The Obesity Paradox in ICU Patients

Janice Pan†, Robert Shaffer†, Zeina Sinno∗†, Marcus Tyler∗†, and Joydeep Ghosh†, Fellow, IEEE

Abstract— Excessive weight is connected with an increased risk of certain life-threatening diseases. However, some evidence shows that among patients with chronic diseases such as heart failure (HF), chronic kidney disease (CKD), and COPD, increased weight is paradoxically associated with a decreased risk of mortality. This counterintuitive phenomenon is referred to as the obesity paradox. The obesity paradox has been mostly observed among certain cohorts of patients with HF, but not specific to patients in the Intensive Care Unit (ICU) setting. This paper studies the relationship between obesity and mortality of ICU patients with and without HF and presents evidence supporting the existence of this paradox. The results provide helpful insights for developing more patient-centric care in ICUs. Additionally, we use both the MIMIC-II and MIMIC-III databases, for which few comparative studies exist to date. We demonstrate an aspect of consistency between the databases, providing a significant step towards validating the use of the newly announced MIMIC-III in broader studies.

I. INTRODUCTION

Congestive heart failure (HF) is a disorder of the heart in which the heart has trouble pumping sufficient blood to meet the needs of the body [1]. Though the condition usually develops with age, anyone can contract heart failure, which is considered a serious, chronic (i.e., long-term) condition [2]. A significant risk factor for developing heart disease, among other serious, life-threatening diseases such as high blood pressure, diabetes, and cancer, is obesity [3]. Studies, however, have demonstrated that obese patients with certain severe conditions sometimes have better prognoses and survival rates than their non-overweight or non-obese counterparts. This phenomenon is known as the obesity paradox and has been observed in cohorts of heart failure patients [4, 5]. Paradoxically, though obesity significantly increases the risk for developing new-onset heart failure, overweight and mildly-obese patients with chronic heart failure tend to have better prognosis and survival compared to lean or normal-weight patients [6]. This paradox is not well accepted as certain scholars attribute it to biases in observational studies that ignore confounding factors such as excessive smoking.

In this work, we carefully examine the obesity paradox in the context of Intensive Care Unit (ICU) patients, who are generally in critical condition and in worse health than the non-ICU cohort. We use the MIMIC-II [7,8] and MIMIC-III [9] databases, which contain the ICU records and health-related data of more than 32,000 and 45,000 patients, respectively, from the Beth Israel Deaconess Medical Center, and we consider only HF patients without differentiating between types (e.g. systolic, diastolic, acute). The MIMIC-III database was only recently released on December 10, 2015 [10], so part of our contribution is also to provide an analytic comparison between the results from both databases. MIMIC-III is an augmented and improved version of MIMIC-II, as it contains newly collected data and regenerated entries of previously collected data.

To the best of our knowledge, the obesity paradox has not been observed, explored, or quantified among ICU patients with HF. If there is strong evidence to support the existence of the paradox for this group, then caregivers may find that non-obese patients require special attention or treatments. Since healthcare is moving towards more patient-centric systems, it is crucial to understand how to provide individualized treatments to patients based on pre-existing conditions. Moreover, improved allocation of medical staff and resources may increase patient survival rates.

II. RELATED WORK

In this section, we review relevant work to the obesity paradox for both Chronic Heart Failure (CHF) and ICU patients and provide possible explanations of this paradox.

For the general population, factors such as hypercholesterolemia [11], high BMI [12], and hypertension [13] are deleterious to health. However, these factors seemed to have a more desirable effect in the cohort of patients with CHF in several studies. Lavie, a cardiologist in Jefferson, Louisiana, was one of the first clinicians to observe this paradox [14]. Horwich et al. [15, 16] later showed that hypercholesterolemia was associated with a lower risk factor for mortality of CHF patients. Similar observations were drawn for the cases of high BMI [17-19], hypertension [19-21], and for when these two factors are combined [22].

There are several explanations for the existence of the obesity paradox. One is that HF patients have a mortality risk that is significantly greater than that of the general population [23]. In other words, the severity of the onset of HF can cause mortality in a shorter period of time, which may conceal the long-term adverse effects of risk factors such as obesity. Another explanation is the predisposition of malnourished patients to infections. The absence of these obesity-related risk factors may indicate a state of undernutrition, which may predispose the patients to infections or other inflammatory processes [24], which in some cases can lead to mortality.

A third explanation may be that many CHF patients may have physiological characteristics different than those of the general population [22]. Many patients with heart disease die before reaching the more-severe diagnosis of CHF. Only a
small fraction (roughly one fifth) are diagnosed with CHF [22], which might suggest that those patients suffering from CHF despite conventional risk factors may have stronger resistance to some risk factors, e.g., tissue necrosis. A series of studies conducted by Lavie et al. [4, 6, 25] found that overweight patients may be resistant to tissue necrosis that often accompanies late-stage HF. The authors suggest that overweight patients may have greater metabolic reserves than their healthy-weight counterparts, which allows them to withstand stressful events more easily.

Finally, yet another explanation may be selection bias, which suggests that doctors might monitor obese patients’ cardiovascular health more closely than they monitor the health of non-obese patients [25]. Similarly, doctors may place obese patients on more aggressive treatment plans than healthy-weight patients. In these scenarios, the obesity paradox may actually represent a form of survival bias, in which the unhealthiest patients are more closely monitored than their healthier counterparts.

Less effort has been made towards studying the obesity paradox under ICU setting. Pickkers et al. [26] showed an inverse relationship between obesity and mortality in critically ill patients. Arabi et al. [27] challenged the validity of the obesity paradox concept for critically ill patients with septic shock and found that obesity did not significantly improve survival. Another study by Utzolino et al. [28] suggested that the obesity paradox may exist with surgical peritonitis, as short-term outcomes were improved for patients with obesity, but long term prognoses did not improve. Hutagalung et al. [29] showed that for patients in the surgical ICU, being overweight or obese was associated with a decreased risk of 60-day mortality.

III. DATA COLLECTION

As mentioned, the publicly available MIMIC-II and MIMIC-III databases contain the ICU records and health data of tens of thousands of patients. Because many patients had multiple ICU stays, we take the first-recorded stay information, including the first-recorded weight and height, for analysis. Also, because we require BMI information for determining obesity severity, we only use patients for whom both height and weight information is recorded.

Among these records, we select adults between the ages of 28 and 90. We use 90 years as a maximum age cut-off because the true ages for patients older than 90 years were obscured. Our minimum cut-off of 28 was chosen to be two standard deviations below the mean age among adults.

To compute BMI, we use the equation:

$$\text{BMI} = \frac{\text{Weight}[\text{kg}]}{(\text{Height}[\text{m}])^2}$$  \hspace{1cm} (1)

and filter out records with BMI values outside of the Biologically Informative Value range [12 kg/m², 70 kg/m²].

To classify HF patients in both MIMIC-II and MIMIC-III, we use ICD-9 data and only consider patients with records that contained a code under the 428 class. The ICD-9 code 428 is the umbrella code for HF, and its sub-diseases include congestive heart failure (428.0), systolic heart failure (428.2), diastolic heart failure (428.3), and similar diagnoses. This class of ICD-9 codes also includes that of acute heart failure. We compared results with and without adding acute heart failure records in MIMIC-II and found no significant differences. Therefore, each patient in our HF cohorts may have any of the various types of heart failure.

Additionally, for the MIMIC-III database, in the admissions table, there is a diagnosis field, which contains a preliminary, free-text diagnosis for the patient upon admission to the hospital, which is usually recorded by the admitting clinician. We also use this free-text field to identify patients with heart failure if their diagnosis field contains any of the following keywords: CORONARY, HEART, and MYOCARDIAL. We also implemented negation detection and manually checked over flagged phrases, which included diagnoses of non-Q wave myocardial infarction (MI) and non-ST elevation myocardial infarction, but none negated the HF classification.

Finally, to compute mortality, we use two definitions, both based on the 90-day post-discharge period. One is a binary variable, indicating if the patient died within the 90 days post-discharge. A 1 for this binary variable indicates the patient died during their ICU stay or during the 90 days post-discharge, whereas a 0 indicates the patient survived the 90 days post-discharge. The second is a continuous variable in [0,1], where 1 indicates that the patient died during their ICU stay, and 0 indicates that the patient survived 90 days. Specifically, if a patient survived \(x \in [0,90]\) days post-discharge (\(x = 0\) if the patient died in the ICU), their fractional mortality score \(m\) would be computed as \(m = \frac{90-x}{90}\). The Social Security Death Index (SSDI) was used to determine out-of-hospital mortality in both databases. The MIMIC-II (v2.6) database allows for mortality prediction of 0.75 years. The MIMIC-III (v1.3) consists of data from two database sources, Metavision [30] and Carevue [31], which allow for 90-day and 4-year mortality predictions, respectively. We chose to use the 90-day post-discharge period for mortality prediction so we could use all the available data from both MIMIC-II and MIMIC-III.

IV. METHODS AND ANALYSIS

To gather evidence supporting the obesity paradox, we use both Pearson’s correlation coefficient (PLCC) as well as overall mortality rates. In both cases, we split the cohorts into BMI bands to better see trends along a BMI spectrum. For correlations, we compute the PLCC between BMI and the continuous mortality variable. We use a BMI band size of 9 for the coefficients in an effort to increase our sample sizes and compute meaningful confidence intervals [32]. Figure 1 shows the correlation results for patients with and without HF, with 95% confidence intervals. In our work, the null hypothesis is that HF does not affect the correlations, i.e. that the HF and Non-HF cohorts are statistically
equivalent. In Figure 1, we demonstrate the statistical significance when the confidence interval for one cohort does not overlap with the estimate for the other cohort. This occurs mostly with lower BMI values, e.g., BMIs of 17 (MIMIC-II), 26 (MIMIC-III), and 35 (MIMIC-II). Of these statistically significant values, we see that lower BMIs (<30) are more highly correlated with mortality for the HF cohort, whereas the trend is reversed for higher BMIs (>35).

For overall mortality, we compute mortality percentages using the binary mortality variable. We use a band size of 6 for the percentages, which maximizes resolution while minimizing discrepancies (likely due to noise from the smaller sample sizes) between MIMIC-II and MIMIC-III. Figure 2 shows the mortality percentages for patients with and without HF. In either dataset, mortality steadily decreases in the HF group as BMI increases. However, for the non-HF population, mortality drops until BMI is about 30 and then starts to increase with increasing BMI. The observed differences between the two populations in both datasets are strong evidence of a paradox.

Figure 1. Correlations between BMI and mortality for HF and non-HF with 95% confidence intervals.

Figure 2. Mortality percentages mortality for HF and non-HF patients.

A. Matching
To account for possible selection bias, we generated four matching estimates corresponding to the average treatment effect induced by moving from healthy weight to overweight after matching on two sets of features. In matching problems, the goal is to pair treated and non-treated observations such that the difference between the treatment (BMI > cutoff) and control (BMI < cutoff) groups on observed covariates is as small as possible.

If successful (and if matching covariates are chosen appropriately), any remaining differences between the groups can be attributed solely to the treatment. For our analysis, we generated matching estimates using Diamond and Sekhon’s [33] GenMatch algorithm. As before, we used BMI cutoffs of 25 and 30 to divide the cohorts into groups, and matched on two feature sets. In the first set, we solely considered demographic data: ethnicity, gender, and age. The second feature set contains the abovementioned demographic covariates in addition to other diagnostic covariates, which were identified through a literature survey as being connected with late-stage chronic heart failure (e.g., cardiogenic shock, end-stage chronic kidney disease, pneumonia). We also included a set of other diagnosis covariates through a so-called “propensity score.” To generate this propensity score, we regressed all ICD-9 diagnosis codes not already included in our covariate set on 90-day mortality codes using a 10-fold cross-validated LASSO model. We then regressed all diagnosis codes with non-zero coefficients in the LASSO model on obesity status (using an ordinary logistic regression model), and estimated a probability of obesity for each individual. This probability (the “propensity score”) refers to the probability that a given individual was obese conditional on observed covariates.

Using these features, we then used GenMatch to generate an optimal set of feature weights and matched each treated observation to the closest control observation (averaging across matches when multiple equally-good matches existed). We then discarded unmatched control observations and treated observations that could not be matched to a control observation within 0.25 standard deviations on each feature. Balance results for each of these matched datasets are shown in Figure 3. In matching problems, the goal is to balance covariates without limit, so any differences in matched covariates between the treatment (BMI > cutoff) and control (BMI < cutoff) groups post-matching is undesirable. At the very least, however, we should not be able to reject the hypothesis that the groups are different on any covariates included in the model. Based on this criterion, all eight matched datasets achieve balance.

Interpreting the results based on the maximal set of covariates, however, should be approached with caution. In the causal inference setting, treatment estimates generated after conditioning on a so-called post-treatment variable (a variable observed after treatment is assigned) usually require heroic assumptions to achieve unbiasedness (see Rosenbaum [35]), even if those post-treatment variables are strongly related to the outcome, e.g., surgical interventions in the obesity/heart failure analysis. Unfortunately, in the obesity case, this post-treatment bias is difficult to address. As Hernan [34] notes, obesity lacks many qualities of an ordinary treatment, such as a drug or surgical intervention. Most notably, the exact time at which a person becomes obese is often unclear, and people can cycle in and out of obesity status over the course of their lives. As a result, it is often unclear whether a person contracted a given biomedical variable pre- or post-obesity.
With these cautionary notes in mind, we used our matched datasets to compute and plotted average mortality rates to further explore the relationship between BMI and mortality. Figure 4 shows average mortality rates and supports the correlations-based results discussed previously. In all specifications, the treatment effect is estimated to be negative, ranging from approximately a 1% to a 10% difference in mean mortality rates across the eight analyses. In all specifications, the effect is slightly larger in MIMIC-III than MIMIC-II. The effect is significant across all specifications and larger in magnitude when the BMI cutoff is 25. Using Abadie-Imbens standard errors as suggested by [33] to incorporate uncertainty induced by the matching process, the $p$-value associated with treatment effect estimates in these analyses ranges from $p < 0.01$ (demographics, MIMIC-III) to $p = 0.045$ (demographics and diagnostics, MIMIC-II). By contrast, the estimated effect is not significantly different from when the BMI cutoff is 30, save for the demographics-only, MIMIC-III specification. As before, these results suggest that the marginal effect of increasing BMI is nonlinear and decreasing.

Again, because assigning causality to obesity status is problematic, we emphasize that these results should be interpreted with caution. However, the consistency of our estimates across different treatment definitions and coefficient sets suggests that the obesity paradox cannot be easily explained away by simple selection bias, at least for overweight rather than obese patients.

B. MIMIC-II vs. MIMIC-III

Throughout this section, we have shown results from both MIMIC-II and MIMIC-III, and the correlations, plots, and interpretations all support the same conclusions regarding the obesity paradox. That is, our analysis using both databases supports the existence of the obesity paradox. For both databases, we also show matching analysis that, should the obesity paradox exist, eliminates explanations for it. Having both databases produce similar results validates our methods and supports our conclusions further. No previous work has compared similar cohorts of patients between MIMIC-II and MIMIC-III. Our results show an aspect of consistency between the databases, which is a step in validating the use of MIMIC-III for a broader set of studies.

V. CONCLUSION

This paper provided further evidence supporting the existence of the obesity paradox for HF patients represented in both the MIMIC-II and MIMIC-III databases for ICU records. Among the analyzed HF patients, the correlation between mortality and BMI decreases or stays constant for overweight or mildly obese patients, while among non-HF patients, the correlation increases. Similarly, when looking at the relationship between mortality percentage and BMI, while the mortality percentage seems to level out or increase for non-HF patients with higher BMI’s, it actually decreases for the HF cohort. The mortality rates remain significantly different after matching between non-overweight and overweight groups on two different covariate sets, though most estimated treatment effects are not significant with a BMI cutoff of 30. These three analyses on both databases support existence of the paradox, and potential causal or other explanatory reasons for these observations have been explored as well. We should keep in mind that additional parameters and covariates must be accounted for when...
considering the causal interpretability of our analyses and to gain a clearer insight into this paradox.

Our methods applied to both MIMIC-II and MIMIC-III also allow us to compare the cohorts in each database. Extensive comparative analysis has not yet been done using both databases, as MIMIC-III was only recently released. Having drawn the same conclusions from the results obtained using both databases, we validate the use of MIMIC-III in the context of our problem. Any future work comparing the databases can then contribute further to a more comprehensive validation of MIMIC-III.

REFERENCES


