<td>148 (17.5)</td>
<td>326 (11.0)</td>
<td>214 (20.2)</td>
<td>1027 (5.1)</td>
</tr>
<tr>
<td></td>
<td>Medicare</td>
<td>12509 (50.2)</td>
<td>273 (32.2)</td>
<td>1375 (46.4)</td>
<td>378 (35.8)</td>
<td>10483 (52.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>10703 (42.9)</td>
<td>426 (50.3)</td>
<td>1261 (42.6)</td>
<td>465 (44.0)</td>
<td>8551 (42.6)</td>
<td></td>
</tr>
<tr>
<td>Elective admission, n (%)</td>
<td>Yes</td>
<td>0</td>
<td>1051 (4.2)</td>
<td>25 (3.0)</td>
<td>55 (1.9)</td>
<td>32 (3.0)</td>
<td>939 (4.7)</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td>Present</td>
<td>0</td>
<td>3141 (12.6)</td>
<td>120 (14.2)</td>
<td>552 (18.6)</td>
<td>189 (17.9)</td>
<td>2280 (11.4)</td>
</tr>
<tr>
<td>Charlson comorbidity index, mean (SD)</td>
<td>0</td>
<td>5.4 (2.9)</td>
<td>5.4 (3.0)</td>
<td>5.9 (3.0)</td>
<td>4.9 (3.1)</td>
<td>5.4 (2.9)</td>
<td></td>
</tr>
<tr>
<td>SOFA, median [Q1,Q3]</td>
<td>0</td>
<td>5.0 [3.0,8.0]</td>
<td>6.0 [3.0,8.0]</td>
<td>6.0 [4.0,8.0]</td>
<td>5.0 [4.0,8.0]</td>
<td>5.0 [3.0,8.0]</td>
<td></td>
</tr>
<tr>
<td>Dose of methylprednisolone*, mean (SD)</td>
<td>22,712</td>
<td>57.6 (94.5)</td>
<td>52.9 (63.2)</td>
<td>46.9 (58.8)</td>
<td>84.2 (131.9)</td>
<td>57.9 (97.6)</td>
<td></td>
</tr>
<tr>
<td>Insulin on day 1, n (%)</td>
<td>Present</td>
<td>0</td>
<td>9623 (38.6)</td>
<td>528 (62.3)</td>
<td>1812 (61.2)</td>
<td>593 (56.1)</td>
<td>12371 (61.7)</td>
</tr>
<tr>
<td>Glucose measurement day 1, n (%)</td>
<td>Present</td>
<td>0</td>
<td>22755 (91.3)</td>
<td>319 (37.7)</td>
<td>1150 (38.8)</td>
<td>464 (43.9)</td>
<td>7690 (38.3)</td>
</tr>
<tr>
<td>Length of stay, median [Q1,Q3]</td>
<td>days</td>
<td>3.2 [1.9,6.3]</td>
<td>3.1 [1.8,6.4]</td>
<td>3.3 [1.9,6.5]</td>
<td>3.1 [1.9,6.3]</td>
<td>3.2 [1.9,6.3]</td>
<td></td>
</tr>
</tbody>
</table>
Legend: Race and Ethnicity includes White, Black, Hispanic, and Asian patients. * Dose of glucocorticoids was transformed to be equipotent with methylprednisolone, and normalized by length of stay

Abbreviations: Abbreviations: IQR; interquartile range; SOFA, sequential organ failure assessment; LOS, length of stay
Table 2. Logistic regression results for outcome of glucose measurement on day 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>2.5 % CI</th>
<th>97.5 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.91</td>
<td>4.33</td>
<td>8.06</td>
</tr>
<tr>
<td>Age</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SOFA</td>
<td>1.09</td>
<td>1.07</td>
<td>1.11</td>
</tr>
<tr>
<td>Charlson comorbidity index</td>
<td>1.05</td>
<td>1.03</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>Insurance:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicaid</td>
<td>0.86</td>
<td>0.72</td>
<td>1.03</td>
</tr>
<tr>
<td>Medicare</td>
<td>0.97</td>
<td>0.87</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>Race and Ethnicity:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0.80</td>
<td>0.62</td>
<td>1.03</td>
</tr>
<tr>
<td>Black</td>
<td>1.00</td>
<td>0.87</td>
<td>1.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.05</td>
<td>0.82</td>
<td>1.35</td>
</tr>
<tr>
<td><strong>Anchor year:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 - 2010</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011 - 2013</td>
<td>0.90</td>
<td>0.81</td>
<td>1.01</td>
</tr>
<tr>
<td>2014 - 2016</td>
<td>1.05</td>
<td>0.93</td>
<td>1.20</td>
</tr>
<tr>
<td>2017 - 2019</td>
<td>0.98</td>
<td>0.84</td>
<td>1.14</td>
</tr>
<tr>
<td><strong>Binary variables:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male sex</td>
<td>0.94</td>
<td>0.86</td>
<td>1.03</td>
</tr>
<tr>
<td>Elective admission</td>
<td>0.82</td>
<td>0.65</td>
<td>1.04</td>
</tr>
<tr>
<td>English proficient</td>
<td>0.98</td>
<td>0.82</td>
<td>1.17</td>
</tr>
<tr>
<td>Major surgery</td>
<td>1.92</td>
<td>1.73</td>
<td>2.14</td>
</tr>
<tr>
<td>Glucocorticoids day one</td>
<td>1.09</td>
<td>0.90</td>
<td>1.31</td>
</tr>
<tr>
<td>English proficient</td>
<td>0.98</td>
<td>0.82</td>
<td>1.17</td>
</tr>
<tr>
<td>COPD present</td>
<td>1.00</td>
<td>0.84</td>
<td>1.20</td>
</tr>
<tr>
<td>Diabetes present</td>
<td>1.72</td>
<td>1.43</td>
<td>2.05</td>
</tr>
<tr>
<td>Asthma present</td>
<td>1.38</td>
<td>0.55</td>
<td>3.47</td>
</tr>
</tbody>
</table>

Abbreviations: COPD, Chronic obstructive pulmonary disease; SOFA, sequential organ failure assessment
Table 3. Negative binomial regression results for outcome of frequency of glucose measurement during length of stay

<table>
<thead>
<tr>
<th>Variable</th>
<th>IRR</th>
<th>2.5 % CI</th>
<th>97.5 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.17</td>
<td>2.82</td>
<td>3.57</td>
</tr>
<tr>
<td>Age</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SOFA</td>
<td>1.03</td>
<td>1.02</td>
<td>1.03</td>
</tr>
<tr>
<td>Charlson comorbidity index</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>Methylprednisolone dose normalized by LOS</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Insurance:**

- Other (Reference)
- Medicare: 0.98, 0.94, 1.02
- Medicaid: 0.94, 0.88, 1.01

**Race:**

- White (Reference)
- Asian: 1.04, 0.94, 1.15
- Black: 1.06, 1.01, 1.12
- Hispanic: 1.11, 1.01, 1.21

**Anchor year:**

- 2008 - 2010 (Reference)
- 2011 - 2013: 0.96, 0.92, 1.00
- 2014 - 2016: 1.01, 0.96, 1.05
- 2017 - 2019: 0.93, 0.88, 0.98

**Binary variables:**

- Male sex: 1.06, 1.03, 1.10
- Elective admission: 1.46, 1.34, 1.58
- Insulin on day 1: 1.69, 1.64, 1.75
- English proficient: 1.08, 1.01, 1.15
- Major surgery: 1.57, 1.51, 1.63
- COPD present: 0.93, 0.87, 0.99
**Diabetes present** | 1.36 | 1.29 | 1.43

**Asthma present** | 0.94 | 0.68 | 1.29

**Abbreviations**: IRR, incidence rate ratio; COPD, chronic obstructive pulmonary disease; SOFA, sequential organ failure assessment; LoS, length of stay

*Glucose for checking potentially spurious correlations*
SUPPLEMENTARY MATERIALS

Supplementary Figure 1. Directed acyclic graph for outcome of a single glucose measurement on day 1 of hospital admission

Legend: Race includes White, Black, Hispanic, and Asian patients.

Abbreviations: IQR; interquartile range; COPD, chronic obstructive pulmonary disease; SOFA, sequential organ failure assessment; CCI, Charlson Comorbidity Index. Lab, laboratory
Supplementary Figure 2. Directed acyclic graph for outcome of frequency of glucose measurement over the hospital admission

Legend: Race includes White, Black, Hispanic, and Asian patients.

Abbreviations: IQR; interquartile range; COPD, chronic obstructive pulmonary disease; SOFA, sequential organ failure assessment; CCI, Charlson Comorbidity Index. Lab, laboratory
Supplementary Figure 3. Boxplot of number of glucose measurements per LOS binned by race-ethnicity and stratified by diabetes status

Legend: Race includes White, Black, Hispanic, and Asian patients. Orange horizontal bar denotes the median, while the box encompasses the first quartile to the third quartile. The whiskers denote 1.5x the interquartile range. Points are outliers.

Abbreviations: #, number; LOS, length of ICU stay.
Glucose for checking potentially spurious correlations

Supplementary Figure 4. Logistic Regression model forest plot

Supplementary Figure 5. Negative Binomial Model Forest Plot
## Supplementary Table 1. Definition of considered covariates

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Description</th>
<th>Handling of missing values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>As provided by dataset</td>
<td>N/A</td>
</tr>
<tr>
<td>Age</td>
<td>At admission</td>
<td>N/A</td>
</tr>
<tr>
<td>Race Ethnicity</td>
<td>Grouped into Asian, Black, White, Hispanic, and Other</td>
<td>Exclusion of group &quot;other&quot;</td>
</tr>
<tr>
<td>Insurance</td>
<td>As provided by dataset, Medicare/Medicaid or other</td>
<td>N/A</td>
</tr>
<tr>
<td>English Proficiency</td>
<td>As provided by dataset, English or limited</td>
<td>N/A</td>
</tr>
<tr>
<td>Year Group</td>
<td>As provided by dataset, bi-yearly bins</td>
<td>N/A</td>
</tr>
<tr>
<td>Elective Admission</td>
<td>As provided by dataset, true or false if admission is elective</td>
<td>N/A</td>
</tr>
<tr>
<td>Major Surgery</td>
<td>As provided by OASIS score</td>
<td>N/A</td>
</tr>
<tr>
<td>SOFA</td>
<td>SOFA score with each of its subcomponents: on admission and for the selected 24 hours, aggregated by the maximum value</td>
<td>Assumption of best possible value in case of missing</td>
</tr>
<tr>
<td>Charlson comorbidity Index</td>
<td>As provided by dataset</td>
<td>N/A</td>
</tr>
<tr>
<td>COPD</td>
<td>ICD-10 codes J41.X-J47.X</td>
<td>N/A</td>
</tr>
<tr>
<td>Asthma</td>
<td>ICD-10 codes J84.1X</td>
<td>N/A</td>
</tr>
<tr>
<td>Diabetes</td>
<td>ICD-10 codes E08.X-E11.X, and E13.X</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>ICD-10 codes L94.0X, L94.1X, L94.3X</td>
<td>N/A</td>
</tr>
<tr>
<td>DKA present</td>
<td>ICD-10 codes E10.1X, E11.1X, E13.1X</td>
<td>N/A</td>
</tr>
<tr>
<td>Insulin</td>
<td>MIMIC-IV Item IDs 229299, 229619, 223257, 223258, 223259, 223260, 223261, 223262</td>
<td>Assume no insulin received if no entries.</td>
</tr>
<tr>
<td>Glucose</td>
<td>MIMIC-IV Item IDs 220621, 228388, 25664, 226537</td>
<td>Assume no glucose measurement if no entries.</td>
</tr>
<tr>
<td>Methylprednisolone dose normalized by LOS</td>
<td>Only non-topical formulations of Dexamethasone, Hydrocortisone, Methylprednisolone, and Prednisone were considered. Methylprednisolone equivalent doses were calculated by multiplying the dose with following factors: Dexamethasone x5 Hydrocortisone x0.2 Methylprednisolone x1 Prednisone x1</td>
<td>Assume no glucocorticoid received if no entries.</td>
</tr>
</tbody>
</table>